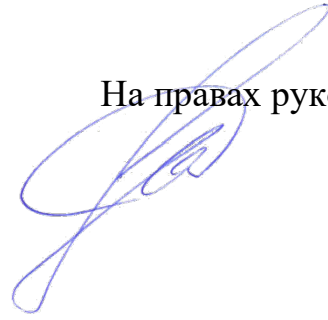


На правах рукописи



КАЛИНИН МАКСИМ ЮРЬЕВИЧ

**МАРКОВСКИЕ МОДЕЛИ И АЛГОРИТМЫ КЛАССИФИКАЦИИ
ИНФОРМАЦИОННЫХ СИГНАЛОВ С МНОГОМЕРНЫМИ
ВЕРОЯТНОСТНЫМИ СВОЙСТВАМИ**

2.3.1. Системный анализ, управление и обработка информации, статистика

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
кандидата технических наук

Москва – 2024

Работа выполнена в АО «Научно-исследовательский институт вычислительных комплексов им. М.А. Карцева», г. Москва, Россия

Научный руководитель: Мутин Денис Игоревич, доктор технических наук, доцент, ведущий специалист АО «Научно-исследовательский институт вычислительных комплексов им. М.А. Карцева», г. Москва, Россия

Официальные оппоненты:

1) Атласов Игорь Викторович, доктор физико-математических наук, профессор, профессор кафедры естественнонаучных дисциплин учебно-методического комплекса информационных технологий (УМК ИТ) Федерального государственного казенного образовательного учреждения высшего образования «Московский университет Министерства внутренних дел Российской Федерации имени В.Я. Кикотя», г. Москва, Россия.

2) Шмырин Анатолий Михайлович, доктор технических наук, профессор, заведующий кафедрой высшей математики Федерального государственного бюджетного образовательного учреждения высшего образования «Липецкий государственный технический университет», г. Липецк, Россия.

Ведущая организация:

Федеральное государственное бюджетное образовательное учреждение высшего образования «Воронежский государственный технический университет», г. Воронеж, Россия.

Защита диссертации состоится 21 мая 2025 г. в 16 часов на заседании диссертационного совета 75.1.026.01 в АО «Научно-исследовательский институт вычислительных комплексов им. М.А. Карцева» по адресу: 117437, г. Москва, ул. Профсоюзная, д. 108.

С диссертацией можно ознакомиться в библиотеке АО «Научно-исследовательский институт вычислительных комплексов им. М.А. Карцева» и на сайте <https://niivk.ru/>

Автореферат разослан «___» _____ 2025 г.

Ученый секретарь
диссертационного совета 75.1.026.01
канд. техн. наук, доцент

 Е.И. Мутина

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Информационные сигналы возникают в различных областях человеческой деятельности, например, при обмене сообщениями, контроле и диагностике состояния технических объектов и окружающей среды, в медицинских исследованиях. Одной из актуальных задач является выявление и идентификация информационных сигналов с целью последующей классификации состояний рассматриваемых объектов, (устройств, технических систем, сердечно-сосудистой системы человека и т.д.).

Необходимость классификации сигналов возникает в различных областях науки, техники и медицины. Работа авиационного двигателя или турбины электростанции порождает вибрации, которые с помощью датчиков преобразуются в электрические сигналы, по которым можно принимать решения о состоянии контролируемого устройства (например, нормальное, предаварийное и авария). Изучение кардиографических сигналов (кардиограмм) дает возможность выявлять состояние сердечно-сосудистой системы человека (норма, патология). В радиотехнической разведке по принимаемым сигналам можно определять их свойства (например, вид модуляции).

В математике и технической кибернетике аналогичные общие задачи решаются в теории распознавания образов, в том числе на основе обучаемых нейронных сетей. Случайные сигналы изучаются в теории вероятностей и теории случайных процессов. Процедуры обучения и самообучения технических систем изучаются, например, в системах искусственного интеллекта. Подобные задачи классификации рассматриваются в теории многомерной статистической классификации, кластерного и факторного анализа.

Марковские одномерные и двухсвязные модели случайных процессов обладают широкой универсальностью, гибкостью и удобны для использования в задаче классификации.

Одной из проблем в исследовании сигналов и методов их обработки является моделирование и имитация случайных сигналов с заданными вероятностными свойствами. Алгоритмы имитации случайных сигналов на основе их марковских моделей позволяют с высокой скоростью и точностью формировать потоки случайных чисел с произвольной заданной двумерной или трехмерной плотностью вероятностей, которые целесообразно использовать для исследования и отладки алгоритмов классификации случайных процессов.

Известные подходы к задаче классификации случайных процессов на основе простых (например, гауссовских) или специализированных (Релея, Накагами и др.) математических моделей не позволяют обеспечить достоверное различение классов, особенно в условиях априорной неопределенности в статистических свойствах сигналов. Марковские модели лишены указанных недостатков, они имеют большое число независимых параметров (элементов матрицы переходных вероятностей) и моделируют случайные процессы с произвольными двумерными и трехмерными плотностями вероятностей.

Актуальность исследования обосновывается необходимостью разработки и исследования новых эффективных (оптимальных по критерию максимальной

апостериорной вероятности) алгоритмов классификации случайных процессов на основе их марковских моделей.

Работа выполнена в соответствии с научным направлением разработки математического обеспечения классификации состояний рассматриваемых объектов.

Степень разработанности темы. Фундаментальными исследованиями в теории распознавания образов посвящены работы Ю.И. Неймарка, В.Н. Фомина, С. Хайкина, У. Гренандера, Э. Патрика, К. Фукунаги, в теории вероятностей, математической статистики и случайных процессов Е.С. Вентцеля, И.И. Гихмана, Е.Б. Дынкина, В.А. Казакова, В.С. Пугачева, А.В. Скорохода, в области факторного, дискриминантного и кластерного анализа И.Д. Мандела, Б. Дюрана, Дж.-О. Кима, П. Оделла, Г. Хармана и др. Исследованиям в области теории сигналов и их обработки посвящены работы Б.Р. Левина, А.Б. Сергиенко, Л.М. Финка, Э. Айфичера, Б. Джервиса, Р. Лайонса, Л. Френкса, фундаментальные результаты в теории информации получены в трудах Б.Д. Кудряшова, Р. Галагера, Я. Кёрнера, С. Кульбака, И. Чисара, в теории обучающихся систем М.Б. Беркинблита, С.И. Николенко, А.Л. Тулупьева, Э.Г. Барто, Р.С. Саттона, С. Симрнза, Б. Уидроу.

Задача классификации случайных процессов в различных вариантах касается всех указанных направлений исследований и рассматривается либо применительно к специализированным моделям сигналов (например, для моделей узкополосных сигналов в радиотехнических алгоритмах их различения), либо к достаточно общим моделям (например, в теории распознавания образов), что приводит к сложным и трудно реализуемым вычислительным алгоритмам формирования решения.

Вопросы разработки оптимальных по критерию максимальной апостериорной вероятности алгоритмов классификации информационных сигналов на основе их марковских моделей рассмотрены недостаточно и имеют существенные отличия от известных исследований. Предлагается разработка односвязной и двухсвязной марковских моделей случайного процесса для формирования моделей классов, обучаемых алгоритмов классификации сигнала и имитации случайных сигналов с заданными двумерными и трехмерными вероятностными свойствами. Для исследования свойств алгоритмов классификации предлагается использование теории марковских процессов, математической статистики и теории информации.

Объектом исследования является информационный сигнал и процедура принятия решения о его принадлежности к одному из классов, связанных с состоянием рассматриваемой системы.

Предметом исследования являются математические модели случайных процессов, алгоритмы классификации информационных сигналов, процедуры их обучения, алгоритмы имитации случайных сигналов с заданными вероятностными свойствами.

Цель и задачи исследования. Целью диссертационного исследования является разработка на основе простой и двухсвязной марковской модели случайных процессов алгоритмов классификации информационных сигналов,

необходимого для этого программного обеспечения и алгоритмов имитации случайных сигналов с заданными многомерными вероятностными свойствами.

Для достижения поставленной цели решены следующие частные научные задачи:

1) анализ проблем построения и разработка многомерных марковских моделей потоков данных, максимизирующих апостериорные вероятности принимаемых решений при динамической идентификации потоков данных;

2) разработка алгоритма формирования эталонных потоков данных с заданными вероятностными свойствами на основе многомерных марковских моделей;

3) разработка и исследование алгоритма классификации потоков данных для системы машинного обучения;

4) построение системы классификации потоков данных с заданной достоверностью.

Методология и методы исследования. Для решения поставленных задач используются методы теории вероятности, математической статистики, марковских случайных процессов, теория распознавания образов, статистическая теория принятия решений, теория самообучающихся технических систем, теории информации. Общей методологической основой является системный подход.

Научная новизна. В диссертации получены следующие результаты, характеризующиеся научной новизной:

двухмерные и трехмерные марковские модели потоков данных, отличающиеся максимизацией апостериорных вероятностей принимаемых решений и обеспечивающие динамическую идентификацию потоков данных;

алгоритм формирования эталонных потоков данных с заданными вероятностными свойствами, отличающийся возможностью получения потока данных с произвольным двухмерным распределением вероятностей, определяемым соответствующей марковской моделью, и позволяющий формировать обучающие выборки для систем машинного обучения;

алгоритм классификации потоков данных для системы машинного обучения, основанный на формировании решающих статистик, получаемых с помощью марковских моделей потоков данных, обеспечивающий максимум отношения правдоподобия;

архитектура системы классификации потоков данных, отличающаяся интеграцией модуля генерации потоков данных с заданными вероятностными свойствами в подсистему обработки потоков данных и обеспечивающая классификацию сигналов по поступающей выборке отсчетов с заданной достоверностью.

Теоретическая значимость заключается в том, что новые предлагаемые методы формирования моделей дискретизированных случайных сигналов и разработанные алгоритмы классификации потоков данных, основанные на односвязных и двухсвязных цепях Маркова, могут быть применены при формировании теоретических и экспериментальных моделей информационных сигналов и для разработки программных комплексов их классификации.

Практическая значимость работы заключается в повышении

достоверности принимаемых решений по выявлению, идентификация и классификации информационных сигналов, возникающих в различных областях человеческой деятельности. Разработан комплекс программ, позволяющий осуществлять классификацию информационных сигналов (на примере радиосигналов). Предложены рекомендации по исследованию ритмограмм для решения задачи классификации состояний сердечно-сосудистой системы пациента.

Достоверность результатов подтверждается использованием при разработке моделей известных математических методов, результатами вычислительных экспериментов и статистического имитационного моделирования.

Положения, выносимые на защиту:

1) многомерные марковские модели потоков данных для систем машинного обучения позволяют повысить достоверность принимаемых решений и обеспечить динамическую идентификацию потоков данных;

2) алгоритм формирования эталонных потоков данных с заданными вероятностными свойствами обеспечивает формирование потока данных с произвольным двухмерным и трехмерным распределением вероятностей на основе марковской модели и обучающих выборок для систем машинного обучения;

3) алгоритм классификации потоков данных для системы машинного обучения на основе решающих статистик, получаемых с помощью марковских моделей потоков данных, позволяет повысить достоверность принимаемых решений;

4) архитектура системы классификации потоков данных с интеграцией модуля генерации потоков данных с заданными вероятностными свойствами позволяет выполнять «обучение с учителем» для формирования моделей классов.

Апробация работы. Основные результаты диссертационной работы докладывались и обсуждались на следующих конференциях и семинарах: Международная научно-практическая конференция «Общественная безопасность, законность и правопорядок в III тысячелетии» (Воронеж, 2016 г.); Международная научно-практическая конференция «Интеллектуальные информационные системы» (2016 г.); Международная научно-практическая конференция «Охрана, безопасность, связь» (Воронеж, 2017 г., 2018 г.); Международная научно-практическая конференция «Измерения: состояние и перспективы развития» (Челябинск 2017 г.), Международная научно-техническая конференция «Актуальные проблемы прикладной математики, информатики и механики» (Воронеж 2020 г.).

Реализация и внедрение результатов работы. Результаты диссертации использованы в опытно-конструкторской работе «Изумруд», выполненной РУНЦ «Безопасность» Московского государственного технического университета им Н.Э. Баумана, внедрены в практическую деятельность ООО «КАСКАД» (г. Москва), а также в образовательный процесс Воронежского института высоких технологий.

Соответствие паспорту специальности. Содержание диссертации

соответствует п. 3 «Разработка критериев и моделей описания и оценки эффективности решения задач системного анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта», п. 4 «Разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта», п. 5 «Разработка специального математического и алгоритмического обеспечения систем анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта» паспорта специальности 2.3.1. Системный анализ, управление и обработка информации, статистика.

Публикации. По материалам диссертации опубликовано 20 печатных работ, в том числе 7 статей опубликовано в изданиях, рекомендованных ВАК при Минобрнауки России, 4 статьи в изданиях, индексируемых в Scopus, разработаны 2 программы для ЭВМ, зарегистрированные в Федеральной службе по интеллектуальной собственности, получен 1 патент на изобретение РФ.

В работах, опубликованных в соавторстве, лично соискателем предложены: [1] – использование комбинационных подходов для формирования двумерной плотности вероятности; [2, 8, 18] – алгоритм формирования отсчетов случайного процесса при заданной функции распределения вероятностей; [3] – оптимальный алгоритм классификации потоков данных; [5, 15, 17] – математическое представление потоков данных для формирования двумерных и трехмерных марковских моделей; [6] – алгоритм формирования эталонных потоков данных; [7] – архитектура системы классификации потоков данных; [9] – процедура интеграции имитатора потока данных в подсистему классификации; [10, 11] – алгоритм формирования решения с требуемой достоверностью по полученным решающим статистикам; в [16] – структурная схема демодулятора; [19] – подходы к выбору статистических характеристик; [20] – структура цифрового имитатора.

Структура и объем работы. Диссертация состоит из введения, четырех глав, заключения, списка литературы из 110 наименований и приложения. Работа изложена на 155 страницах машинописного текста (основной текст занимает 136 страниц, содержит 114 рисунков, 9 таблиц, 1 приложение).

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационного исследования, сформулированы цель и частные задачи, показана научная новизна и практическая значимость полученных результатов, выносимых на защиту.

В первой главе проведен анализ информационных сигналов, рассмотрены их свойства и задачи классификации. Рассмотрены существующие математические модели радиосигналов и ритмограмм. Показано, что марковские модели случайных процессов с дискретным множеством состояний являются достаточно универсальным математическим аппаратом вероятностного моделирования. Применительно к квантованным по времени и уровню сигнала простейшая марковская модель представляет собой дискретную цепь Маркова.

Цепь Маркова моделирует взаимосвязанные дискретные отсчеты z_n сигнала,

принимаящие целочисленные значения i от 1 до M , равные номерам уровней квантования, в момент времени t_n . Для m -связной цепи вероятность значений отсчета z_n зависит от ранее полученных отсчетов z_{n-k} , $k = \overline{1, m}$ и не зависит от более ранних отсчетов. Поскольку исследование случайных информационных сигналов и алгоритмов их обработки предполагает необходимость их имитации, т. е. формирования последовательности отсчетов с заданными вероятностными свойствами, предложены алгоритмы имитации случайных процессов с заданными двумерными и трехмерными плотностями вероятностей на основе их марковских моделей.

В разработанном алгоритме формирования эталонных потоков данных с заданным двумерным распределением вероятностей на основе матрицы $[P_{ij}]$ вычисляется матрица распределений вероятностей перехода с элементами

$$F_{ij} = \sum_{k=1}^j P_{ik}.$$

Алгоритм формирования эталонных потоков данных (рис. 1) заключается в следующем: датчик псевдослучайных чисел v_n с равномерным распределением вероятностей на интервале от 0 до 1 формирует очередное значение, соответствующее текущему n -му отсчету z_n . Если предыдущее значение отсчета $z_{n-1}=i$, то величина $z_n=j$ выбирается как минимальное значение j , при котором выполняется неравенство

$$v_n < F_{ij}.$$

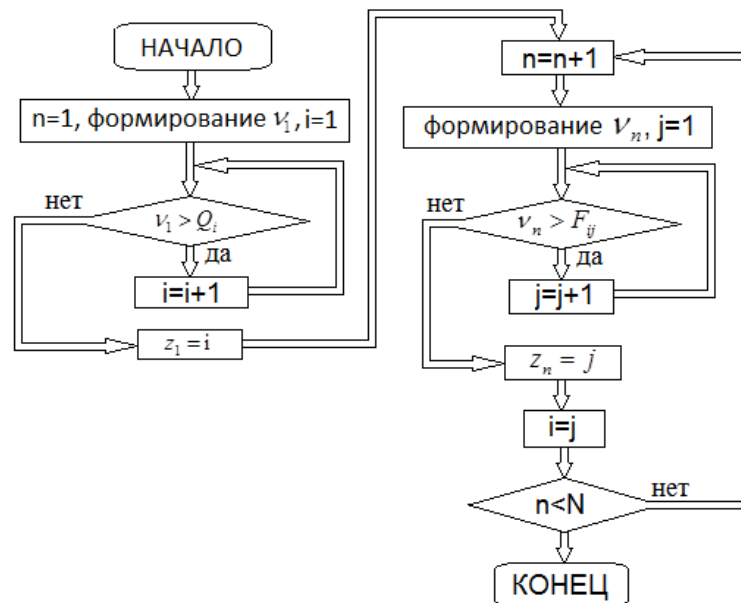


Рисунок 1. Структурная схема алгоритма формирования эталонных потоков

Для двумерного гауссовского распределения вероятностей со средним значением $x_{\text{СР}} = 0$, дисперсией $\sigma^2 = 4$ и коэффициентом корреляции $r = 0,8$ при $M = 32$ на рис. 2 представлены трехмерные диаграммы матрицы переходных вероятностей $[P_{ij}]$ (рис. 2а) и распределений вероятностей перехода $[F_{ij}]$ (рис. 2б).

Результаты моделирования имитатора показаны на рис. 3: реализация отсчетов сигнала (рис. 3а), гистограмма переходных вероятностей (рис. 3б),

гистограмма одномерного распределения вероятностей (рис. 3в, пунктиром отображается теоретическая гистограмма) и зависимость коэффициента корреляции r_k от смещения выборок k (рис. 3г). Как видно, алгоритм обеспечивает достаточно точное соответствие результатов имитации и теоретических статистических характеристик, кроме того, позволяет получить поток данных с произвольным двухмерным распределением вероятностей, определяемым соответствующей марковской моделью.

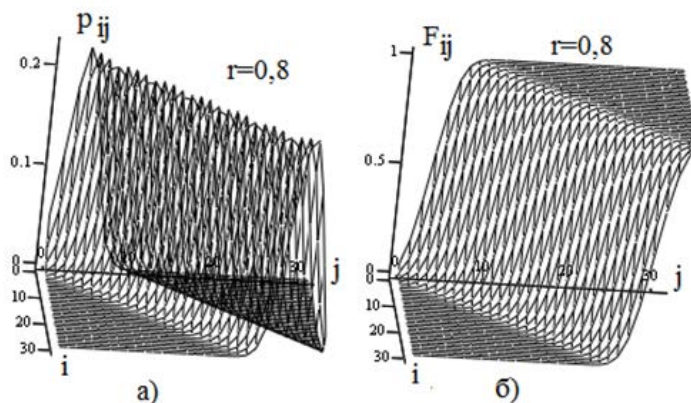


Рисунок 2. Трехмерные диаграммы вероятностных матриц

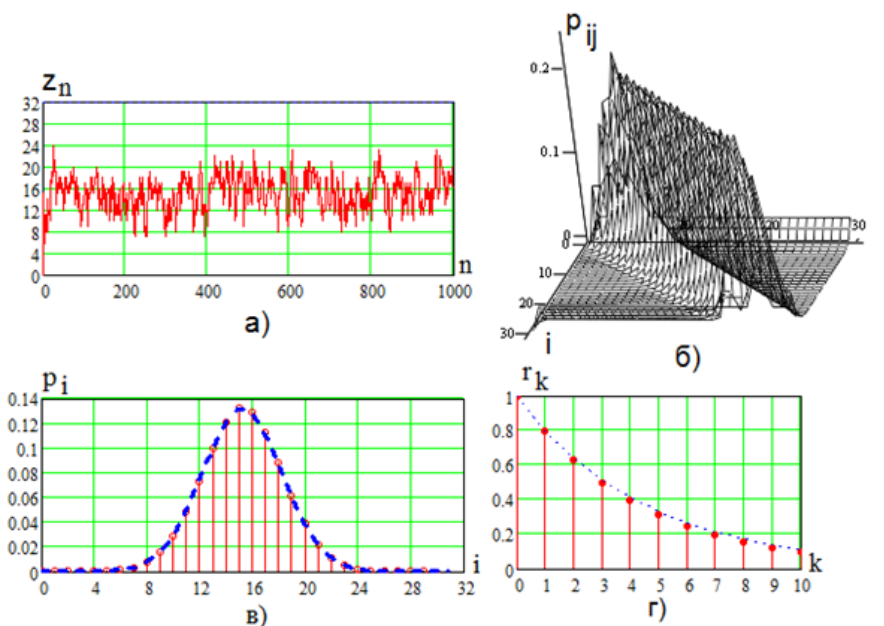


Рисунок 3. Результаты моделирования имитатора

Сформулированы цели и задачи исследования.

Таким образом, на основе марковской модели, отличающейся максимизацией апостериорных вероятностей принимаемых решений и обеспечивающей динамическую идентификацию потоков данных, разработан алгоритм формирования эталонных потоков данных с произвольным двухмерным распределением вероятностей.

Во второй главе предложены одно (1а) и двухсвязная (1б) модели Маркова по заданным плотностям вероятностей и по наблюдаемым экспериментально временным реализациям отсчетов случайного процесса:

$$P_{ij} = \frac{P(z_1=i, z_2=j)}{P(z_1=i)} = \frac{\int_{g_{i-1}}^{g_i} \int_{g_{j-1}}^{g_j} w(x_1 x_2) dx_2 dx_1}{\int_{g_{i-1}}^{g_i} \int_{-\infty}^{\infty} w(x_1 x_2) dx_2 dx_1}, \quad (1a)$$

$$P_{kij} = \frac{P(z_1=k, z_2=i, z_3=j)}{P(z_1=k, z_2=i)} = \frac{\int_{g_{j-1}}^{g_j} \int_{g_{i-1}}^{g_i} \int_{g_{k-1}}^{g_k} w(x_1 x_2 x_3) dx_3 dx_2 dx_1}{\int_{g_{j-1}}^{g_j} \int_{g_{i-1}}^{g_i} \int_{-\infty}^{\infty} w(x_1 x_2 x_3) dx_3 dx_2 dx_1}, \quad (16)$$

где g_i – уровни квантования, $w(x_1 x_2)$ и $w(x_1 x_2 x_3)$ – двухмерная и трехмерная плотности вероятностей.

При формировании марковской модели по результатам статистического имитационного моделирования формируется матрица суммарных чисел перехода l_{ij} значений квантованной последовательности отсчетов z_n от $z_{n-1} = i$ к $z_n = j$ для всех $n = \overline{2, N}$. Структурная схема алгоритма приведена на рис. 4а. На рис. 4б показан пример матрицы чисел перехода при $M = 64$, а ее графическое представление в виде трехмерной диаграммы – на рисунке 4в.

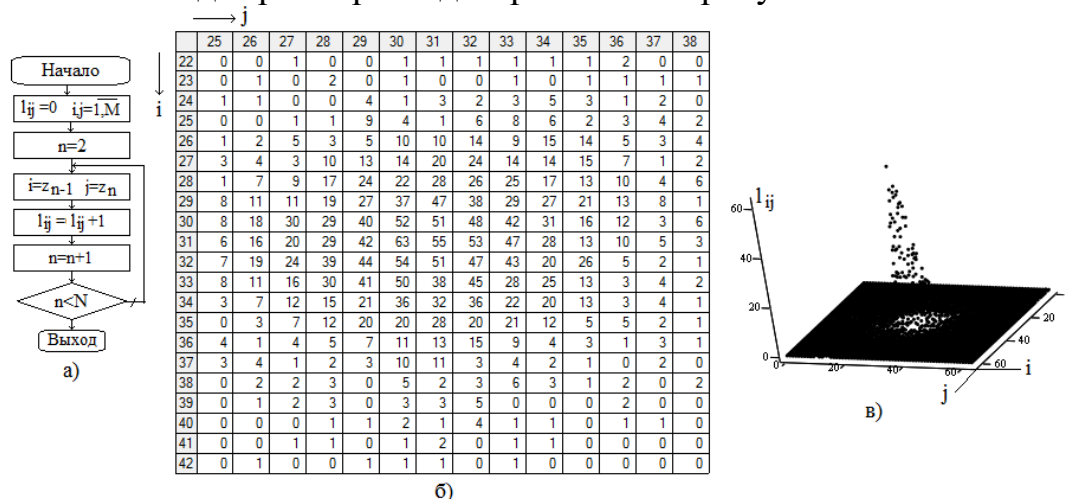


Рисунок 4. Формирование чисел перехода

В главе приводятся результаты статистического имитационного моделирования алгоритмов имитации случайных сигналов по их марковским моделям. Приведены примеры формирования марковских моделей радиосигналов и ритмограмм. Разработана методика формирования моделей классов для простой и двухсвязной цепи Маркова.

Таким образом, на основе трехмерной марковской модели разработан алгоритм формирования эталонных потоков данных с заданными вероятностными свойствами.

Третья глава посвящена разработке алгоритмов классификации сигналов на основе их марковских моделей. Рассмотрены апостериорные вероятности принадлежности полученной выборки отсчетов сигнала соответствующему классу G_m (m – его номер), получено выражение

$$P(G_m / z_1, z_2, \dots, z_N) = \frac{p_{z_1}^{(k)} \cdot P(G_m) \cdot 2^{-L_m(z_1, z_2, \dots, z_N)}}{\sum_{n=1}^K p_{z_1}^{(n)} \cdot P(G_n) \cdot 2^{-L_n(z_1, z_2, \dots, z_N)}}, \quad (2)$$

где решающая статистика для односвязной цепи Маркова имеет вид

$$L_m = L_m(z_1, z_2, \dots, z_N) = - \sum_{i=1}^M \sum_{j=1}^M l_{ij} \cdot \log_2 P_{ij}^{(m)}, \quad (3)$$

а для двухсвязной марковской модели

$$L_m = L_m(z_1, z_2, \dots, z_N) = - \sum_{k=1}^M \sum_{i=1}^M \sum_{j=1}^M l_{kij} \cdot \log P_{kij}^{(m)}, \quad (4)$$

где $P_{***}^{(m)}$ – марковская модель класса, а l_{***} – матрица чисел перехода.

Решающие статистики (3) и (4) представляют собой «расстояние» от принимаемой выборки до соответствующего класса, задавая метрику в пространстве классов.

Оптимальный алгоритм классификации случайного сигнала строится по критерию максимальной апостериорной вероятности и может быть записан в виде

$$L_{k1}(z_1, z_2, \dots, z_N) - L_{k0}(z_1, z_2, \dots, z_N) \geq D, \quad (5)$$

где порог принятия решения равен

$$D = \log_2 \left[\frac{\sum_{k=1, k \neq k_0}^K p_{z_1}^{(k)} \cdot P(G_k)}{p_{z_1}^{(k_0)} \cdot P(G_{k_0})} \frac{P_0}{1 - P_0} \right], \quad (6)$$

P_0 – доверительная вероятность правильного решения, L_{k0} – минимальное значение решающей статистики, L_{k1} – следующее за ним значение.

Разработанный алгоритм классификации потоков данных заключается в следующем: по поступившей выборке z_1, z_2, \dots, z_N из N отсчетов определяются числа l_{ij} переходов отсчетов от значения $z_{n-1} = i$ к $z_n = j$, $n = \overline{2, N}$, которые накапливаются от начальных значений $l_{ij} = 0$ по мере появления новых отсчетов. Для двухсвязной модели аналогично определяются числа l_{kij} перехода от $z_{n-2} = k$ к $z_{n-1} = j$ и затем к $z_n = j$, $n = \overline{3, N}$. Затем определяются решающие статистики $L_k(z_1, z_2, \dots, z_N)$ вида (3) или (4) для каждого класса G_k , $k = \overline{1, K}$, выбирается минимальное значение L_{k0} и минимальное из оставшихся значений L_{k1} и их разность сравнивается с порогом (6). Если порог превышает, то выносятся окончательное решение о принадлежности выборки к классу G_{k_0} с достоверностью не меньше P_0 . Если неравенство (5) не выполняется, то это означает, что выбран недостаточный объем выборки N и его необходимо увеличить.

Для односвязной марковской модели среднее значение решающей статистики L_k пропорционально объему выборки N ,

$$L_k = (N - 1) \cdot H(Q, P^{(k)}), \quad (7)$$

$$H(Q, P^{(k)}) = H_2(Q) - H_1(Q) + D_2(Q, P^{(k)}) - D_1(Q, P^{(k)}), \quad (8)$$

двумерная и одномерная энтропии принимаемого сигнала равны

$$H_2(Q) = - \sum_{i=1}^M \sum_{j=1}^M Q(i, j) \cdot \log_2 [Q(i, j)], \quad (9)$$

$$H_1(Q) = - \sum_{i=1}^M Q(i) \cdot \log_2 [Q(i)]. \quad (10)$$

Двумерная и одномерная информационные дивергенции соответственно

$$D_2(Q, P^{(k)}) = -\sum_{i=1}^M \sum_{j=1}^M Q(i, j) \cdot \log_2 \left[\frac{Q(i, j)}{P^{(k)}(i, j)} \right], \quad (11)$$

$$D_1(Q, P^{(k)}) = -\sum_{i=1}^M Q(i) \cdot \log_2 \left[\frac{Q(i)}{P^{(k)}(i)} \right], \quad (12)$$

$Q(i, j)$ – совместная вероятность появления двух соседних отсчетов $z_{n-1} = i, z_n = j$ в принимаемой выборке, $Q(i)$ – вероятность появления отсчета $z_n = i$.

Разность информационных дивергенций определяет «расстояние» между принимаемым сигналом и моделью соответствующего класса. Аналогичные результаты получены и для двухсвязных марковских моделей. Таким образом, теоретико-информационные свойства моделей классов и принимаемого сигнала могут использоваться для анализа процедуры классификации.

В главе также рассмотрены свойства решающих статистик простой марковской модели, разработана метрика классов, получены их статистические характеристики. Показано, что решающие статистики имеют нормальное распределение вероятностей. Получены оценки их среднего значения и дисперсии. Проведен анализ статистических характеристик решающих статистик двухсвязной марковской модели, получены оценки их среднего значения и дисперсии.

Таким образом, разработан алгоритм классификации потоков данных на основе решающих статистик, основанный на формировании решающих статистик, получаемых с помощью марковских моделей потоков данных, обеспечивающий максимум отношения правдоподобия.

В четвертой главе предложена и реализована в виде комплекса программ архитектура системы классификации потоков данных (рис. 5).

Информационные сигналы поступают от исследуемого объекта, формируются и анализируются их марковские модели.

Модели классов формируются в процессе машинного обучения алгоритма классификации за счёт интеграции модуля генерации потоков данных с заданными вероятностными свойствами в подсистему обработки потоков данных. Для этого используются либо теоретические сведения о многомерных плотностях вероятностей анализируемых случайных процессов, либо экспериментальные реализации (выборки) отсчетов анализируемых процессов с известной принадлежностью выбранному классу (обучение с учителем).

Для односвязной марковской модели в режиме обучения по полученной классифицированной выборке отсчетов случайного процесса формируется матрица суммарных чисел перехода l_{ij} значений квантованной последовательности отсчетов z_n от $z_{n-1} = i$ к $z_n = j$ для всех $n = \overline{2, N}$, на основе которых определяется матрица переходных вероятностей класса (его модель) в виде

$$\tilde{P}_{ij} = \frac{l_{ij}}{\sum_{k=1}^M l_{ik}}. \quad (13)$$

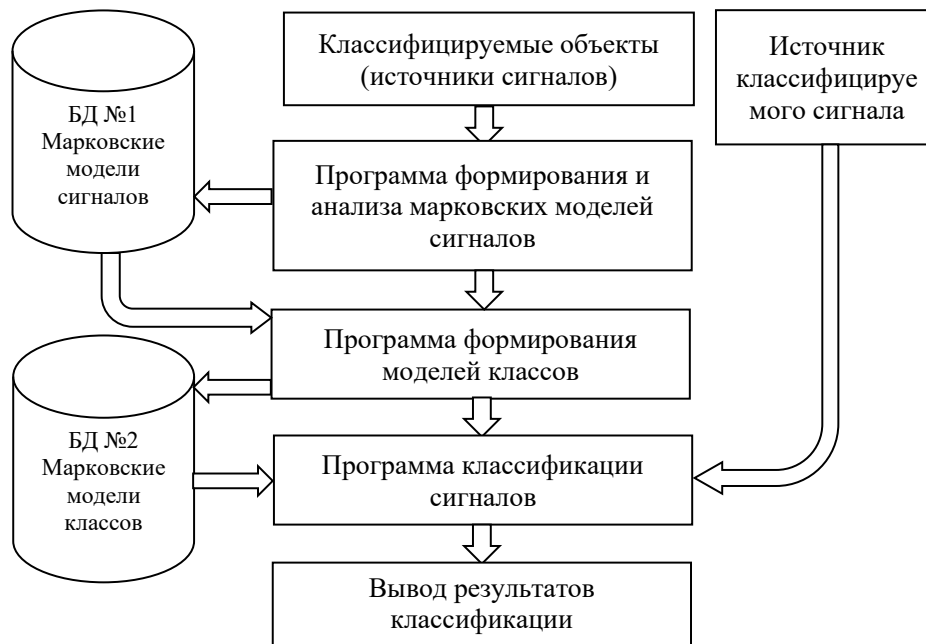


Рисунок 5. Обобщенная структурная схема программного обеспечения

Аналогичные оценки используются и для трехмерной матрицы двухсвязной цепи Маркова.

Программное обеспечение процедуры классификации случайных процессов обеспечивает выполнение следующих операций:

- формирование матриц чисел перехода значений случайного процесса;
- анализ решающих статистик для выбранных моделей классов;
- классификацию случайных процессов.

Программа формирования чисел перехода позволяет для указанного ей файла, содержащего классифицированную реализацию отсчетов сигнала, построить матрицу чисел перехода l_{ij} значений отсчетов z_n от $z_{n-1} = i$ к $z_n = j$ для всех $n = \overline{2, N}$. Эта матрица удобна для накопления информации о свойствах модели и построения ее усредненных вариантов, из нее определяется матрица переходных вероятностей.

Программа анализа решающих статистик позволяет сравнивать их значения для файла, содержащего анализируемую выборку отсчетов, со всеми моделями в выбранном классе.

Программа классификации случайных процессов предназначена для решения задачи классификации реализации сигнала в выбранном входном файле.

На рис. 6 показан пример работы программы классификации радиосигналов по трем классам при обработке выборки отсчетов сигнала с цифровой частотной манипуляцией (GMSK).

Рассмотренное программное обеспечение для анализа и классификации сигналов, позволяет проводить их анализ и классификацию на основе односвязных марковских моделей, обеспечивая их оптимальную обработку и формирование решений с заданной достоверностью.

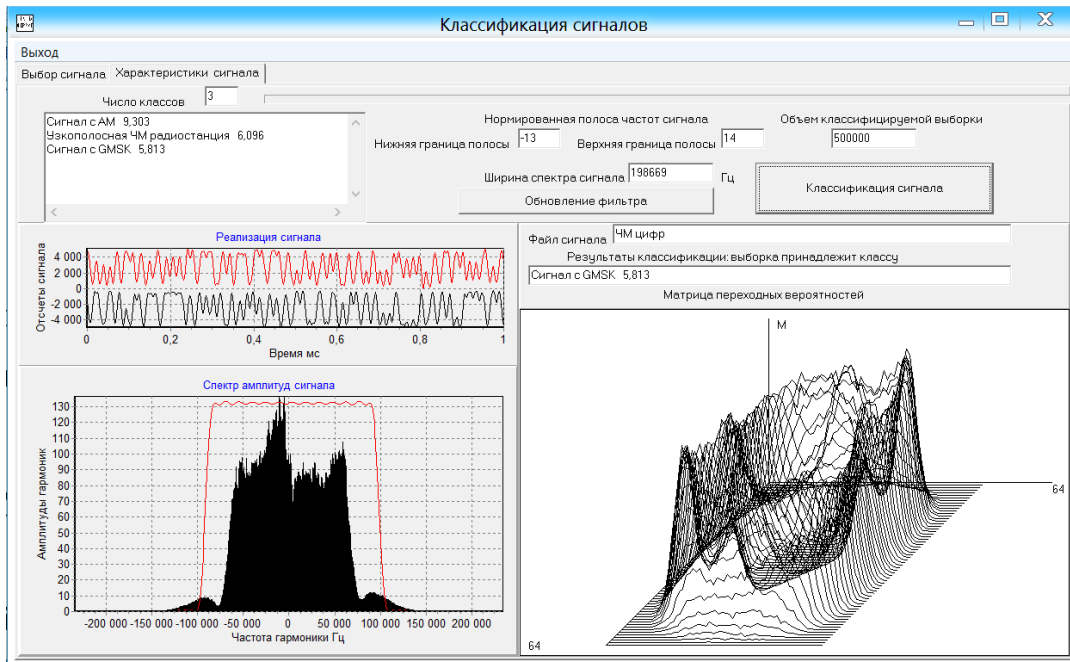


Рисунок 6. Результаты обработки сигнала с GMSK

Таким образом, показана возможность применения предложенных алгоритмов в системах машинного обучения, разработана архитектура системы классификации потоков данных, отличающаяся интеграцией модуля генерации потоков данных с заданными вероятностными свойствами в подсистему обработки потоков данных и обеспечивающая классификацию сигналов по поступающей выборке отсчетов с заданной достоверностью.

В заключении подведены итоги диссертации, сформулированы основные результаты и выводы.

ЗАКЛЮЧЕНИЕ

На основании теоретических исследований и статистического имитационного моделирования сделаны следующие выводы.

1. Для классификации случайных сигналов предложено использовать их односвязную или двухсвязную марковскую модель, позволяющую реализовать универсальный оптимальный алгоритм классификации. Предложены методы формирования марковских моделей случайных процессов на основе теоретических расчетов и по результатам экспериментов, позволяющие формировать модели классов. Приведены примеры их формирования в медицине и радиотехнике.

2. Разработаны алгоритмы формирования эталонных потоков данных с заданными статистическими характеристиками на основе простой и двухсвязной марковских моделей. На примерах нормального случайного процесса показана возможность учета при моделировании «тонкой структуры» статистических связей. Рассмотрена возможность построения имитатора по экспериментальной реализации сигнала.

3. Предложены методы формирования марковских моделей классов сигналов, предложении метрика оценки «расстояния» между классами с помощью

решающих статистик. Предложен метод формирования модели класса с указанием его «центра» и «границ».

4. Разработан алгоритм классификации случайных процессов на основе односвязной цепи Маркова, получены выражения для решающих статистик и решающего правила. Предложен алгоритм классификации случайных процессов на основе их двухсвязной марковской модели, определены решающие статистики и решающее правило, обеспечивающее максимум отношения правдоподобия. Получены выражения для порога сравнения решающих статистик при заданной достоверности результата классификации.

5. Разработано программное обеспечение алгоритмов классификации случайных процессов на основе марковских моделей. Программа формирования матриц чисел перехода позволяет формировать марковские модели реализации радиосигнала и проводить анализ его частотных и статистических свойств, полученные матрицы записываются в файл модели. Программа анализа решающих статистик дает возможность сравнивать решающие статистики обрабатываемой реализации сигнала для различных классов с целью формирования центральной и граничных моделей каждого класса. Программа классификации информационных сигналов предназначена для определения принадлежности обрабатываемой выборки отсчетов сигнала одному из заданных классов с оценкой соответствующих значений нормированных решающих статистик. Проверка работоспособности программного обеспечения проведена на примере классификации радиосигналов.

6. Проведен анализ возможностей использования предложенных алгоритмов для классификации состояний сердечно-сосудистой системы пациента по реализациям дифференциальной ритмограммы. Показана необходимость использования часовых записей в разное время суток и необходимость накопления обучающих реализаций от большого числа пациентов.

Дальнейшие разработки по данному направлению могут быть связаны с повышением точности классификации информационных сигналов, разработки методов и алгоритмов теории распознавания образов, в том числе на основе обучаемых нейронных сетей.

ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ

Статьи в изданиях, рекомендованных ВАК

1. Аналитическая двумерная вероятностная модель радиосигнала / Д.А. Глушков, М.Ю. Калинин, Ю.В. Литвиненко // Вестник Воронежского государственного технического университета. – 2023. – Т. 19. – № 1. – С. 57-61.

2. Возможности использования марковской модели для статистического описания дифференциальной ритмограммы / А.Н. Глушков, М.Ю. Калинин, Ю.В. Литвиненко // Вестник Поволжского государственного технологического университета. Серия: Радиотехнические и инфокоммуникационные системы. – 2019. – № 3 (43). – С. 22-28.

3. Классификация случайных сигналов на основе их двухсвязных марковских моделей / М.Ю. Калинин, О.Н. Чопоров, А.М. Бонч-Бруевич // Моделирование, оптимизация и информационные технологии. 2022; 10(3).

Доступно по: <https://moitvivi.ru/ru/journal/pdf?id=1222> DOI: 10.26102/2310-6018/2022.38.3.017.

4. Калинин М.Ю. Цифровой имитатор случайного процесса на основе его многомерной марковской модели // Системы управления и информационные технологии. – 2022. – № 3 (89). – С. 10-14.

5. Энтропийные оценки решающих статистик алгоритма классификации случайных процессов / М.Ю. Калинин, О.Н. Чопоров // Моделирование, оптимизация и информационные технологии. 2020; 8(4). Доступно по: <https://moitvivi.ru/ru/journal/pdf?id=881> DOI: 10.26102/2310-6018/2020.31.4.034

6. Цифровая имитация случайных сигналов с заданными статистическими свойствами / А.Н. Глушков, М.Ю. Калинин // Вестник Воронежского института МВД России. – 2018. – № 4. – С. 43-50.

7. Калинин М.Ю. Управление процессом и разработка системы принятия решений классификации информационных сигналов на основе марковских моделей, Осама Адил Рахим, М.Ю. Калинин, Д.И. Мутин // Моделирование, оптимизация и информационные технологии. 2024; 12(2). Доступно по: <https://moitvivi.ru/ru/journal/pdf?id=1543> DOI: 10.26102/2310-6018/2024.45.2.001

Публикации в изданиях, индексируемых в 1 квартале базы РИНЦ, базах «Scopus» и «Web of Science»

8. Digital simulator of a random process using Markov model. / A.N. Glushkov, M.Y. Kalinin, V.P. Litvinenko, Y.V. Litvinenko // Applied Mathematics, Computational Science and Mechanics: Current Problems. IOP Conf. Series: Journal of Physics: Conf. Series 1479 (2020). – P. 012056. DOI: 10.1088/1742-6596/1479/1/012056

9. Gaussian signals simulation using biconnected markov chain / A.N. Glushkov, V.V. Menshikh, N.S. Khohlov, O.I. Bokova, M.Y. Kalinin // 2nd International Ural Conference on Measurements (URALCON) Proceedings. – 2017. – P. 245-250. DOI: 10.1109/URALCON.2017.8120718

10. Markov model based mathematical representation of radio signals / A.N. Glushkov, V.V. Menshikh, N.S. Khohlov, O.I. Bokova, M.Y. Kalinin // 2nd International Ural Conference on Measurements (URALCON) Proceedings. – 2017. – P. 239-244. DOI: 10.1109/URALCON.2017.8120717

11. Random Signal Digital Simulator Based on Its Double Chain Markov Model. / O.V. Chernoyarov, A.N. Glushkov, M.Y. Kalinin, Kaung Myat San, V.P. Litvinenko // IAENG International Journal of Computer Science. – Volume 48. – Issue 4: December 2021. – P. IJCS_48_4_29.

Свидетельства о регистрации программы для ЭВМ

12. Калинин М.Ю. Программа классификации информационных сигналов : свидетельство о государственной регистрации программы для ЭВМ от 23.07.2018 № 2018618911 / М.Ю. Калинин. – Москва : ФИПС, 2018.

13. Калинин М.Ю. Программа формирования матриц переходных вероятностей : свидетельство о государственной регистрации программы для ЭВМ от 02.8.2018 № 2018619251 / М.Ю. Калинин. – Москва : ФИПС, 2018.

Патент на изобретение РФ:

14. Калинин М.Ю. Цифровой имитатор случайных сигналов : патент на изобретение от 05.06.2019 № RU 2690780 C1 / А.Н. Глушков, М.Ю. Калинин, В.П. Литвиненко, Ю.В. Литвиненко. – Москва : ФИПС, 2019.

Статьи и материалы конференций

15. Аспекты применения марковских моделей информационных сигналов при разработке устройств имитации информационных сигналов / А.Н. Глушков, М.Ю. Калинин // В сборнике: Охрана, безопасность, связь. – Воронеж : ВИ МВД России. – 2017. – № 1-1. – С. 272-276.

16. Исследование цифрового алгоритма когерентной демодуляции двоичных фазоманипулированных сигналов / М.Ю. Калинин, А.Н. Глушков // В сборнике: Общественная безопасность, законность и правопорядок в III тысячелетии. – Воронеж : ВИ МВД России. – 2016. – № 1-2. – С. 132-137.

17. Калинин М.Ю. Обучение алгоритма классификации случайного процесса / М.Ю. Калинин, О.Н. Чопоров // Вестник Воронежского института высоких технологий. – 2021. – № 1 (36). – С. 19-22.

18. Марковская модель ритмограммы / А.Н. Глушков, М.Ю. Калинин, Ю.В. Литвиненко // В сборнике: Интеллектуальные информационные системы. – Воронеж : ВГТУ. – 2016. – С. 127-131.

19. Статистические характеристики дифференциальной ритмограммы / А.Н. Глушков, М.Ю. Калинин, Ю.В. Литвиненко // В сборнике: Интеллектуальные информационные системы. – Воронеж : ВГТУ. – 2016. – С. 194-198.

20. Цифровой имитатор случайного процесса на основе марковской модели / А.Н. Глушков, М.Ю. Калинин, В.П. Литвиненко, Ю.В. Литвиненко // В сборнике: Актуальные проблемы прикладной математики, информатики и механики. Сборник трудов Международной научно-технической конференции. – Воронеж : ВГУ. – 2020. – С. 760-767.

Подписано в печать: 25.03.2025
Заказ № 20605. Тираж 100 экз.
Бумага офсетная. Формат 60x90/16.
Типография «Автореферат.ру»
ОГРНИП 320774600073831
119313, Москва, ул. Марии Ульяновой, д.3, к.1
(977) 518-13-77 (499) 788-78-56
www.avtoreferat.ru riso@mail.ru

