На правах рукописи

Мельников Сергей Юрьевич

МЕТОДЫ РАСПОЗНАВАНИЯ И ИДЕНТИФИКАЦИИ КОНЕЧНЫХ АВТОМАТОВ ПО СТАТИСТИЧЕСКИМ ХАРАКТЕРИСТИКАМ ВЫХОДНЫХ И ВХОДНЫХ ПОСЛЕДОВАТЕЛЬНОСТЕЙ

Специальность: 05.13.17 – Теоретические основы информатики

АВТОРЕФЕРАТ

диссертации на соискание ученой степени
доктора физико-математических наук

Москва 2021

Работа выполнена на кафедре прикладной информатики и теории вероятностей Федерального государственного автономного образовательного учреждения высшего образования «Российский университет дружбы народов»

Научный консультант Доктор технических наук, профессор, заведующий кафедрой прикладной информатики и теории вероятностей Федерального государственного автономного образовательного учреждения высшего образования «Российский университет дружбы народов» **Самуйлов Константин Евгеньевич**

Официальные оппоненты:

Крук Евгений Аврамович,
доктор технических наук, профессор, Московский институт электроники и математики имени А. Н. Тихонова Национального исследовательского университета «Высшая школа экономики», и.о. директора, научный руководитель

Тимонина Елена Евгеньевна,
доктор технических наук, профессор, Федеральный исследовательский центр «Информатика и управление» Российской академии наук, ведущий научный сотрудник

Пудовкина Марина Александровна,
доктор физико-математических наук, Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский ядерный университет «МИФИ», профессор кафедры криптологии и дискретной математики

Орлов Юрий Николаевич,
доктор физико-математических наук, доцент, Федеральное государственное учреждение «Федеральный исследовательский центр Институт прикладной математики им. М.В. Келдыша Российской академии наук», заведующий отделом №6

Защита диссертации состоится «03» декабря 2021 г. в 15 час. 00 мин. на заседании диссертационного совета ПДС 0200.001 на базе Российского университета дружбы народов (117198, г. Москва, ул. Миклухо-Маклая, д.6).

С диссертацией можно ознакомиться в научной библиотеке Российского университета дружбы народов по адресу: 117198, Москва, ул. Миклухо-Маклая, дом 6 (отзывы на автореферат просьба направлять по указанному адресу) или на официальном сайте диссертационных советов РУДН по адресу: http://dissovet.rudn.ru

Автореферат разослан «____» октября 2021 г.

Ученый секретарь диссертационного совета
ПДС 0200.001, доцент А.В. Демидова

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность темы исследования. Интенсивно развивающиеся сети связи являются неотъемлемой частью жизни информационного общества. Современные сети используют огромное число разновидностей оборудования приема, передачи и обработки информации. При разработке и анализе такого оборудования необходимо решение целого ряда идентификационных задач. Общепринятой моделью дискретных устройств являются конечные автоматы. Диссертационная работа посвящена проблеме идентификации конечных автоматов, которые обрабатывают случайные или псевдослучайные данные, по наблюдениям над выходной и входной последовательностями.

Проблема идентификации конечных автоматов относится к классической задаче математической кибернетики о «черном ящике». Она включает в себя вопросы распознавания, восстановления, контроля и диагностики автоматов.

В диссертации под распознаванием автоматов понимается указание таких автоматов из заданного класса, которые соответствуют наблюдаемым свойствам выходной и входной последовательностей, под идентификацией автомата понимается проверка гипотезы о том, что последовательности, свойства которых наблюдаются, были получены с помощью эталонного автомата.

Перечислим области, в которых исследуемая проблема является актуальной.

В области технической диагностики дискретных устройств важное место занимает вероятностное тестирование. На вход тестируемого и исправного устройств подаются случайные данные с заданными вероятностными характеристиками, а обнаружение неисправности достигается путем сравнения статистических характеристик выходных последовательностей. Решение задачи распознавания автоматов позволит указать конкретный тип неисправности устройства.

В ряде криптоаналитических приложений необходимо проверять гипотезы о том, что анализируемый узел реализует тот или иной алгоритм шифрования, ключ которого неизвестен. Задачи идентификации и распознавания возникают, если к самому узлу доступа у исследователя нет, а для анализа может быть доступна полная или частичная информация о выходной и, в ряде случаев, о входной последовательностях.

При разработке генераторов случайных последовательностей (ГСП) необходимо обеспечить неотличимость получаемой последовательности от случайной. Ситуация, при которой по выходной последовательности возможна идентификация генератора или его отдельных узлов, свидетельствует о плохом качестве ГСП.

При поиске скрытых каналов в информационных системах, когда требуется узнать, не произведено ли несанкционированное изменение алгоритмов их функционирования, необходима идентификация систем и их отдельных узлов по статистическим свойствам входных и выходных потоков данных. Эта задача соответствует п.2.1.3 «Основных направлений научных исследований в области обеспечения информационной безопасности Российской Федерации», утвержденным в 2017 г.

Разработка новых методов идентификации автоматов актуальна также для задач обработки материалов на естественных языках. Аппарат детерминированных и вероятностных конечных автоматов в последние годы активно используется для моделирования естественных языков, генерации искусственных текстов, автоматического перевода, коррекции ошибок в текстах, построения моделей интерактивного общения «человек-компьютер». Потребность в эффективно работающих системах, решающих идентификационные задачи по наблюдениям над фрагментами текста или речевого сигнала, в последние годы сильно выросла и продолжает расти, что связано с глобализацией систем и сетей связи и развитием технологий контроля трафика. Одной из практически важных задач здесь является задача распознавания языка текстовых и речевых сообщений. Актуальность теоретико-автоматного похода к задаче

идентификации языка возрастает с развитием многоязычных генераторов искусственных текстов, которые моделируются конечными автоматами.

При идентификации автомата важно определить, с какой точностью имеющиеся данные о выходных и входных последовательностях позволяют идентифицировать автомат. При диагностировании устройств это позволит ограничить круг возможных неисправностей, в криптоаналитических приложениях это необходимо для анализа стойкости криптографических преобразований.

Говоря о точности идентификации автомата в содержательном смысле, необходимо ввести ограничения на класс рассматриваемых автоматов. Такие ограничения могут касаться, например, числа состояний автомата, множеств алфавитов, вида функции переходов или выходов. Частным случаем применения таких ограничений является класс автоматов с фиксированными алфавитами, множеством состояний и функцией переходов. Важнейшими классами автоматов, часто используемыми в ГСП, являются регистры сдвига и их модификации. Функции выходов таких автоматов предназначены для усложнения последовательности состояний. В этом случае автомат может быть идентифицирован с точностью до некоторого набора функций выходов, что приводит к той или иной классификации множества дискретных функций.

Для решения перечисленных задач используется два подхода: детерминированный и вероятностный. Детерминированный подход связан с поочередной установкой исследуемых устройств и систем во все возможные начальные состояния и последующей проверкой свойств генерируемых последовательностей. Вероятностный подход использует вероятностное моделирование входных последовательностей и расчет вероятностных характеристик выходных последовательностей.

С ростом сложности анализируемых устройств и систем оба подхода сталкиваются с серьезными трудностями. Детерминированный подход подразумевает перебор состояний автомата и ограничен доступными вычислительными ресурсами. Так, в криптоаналитических приложениях в качестве ключей шифрования, как правило, используются начальные состояния, число которых выбирается разработчиками из соображений невозможности перебора на современных и перспективных средствах вычислительной техники. При вероятностном подходе, как правило, используются базовые модели зависимости между членами анализируемых последовательностей - независимые испытания и цепи Маркова. В ряде случаев они оказываются слишком грубыми для описания свойств реальных последовательностей.

Степень разработанности темы. На сегодняшний день проблеме идентификации конечных автоматов посвящено значительное количество работ, среди которых можно выделить два направления в зависимости от использования модели случайности входной последовательности.

Направление, не предполагающее случайности входной последовательности, является исторически первым и принадлежит интенсивно развивающейся теории экспериментов с автоматами. В рамках этого направления предполагается, что известны отрезки входной и выходной последовательностей анализируемого автомата, и в ряде случаев возможно управление входной последовательностью. Теоретические и прикладные основы исследований в этом направлении базируются, в основном, на результатах в области перечислительных методов дискретной математики, теории графов, теории дискретных функций, теории конечных автоматов. В числе отечественных исследователей следует назвать Богомолова А.М., Грунского И.С., Кудрявцева В.Б., Никонова В.Г., Погорелова Б.А., Пудовкину М.А, Сачкова В.Н., Сумарокова С.Н., Трахтенброта Б.А. и др., а наиболее значимыми зарубежными авторами являются Cerny J., Gill A., Ginsburg S., Moore E.F., Rabin M.O. и др.

Второе направление, связанное с заданием той или иной вероятностной меры на пространстве входных последовательностей, развивается в исследованиях автоматов со случайным входом и теории вероятностных автоматов. В рамках этого направления, как

правило, предполагается, что вероятностная мера на пространстве входных последовательностей задана, и известны значения вероятностей набора мультиграмм в выходной последовательности анализируемого автомата. В исследованиях в этом направлении возникают разновидности отношений статистической эквивалентности между автоматами. В частности, в предположении независимости членов входной последовательности, статистическая эквивалентность может рассматриваться как при произвольном полиномиальном распределении на входном алфавите, так и при заданном, обычно равновероятном. Кроме этого, в ряде работ ограничивается размер мультиграмм.

Теоретические и прикладные основы исследований в этом направлении базируются, в основном, на результатах в области вероятностных методов дискретной математики, теории кодирования, теории вероятностей, теории случайных процессов. В числе отечественных исследователей следует назвать Балакина Г.В., Барашко А.С., Башарина Г.П., Бухараева Р.Г., Грушо А.А., Крука Е.А., Орлова Ю.Н., Рожкова М.И., Самуйлова К.Е., Севастьянова Б.А., Тимонину Е.Е. и др., а наиболее значимыми зарубежными авторами являются Hadjicostis C.N., Marsaglia G., Paz A., Rabin M.O., Vadhan S.P. и др.

Методы идентификации автоматов, разрабатываемые в обоих указанных направлениях, обладают определенными ограничениями. Они могут требовать подачи на вход анализируемого автомата специальным образом подобранных детерминированных или случайных последовательностей с заданной вероятностной мерой и/или установки его в известное начальное состояние, что в ряде приложений невозможно.

Таким образом, необходимы методы идентификации конечных автоматов, которые:

- основаны на наблюдениях над входными и выходными последовательностями,
- не используют информацию о начальном состоянии анализируемого автомата,
- не требуют подачи на вход автомата специально подобранных последовательностей.

Настоящая диссертационная работа посвящена созданию именно таких методов, что и обеспечивает ее актуальность.

Цель и задачи исследования. Диссертация посвящена решению фундаментальной научной проблемы – созданию теоретических основ, моделей и методов распознавания и идентификации конечных автоматов по входным и выходным данным, без использования информации о начальном состоянии автомата.

Для достижения цели в диссертационной работе решены следующие основные **задачи**.

1. Разработка методов распознавания автомата по отрезку выходной последовательности, не требующих установки автомата в фиксированное состояние, с использованием вероятностных моделей входной последовательности, и получение оценок точности распознавания.
2. Исследование совместных статистических свойств входной и выходной последовательностей автомата и построение способа их геометрического описания.
3. Разработка методов идентификации автомата по отрезкам входной и выходной последовательностей, не требующих установки автомата в фиксированное состояние, без использования вероятностных моделей входной последовательности, и получение оценок неоднозначности идентификации.
4. Создание новых принципов классификации дискретных функций, основанных на анализе статистических свойств автоматов с данными функциями выходов.
5. Исследование точности методов распознавания языка дискретных последовательностей естественного происхождения, таких, как речь, текст, искаженный текст, в условиях случайных искажений.
6. Экспериментальное подтверждение разработанных методов идентификации.

Предмет исследования - методы идентификации и распознавания конечных автоматов.

Методы исследования. Для решения поставленных задач в диссертационной работе использовались методы теории автоматов, теории графов, комбинаторного анализа, теории вероятностей и математической статистики, теории дискретных функций. Вероятностные распределения на состояниях автоматов исследовались с помощью элементов теории случайных процессов. Анализ распределений биграмм в выходных последовательностях автоматов проводился с привлечением положений теории спектров графов. При описании свойств автоматных преобразований бесконечных последовательностей применялись методы функционального анализа. Теоретические результаты по идентификации автоматов подтверждались компьютерными экспериментами. Для оценки точности распознавания языков искаженных текстов использовался аппарат теории информации.

Научная новизна

1. Предложен метод распознавания функции выходов автомата со случайным полиномиальным входом с точностью до класса статистической эквивалентности. Отличие метода от известных состоит в использовании целочисленной минимизации модуля линейной формы для определения класса эквивалентности.
2. Предложен метод распознавания функции выходов вероятностного автомата по значениям вероятностей биграмм в его выходной последовательности. Метод отличается сведением задачи распознавания к решению системы квадратичных уравнений, которая в случае автоматов с симметрической матрицей переходов допускает линеаризацию.
3. Предложен метод вычисления спектров обобщенных графов де Брейна, неориентированных степеней графов де Брейна и рангов квадратичных форм, соответствующих вероятностям выходных биграмм регистра сдвига со случайным входом. Метод отличается от известных использованием преобразования унитарного подобия матрицы смежности графа автомата.
4. Разработан метод идентификации автоматов с использованием многогранников, описывающих геометрические области возможных совместных статистических характеристик входной и выходной последовательностей. Метод идентификации с использованием многогранников существенно обобщает известный метод идентификации автоматов по запретам в выходной последовательности.
5. Разработаны способы решения задачи эффективного построения многогранников автоматов для нескольких классов автоматов, основанные на известных комбинаторных алгоритмах, которые возникают при идентификации автоматов по их многогранникам.
6. Предложен новый принцип классификации булевых функций, отличающийся от известных использованием многоугольников автоматов, в которых эти функции являются выходными. Выделен класс булевых функций, которые в схеме регистра сдвига обладают уникальными свойствами, гарантирующими наследование значковых статистических свойств входной последовательности.
7. Предложена модель случайных искажений для оценивания точности методов идентификации языка искаженного текста и проведен ее теоретико-информационный анализ.

Положения, выносимые на защиту

1. Анализ структуры отношения статистической эквивалентности, которое возникает при равенстве пределов относительных частот встречаемости фиксированного слова, позволяющий строить методы распознавания функций выходов автоматов со случайным входом по статистическим свойствам выходной последовательности.
2. Анализ спектральных свойств неориентированных степеней графа переходов автомата, позволяющий разрабатывать методы распознавания функции

выходов конечного вероятностного автомата по значениям вероятностей биграмм в его выходной последовательности.

3. Способы получения спектральных характеристик обобщенных графов де Брейна и неориентированных степеней q-ичных графов де Брейна, позволяющие вычислять значения ряда их теоретико-графовых констант.

4. Описание совместных статистических свойств входной и выходной последовательностей с помощью многогранников, позволяющее разрабатывать методы идентификации автоматов по отрезкам выходных и, возможно, входных последовательностей без перебора начальных состояний.

5. Методы, позволяющие исследовать свойство устойчивости относительных частот встречаемости произвольных слов в растущих начальных отрезках выходной последовательности регистров сдвига и близких к ним автоматов.

6. Новый принцип классификации дискретных функций, основанный на использовании многогранников автоматов. Определены автоматы, сохраняющие значковые статистические характеристики двоичной входной последовательности. Разработанные методы позволяют исследовать наличие такого свойства у регистров сдвига.

7. Экспериментальные оценки точности мультиграммных методов идентификации языка текста со случайными искажениями в зависимости от уровня искажений и их теоретико-информационный анализ.

Теоретическая и практическая значимость. В диссертационной работе сделан вклад в теоретические основы методов идентификации автоматов со случайным и псевдослучайным входом. Предложенные в работе методы могут быть применены для разработки и анализа генераторов случайных чисел, для разработки методов поиска скрытых каналов в информационных системах, для решения задач технической диагностики дискретных устройств, для построения систем автоматической обработки текстовых материалов. Ряд результатов диссертационной работы использован при выполнении НИОКР, проводимых ООО «Лингвистические и информационные технологии» в интересах государственных заказчиков, где автор являлся научным руководителем и исполнителем, в том числе по гранту РФФИ и проекту «Семантика» ФПИ. Получено свидетельство о государственной регистрации разработанного программного средства.

Достоверность полученных результатов подтверждается математически корректными выводами и доказательствами теорем и других утверждений, корректностью разработанных математических моделей, использованием известных положений фундаментальных методов, соответствием полученных теоретических результатов данным проведенных вычислительных экспериментов и опубликованными результатами исследований других авторов.

Соответствие паспорту специальности. Диссертационное исследование выполнено в соответствии с паспортом специальности 05.13.17 «Теоретические основы информатики» и включает оригинальные результаты в области исследования процессов накопления и обработки информации, исследования методов преобразования информации в данные, создания и исследования информационных моделей, моделей данных, а также исследования принципов создания и функционирования аппаратных и программных средств автоматизации указанных процессов. Диссертационное исследование соответствует следующим разделам паспорта специальности 05.13.17 «Теоретические основы информатики»:

5. Разработка и исследование моделей и алгоритмов анализа данных, обнаружения закономерностей в данных и их извлечениях, разработка и исследование методов и алгоритмов анализа текста, устной речи и изображений.

10. Разработка основ математической теории языков и грамматик, теории конечных автоматов.

11. Разработка методов обеспечения высоконадежной обработки информации и обеспечения помехоустойчивости информационных коммуникаций для целей передачи, хранения и защиты информации; разработка основ теории надежности и безопасности использования информационных технологий.

Апробация работы. Научные положения и результаты диссертационной работы докладывались и получили одобрение на следующих семинарах и конференциях: Всероссийских симпозиумах по прикладной и промышленной математике (2004, 2006, 2007, 2009-2020), I, II и III международных конференциях «Системный анализ и информационные технологии» (2005-2009), Международных конференциях «Информационные технологии и искусственный интеллект» (Кацивели, 2006), (Дивноморское, 2007), XII и XIV Международных конференциях “Speech and Computer” SPECOM 2007, (Москва, 2007), SPECOM 2011 (Казань, 2011), X Международной научно-практической конференции «Информационная безопасность» (Таганрог, 2008), 4-й и 6-й Всероссийских научно-технических конференциях «Суперкомпьютерные технологии СКТ-2016, СКТ-2018, СКТ-2020» (Дивноморское, 2016, 2018, 2020), 8-й и 9-й Всероссийских мультиконференциях по проблемам управления «МКПУ-2015», «МКПУ-2017», «МКПУ-2019» (Дивноморское, 2015, 2017, 2019), II Балтийском Международном Симпозиуме по прикладной и промышленной математике и XXXIV Международном Семинаре по проблемам устойчивости стохастических систем (Светлогорск, 2016 г.), научно-практических конференциях «Комплексная защита информации» (Смоленск, 2016, Минск, 2021), на Международных конференциях «Распределенные компьютерные и телекоммуникационные сети: теория и приложения» DCCN-2018, DCCN-2020 (Москва, 2018, 2020), на III Международной конференции по инженерной и прикладной лингвистике «Пиотровские чтения-2019» (Санкт-Петербург, 2019), на X конференции с международным участием «Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем» ITTMM-2020 (Москва, 2020), на XX Международной конференции по проводным и беспроводным сетям и системам следующего поколения NEW2AN/ruSMART-2020 (Санкт-Петербург, 2020), IV Международной конференции по сетям будущего и распределенным вычислительным системам ICFNDs-2020 (Санкт-Петербург, 2020), на семинаре института экономики, математики и информационных технологий РАНХиГС (Москва, 2018), на межвузовских семинарах «Современные телекоммуникации и математическая теория телетрафика» (учредители: РУДН, МТУСИ, ИППИ РАН, ТГУ) (Москва, 2018-2021).

Публикация результатов работы. Основные результаты диссертации изложены в 48 опубликованных работах, в том числе в 1 монографии; в 17 статьях, опубликованных в журналах, включенных в Перечень РУДН/ВАК, в 11 статьях, опубликованных в источниках, индексируемых Web of Science и Scopus.

Личный вклад автора. Результаты, вынесенные на защиту, получены автором самостоятельно. Результаты по обобщенным графам де Брейна получены в соавторстве с Максимовским А.Ю., автору принадлежит метод доказательства. Решения по распознаванию языка искаженных текстов разработаны в соавторстве с Кулаем А.Ю., автору принадлежат постановка задач и методика исследований. Направления исследований диссертационной работы, формулировки проблем и постановки задач обсуждались с научным консультантом проф. К.Е. Самуйловым, что отражено в совместных публикациях, в которых основные результаты и их доказательства принадлежат диссертанту.

Структура и объем работы. Диссертация состоит из введения, пяти глав, заключения, списка литературы, содержащего 337 наименований. Основное содержание работы изложено на 265 страницах машинописного текста. Работа содержит 68 рисунков и 16 таблиц.

СОДЕРЖАНИЕ РАБОТЫ

Главы 1 и 2 работы посвящены методам распознавания автоматов со случайным входом, главы 3 и 4 посвящены методам идентификации автоматов в случае, когда вероятностная мера на множестве входных последовательностей не задается, в главе 5 сопоставляются методы идентификации конечных автоматов и методы распознавания естественного языка.

Во введении раскрыты актуальность, новизна, научная и практическая значимость диссертационной работы, сформулированы ее цель и задачи, перечислены основные научные результаты, выносимые на защиту, и раскрыто содержание основных глав диссертации.

В первой главе анализируются конечные автоматы со случайным входом, функция выходов которых неизвестна и подлежит распознаванию по наблюдениям над выходной последовательностью.

В **разделе 1.1** рассматриваются произвольные сильносвязные автоматы Мура. Пусть $A=(X,Y,Q,h,f)$ - конечный детерминированный сильносвязный автомат Мура, где $X=\{x_1,x_2,...,x_m\}$ - входной алфавит, $Y=\{y_1,y_2,...,y_k\}$ - выходной алфавит, $Q=\{q_1,q_2,...,q_r\}$ - множество состояний, $h:Q\times X\to Q$ - функция переходов, $f:Q\to Y$ - функция выходов, $r=1,2,...$, $m,k=2,3,...$. Пусть на множестве Q задано начальное вероятностное распределение $q^{(0)}=\left(q_1^{(0)},q_2^{(0)},...,q_r^{(0)}\right)$, на вход автомата A поступает последовательность независимых случайных величин (с.в.) $x^{(i)}$, $i=1,2,...$ с распределением $P\left(x^{(i)}=x_j\right)=p_j$, $p_j>0$, $j=1,2,...,m-1$, $p_m=1-\sum_{j=1}^{m-1}p_j>0$, $y^{(1)},y^{(2)},...$ - выходная последовательность. Пусть $\beta=y_0^{(1)}y_0^{(2)}...y_0^{(l)}$, $y_0^{(i)}\in Y$, $i=1,2,...,l$ - слово в выходном алфавите.

Относительной частотой встречаемости слова β в последовательности $y^{(1)},y^{(2)},...,y^{(N)}$, $N=l,l+1,...$ называется величина

$$Y_N=\frac{1}{N-l+1}\sum_{t=l}^{N}g^{(t)}, \qquad (1)$$

где через $g^{(t)}$ обозначена функция, вычисляемая по выходной последовательности длины $t\geq 1$ и слову β следующим образом:

$$g^{(t)}=\begin{cases}1, & \text{если последние } l \text{ символов выходной последовательности длины } t\\ & \text{составляют слово } \beta,\\ 0 & \text{в противном случае.}\end{cases}$$

При сформулированных условиях для с.в. Y_N при $N\to\infty$ существует предел по вероятности, который будем обозначать $P_\beta(\mathbf{p})$ и называть вероятностной функцией автомата A для слова β. Функция $P_\beta(\mathbf{p})$ определена на множестве

$$D=\left\{\mathbf{p}=(p_1,p_2,...,p_{m-1}),\ \sum_{j=1}^{m-1}p_j<1,\ p_j>0,\ j=1,2,...,m-1\right\}. \qquad (2)$$

Пусть $\mathrm{A}=\{A=(X,Y,Q,h,f),\ f\in F_{r,k}\}$ - класс описанных выше конечных сильносвязных автоматов Мура, функции выходов которых принадлежат множеству $F_{r,k}$ всех функций из Q в Y. Для заданного распределения $\mathbf{p}\in D$ задача распознавания

автомата $A \in \mathrm{A}$ исследуется в двух вариантах: по значению вероятностной функции $P_{\beta}(\mathbf{p})$ и по статистике Y_N. Эту задачу можно сформулировать как задачу распознавания неизвестной функции f выходов автомата в случае, когда известны его функция переходов и распределение с.в. $x^{(i)}$, $i = 1, 2, \ldots$

Две функции выходов назовем статистически эквивалентными, если соответствующие им вероятностные функции тождественно равны на D. В *Утверждениях 1.5 - 1.8* выведен критерий статистической эквивалентности функций выходов, установлены верхняя и нижняя границы числа классов эквивалентности.

По значению вероятностной функции статистически эквивалентные функции выходов неразличимы, поэтому рассматривается задача распознавания класса эквивалентности неизвестной функции выходов. Показано (*Утверждение 1.4*), что для почти всех $\mathbf{p} \in D$ (по мере Лебега в R^{m-1}) значение вероятностной функции позволяет однозначно определить класс эквивалентности функции выходов. *Утверждение 1.9* сводит задачу распознавания класса эквивалентности неизвестной функции выходов по статистике Y_N к задаче дискретной минимизации модуля линейной формы. Размерность задачи минимизации не превосходит $|Q|^l$. Приведена оценка необходимой длины последовательности.

В **разделе 1.2** результаты, полученные в первом разделе, уточняются для случая двоичного регистра сдвига и значковых свойств выходной последовательности. Пусть $V_n(q)$ - множество всех n-мерных q-ичных векторов, $q = 1, 2, \ldots$, $V_n = V_n(2)$, F_n - множество всех булевых функций от n аргументов, $n = 1, 2, \ldots$. Для $f(x_1, x_2, \ldots, x_n) \in F_n$ через $A_f = (X = \{0,1\}, V_n, Y = \{0,1\}, h, f)$ обозначим двоичный регистр сдвига с накопителем размера n, то есть автомат Мура с множеством состояний V_n, функцией переходов h, определяемой по правилу $h((a_1, \ldots, a_n), x) = (a_2, \ldots, a_n, x)$, где $x, a_i \in \{0,1\}$, $i = 1, 2, \ldots, n$, функцией выходов $f(x_1, x_2, \ldots, x_n)$.

Предположим, что на вход A_f поступает бернуллиевская последовательность независимых двоичных с.в. $x^{(i)}$, $i = 1, 2, \ldots$ с распределением $P\{x^{(i)} = 1\} = p$, $P\{x^{(i)} = 0\} = 1 - p$, $0 < p < 1$. Вероятностная функция автомата A_f для знака "1" выходной последовательности есть полином

$$P_{A_f}(p) = \sum_{j=0}^{n} s_j p^j (1-p)^{n-j}, \qquad (3)$$

где $s_k = \sum_{\substack{(x_1,\ldots,x_n) \\ \sum x_i = k}} f(x_1, \ldots, x_n)$, $k = 0, 1, \ldots, n$. Доказано (*Утверждение 1.11*), что статистическая эквивалентность булевых функций f и g равносильна системе равенств:

$$\sum_{\substack{(x_1,\ldots,x_n) \\ \sum x_i = k}} f(x_1, \ldots, x_n) = \sum_{\substack{(x_1,\ldots,x_n) \\ \sum x_i = k}} g(x_1, \ldots, x_n), \; k = 0, 1, \ldots, n. \qquad (4)$$

Получены (*Утверждение 1.14*) точные выражения для числа и мощности классов статистической эквивалентности. Показано (*Утверждение 1.15*), что число классов эквивалентности равно $\exp\left(n^2/2 + O(n)\right)$. Показано, что для почти всех $p \in (0,1)$, за исключением конечного множества «особенных распределений», значение вероятностной функции позволяет однозначно определить класс эквивалентности функции выходов. В частности, это справедливо при $0 < \left|p - 1/2\right| < 1/2 \cdot 4^{-n}$ (*Теорема 1.16*). Следовательно, если

при $p=1/2$ по значковой статистике можно указать только сумму значений функции f, то при "незначительном" отклонении от $1/2$ эта статистика позволяет определить класс статистической эквивалентности функции f, т.е. является существенно более информативной. Исследовано строение множества «особенных распределений». Показано (*Утверждение 1.12*), что его мощность не превосходит $\exp\left(n^2+O(n)\right)$. Задача распознавания f по значковой статистике сведена (*Утверждение 1.17*) к задаче целочисленной минимизации модуля линейной формы при линейных ограничениях:

$$\begin{cases} \left| Y_N - \sum_{i=0}^{n} s_j p^j (1-p)^{n-j} \right| \to \min \\ 0 \le s_j \le \binom{n}{j}, \; s_j \text{ - } \textit{целые}, \; 0 \le j \le n, \end{cases} \tag{5}$$

где $s_0,\ldots,s_n$ - вектор неизвестных, задающий класс статистической эквивалентности функции выходов. *Утверждение 1.18* дает оценку длины последовательности, при которой задача имеет единственное решение.

В **разделе 1.3** задача распознавания функции выходов исследуется для случая, когда на вход A_f поступает последовательность двоичных с.в., связанных в простую однородную стационарную цепь Маркова с матрицей переходных вероятностей

$$\begin{pmatrix} 1-\lambda & \lambda \\ \xi & 1-\xi \end{pmatrix}, \; 0<\lambda,\xi<1. \tag{6}$$

Функция $P_f(\lambda,\xi)=P\left\{ f\left(x^{(i)},x^{(i+1)},\ldots,x^{(i+n-1)}\right)=1 \right\}$ в диссертации названа вероятностной функцией автомата A_f для знака "1" при марковской входной зависимости. В *Утверждении 1.20* приведен явный вид $P_f(\lambda,\xi)$.

Статистическая эквивалентность двух функций выходов в этом случае равносильна (*Утверждение 1.21*) равенству сумм значений этих функций на $d(n)=\frac{3}{4}n^2-n+o(n)$ непересекающихся подмножествах n-мерного единичного куба. Получены (*Утверждение 1.22*) выражения для числа классов эквивалентности и их мощностей. Число классов эквивалентности асимптотически равно $\exp\left(\frac{5}{4}n^3\ln n+O(n^3)\right)$.

Показано (*Утверждение 1.24*), что для почти всех $(\lambda,\xi)\in(0,1)^2$, за исключением конечного числа гладких кривых, значение вероятностной функции $P_f(\lambda,\xi)$ позволяет однозначно определить класс эквивалентности функции выходов.

Задача распознавания f по значковой статистике сведена (*Утверждение 1.25*) к задаче целочисленной минимизации модуля линейной формы при линейных ограничениях. Размерность задачи минимизации равна $d(n)$.

В **разделе 1.4** рассматривается задача распознавания неизвестной функции выходов по значковым характеристикам выходных последовательностей для нескольких обобщений регистров сдвига.

Двоичный обобщенный по Imase и Itoh регистр сдвига (ОРС) порядка m, $m=1,2,\ldots$ – это автомат Мура $A_f^{(m)}=(X,Y,Q,h,f)$, где входной и выходной алфавиты есть $X=Y=\{0,1\}$, множество состояний $Q=\{0,1,\ldots,m-1\}$, функция переходов задана правилом $h(q,\varepsilon)=(2q+\varepsilon)\bmod m$, $\varepsilon=0,1$, функция выходов есть отображение

$f: Q \to \{0,1\}$. При $m = 2^t$ двоичный ОРС является обычным проходным регистром сдвига с накопителем размера t. Графом переходов ОРС является обобщенный по Imase и Itoh граф де Брейна.

Пусть $m = s2^k$, s - нечетно, $k \geq 0$. Для $0 \leq q \leq 2m-1$ обозначим $b(q)$ - количество единиц в двоичной записи числа $q \bmod 2^k$. Предположим, что на вход автомата $A_f^{(m)}$ поступает бернуллиевская последовательность с.в. с параметром $p, 0 < p < 1$. Вероятностная функция автомата $A_f^{(m)}$ для знака "1" имеет вид (*Утверждение 1.27*):

$$P_{A_f^{(m)}}(p) = \frac{1}{s} \sum_{i=0}^{k} \sum_{q \in S_i} f(q)\, p^i (1-p)^{k-i}, \text{ где } S_i = \{ q \in Q, \text{с условием } b(q) = i \},\ 0 \leq i \leq k. \quad (7)$$

Таким образом, вероятностные функции обобщенного по Imase и Itoh регистра сдвига при четном m совпадают с точностью до постоянного множителя с вероятностными функциями обычного регистра сдвига. Это означает, что к таким регистрам применимы построения раздела 1.2. При этом роль размера накопителя регистра сдвига, который определяет количество классов эквивалентности и размерности возникающих задач оптимизации, играет число k. При нечетном m ситуация является вырожденной, распределение на состояниях регистра равномерно, вероятностная функция не зависит от p и определяется только суммой значений функции f, которая легко устанавливается по значковой статистике.

Регистром сдвига с внутренним суммированием называется автомат Мура с двоичными входным $X = \{0,1\}$ и выходным $Y = \{0,1\}$ алфавитами, множеством состояний V_n, $n \geq 1$, булевой функцией выходов $f(x_1, x_2, ..., x_n)$, который под действием входного символа $a_0 \in \{0,1\}$ из состояния $(a_1, a_2, ..., a_n)$ переходит в состояние $(a_0 \oplus a_1, a_1 \oplus a_2, ..., a_{n-1} \oplus a_n)$, где $\oplus$ - суммирование по модулю 2. Такой автомат будем обозначать $A_f^{\oplus}$. Доказан (*Утверждение 1.28*) изоморфизм графов переходов автоматов A_f и $A_f^{\oplus}$.

Предположим, что на вход автомата $A_f^{\oplus}$ поступает бернуллиевская последовательность с.в. с параметром $p, 0 < p < 1$. Вероятностная функция автомата $A_f^{\oplus}$ имеет вид (*Утверждение 1.32*):

$$P_{A_f^{\oplus}}(p) = \frac{1}{2^n} \sum_{(x_1, x_2, ..., x_n) \in V_n} f(x_1, x_2, ..., x_n). \quad (8)$$

Этот результат означает, что вероятностная функция регистра сдвига с внутренним суммированием не зависит от p и определяется только суммой значений функции выходов, которая легко устанавливается по значковой статистике выходной последовательности. Поэтому число различных классов эквивалентности функций выходов в данном случае равно $2^n + 1$.

Глава написана на основании публикаций автора [3, 5, 6, 7, 10, 11, 20, 24, 37, 41].

Вторая глава посвящена использованию спектров графов в задаче распознавания неизвестной функции выходов автоматов и других теоретико-автоматных задачах. В главе показано, что вероятностные характеристики биграмм в выходной последовательности автомата со случайным входом можно описывать с помощью спектральных свойств неориентированного графа этого автомата. В **разделе 2.1** анализируется задача нахождения неизвестной функции выходов конечного вероятностного автомата по значениям вероятностей биграмм в его выходной последовательности. Пусть A - конечный автономный вероятностный автомат Мура, $A = (S, Y = \{0,1\}, H, f)$, где

$S = \{ s_1, s_2, ..., s_n \}$ - множество состояний, H - дважды стохастическая квадратная порядка n матрица, начальное распределение на множестве состояний является равномерным. Функция выходов $f : S \to [0,1]$ задается вектором $\mathbf{f} = (f(s_1), f(s_2), ..., f(s_n))$, $0 \le f(s_j) \le 1$. Автомат выдает на выходе символ 1 с вероятностью $f(s)$, символ 0 с вероятностью $1 - f(s)$. В этих условиях вероятность события $I_t = \{ y^{(j)} = 1, y^{(j+t)} = 1 \}$ в выходной последовательности (такое событие будем называть «биграммой I_t») не зависит от j и представляется (*Утверждение 2.1*) квадратичной формой

$$P(I_t) = \frac{1}{2n} \mathbf{f} \left(H^t + (H^t)^T \right) \mathbf{f}^T \qquad (9)$$

В случае, когда матрица H является симметрической и все ее собственные числа различны, матрицы квадратичных форм, описывающих вероятности $P(I_t)$, одновременно диагонализуемы. Это позволяет доказать, что системы уравнений относительно вектора неизвестных $\mathbf{f}$ вида

$$\left\{ P(I_t) = \frac{1}{2n} \mathbf{f} \left(H^t + (H^t)^T \right) \mathbf{f}^T, \; t = 1, 2, ...m, \right. \qquad (10)$$

при $m = n, n+1, n+2, ...$ эквивалентны между собой. При $m \ge n$ количество решений системы не превосходит 2^n (*Утверждение 2.2*). Известные условия одновременной диагонализуемости семейства матриц связаны с их коммутированием между собой. Для случая регистра сдвигов эти условия не выполняются, и воспользоваться предложенным методом линеаризации не удается.

Сложность квадратичного уравнения характеризуется рангом квадратичной формы, т.е. минимальным числом переменных квадратичной формы, которое может быть получено при всевозможных невырожденных линейных преобразованиях вектора неизвестных. Это число равно количеству ненулевых характеристических чисел матрицы квадратичной формы. Под спектром графа (орграфа, мультиграфа) будем понимать множество характеристических чисел его матрицы смежности. Графом G_n де Брейна степени n называется ориентированный граф с множеством вершин $V_n(q)$, $q = 2, 3, ...$, $n = 1, 2, ...$ содержащий дугу, выходящую из вершины $(a_1, a_2, ..., a_n)$ и заходящую в вершину $(b_1, b_2, ..., b_n)$ в том и только в том случае, когда $(a_2, a_3, ..., a_n) = (b_1, b_2, ..., b_{n-1})$. Граф де Брейна обычно интерпретируется как граф переходов q-ичного регистра сдвигов с накопителем размера n. Пусть G_n' - орграф, полученный из G_n изменением направления всех дуг на противоположное. Псевдограф (содержащий петли и кратные ребра) $\tilde{G}_n = G_n \cup G_n'$ называется неориентированным графом де Брейна. Пусть G_n^t - t-я степень графа де Брейна, $\tilde{G}_n^{(t)} = (G_n)^t \cup (G_n')^t$. В **разделе 2.2** предложено преобразование унитарного подобия матрицы смежности графа $\tilde{G}_n^{(t)}$, которое приводит к разреженной матрице, являющейся, в свою очередь, матрицей смежности нового графа. Этот граф является распавшимся (см. Рис.1), и его спектр легко вычисляется.

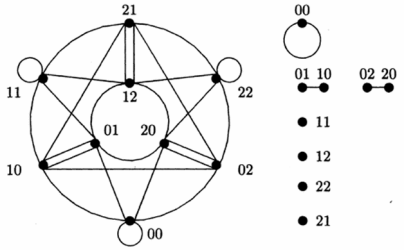


Рис.1. Граф $\tilde{G}_2$ и построенный новый граф при $q=3$.

Утверждение 2.6. Пусть $n=lt+r, 0\le r<t$. Характеристический многочлен графа $\tilde{G}_n^{(t)}$ имеет вид:

$$\chi_{\tilde{G}_n^{(t)}}(\lambda)=q^{t(q^n-1)}(\lambda-2q^t)\prod_{k=1}^{\left[\frac{n}{t}\right]+1}\left\{U_k\left(\frac{\lambda}{2q^t}\right)\right\}^{L_k}, \qquad (11)$$

где $U_k(x)=\frac{1}{\sqrt{1-x^2}}\sin((k+1)\arccos x)$ - многочлен Чебышева второго рода,

$$L_k=\begin{cases}(q^t-1)^2 q^{n-t(k+1)} & при\ 1\le k<l,\\ q^{t+r}-2q^r+1 & при\ k=l,\\ q^r-1 & при\ k=l+1.\end{cases}$$

В разделе рассматриваются q-ичные проходные регистры сдвига с накопителем размера n, $n=1,2,...$ со случайным равновероятным входом и независимой от него случайной равновероятной двоичной последовательностью управления движением. Сдвиг регистра в зависимости от символа управляющей последовательности осуществляется на a или b тактов. В *Утверждениях 2.8 и 2.9* вычислены ранги квадратичных форм, описывающих вероятности биграмм I_1 для вариантов $(a,b)=(0,1)$ и $(a,b)=(1,2)$.

Найденный спектр графа $\tilde{G}_n^{(t)}$ позволил получить (*Утверждение 2.7*) верхнюю границу числа $\alpha(G_n^t)$ независимости (в другой терминологии – числа внутренней устойчивости) графа. Это число определяется как максимальное количество несмежных вершин в графе. Для классического графа де Брейна ($t=1$) оценкам этого числа посвящен ряд работ, доказанный в диссертации результат улучшает известную при $q=2$ верхнюю границу $\alpha(G_n)<2^{n-1}$. В асимптотическом виде он формулируется следующим образом. При $n\to\infty$ для некоторой последовательности $\delta_n\to 0$ справедливо

$$\alpha(G_n)\le(1+\delta_n)\left(1-\frac{\pi^2}{2n^2}\right)\frac{q^n}{2}. \qquad (12)$$

В случае $q=2$ доказанный результат приводит к следующей последовательности границ для $\alpha(G_n)$: {1, 2, 3, 7, 14, 30, 61, 124, 249, 501, 1006, …}, в то время как известные сегодня первые члены истинной последовательности имеют вид {1, 2, 3, 7, 13, 28, 55, 114, 227, 466, …}.

В **разделе 2.3** изучается подкласс обобщенных по Imase и Itoh графов де Брейна, т.н. «редуцированные» графы де Брейна. Этот подкласс соответствует наиболее содержательной ситуации, когда количество вершин графа кратно степени его регулярности. Такие графы в семействе всех обобщенных графов являются в определенном смысле «самыми близкими» к классическим графам де Брейна.

Редуцированным графом $G(n,m)$ де Брейна назван граф на множестве вершин Ω_{nm}, в котором из вершины $i\in\Omega_{nm}$ выходит ровно n дуг, заходящих в вершины $(in+\varepsilon)\bmod nm$, $\varepsilon\in\Omega_n$, $n,m=1,2,...$. В разделе исследуются различные свойства таких графов, приведен вид их характеристического многочлена. Отметим, что спектр обобщенных графов получен Li и Zhang в 1991 г. Приводимое в диссертации доказательство является более кратким и основано на теореме о виде характеристического многочлена k-циркулянтной матрицы. Спектральными методами для этих графов

вычислены числа петель, циклов длины 2, остовных деревьев и эйлеровых циклов, диаметр и указано множество значений параметров n и m, для которых графы обладают свойствами единственности пути заданной длины. Доказано, что в семействе графов $G(n,m)$, $n \geq 3$, $m = 1,2,\ldots$ планарными являются только графы $G(3,2), G(3,3), G(3,4)$. Сформулирована гипотеза: граф $G(2,m)$ планарен при $m = 1\text{-}7, 9$; при $m = 8$ и $m \geq 10$ граф $G(2,m)$ непланарен. Гипотеза подтверждена с помощью компьютерных вычислений в среде Mathematica для $m = 1, 2, \ldots, 257$.

Глава написана на основании публикаций автора [2, 16, 24, 26, 34, 42, 45, 46].

В **третьей** главе развит подход к задаче идентификации, связанный с построением в действительном кубе подходящей размерности специального многогранника, соответствующего автомату. В **разделах 3.1-3.3** даны основные определения. Пусть B - конечное множество (алфавит), B^* - множество всех слов в алфавите B, Ω_B - множество всех бесконечных последовательностей над B. Для $\alpha = a_0 a_1 \ldots a_{m-1} \in B^*$, $\omega = (w_0, w_1, \ldots) \in \Omega_B$ обозначим

$$p_\alpha(\omega) = \lim_{t \to \infty} \frac{1}{t} \sum_{j=0}^{t-1} \delta(w_j w_{j+1} \ldots w_{j+m-1}, a_0 a_1 \ldots a_{m-1}), \quad (13)$$

где δ - символ Кронекера, если предел в правой части существует. Величину $p_\alpha(\omega)$ можно интерпретировать как среднюю частоту встречаемости слова α в последовательности ω. Рассматриваемые пределы существуют, например, для бесконечных периодических последовательностей (как чисто периодических, так и последовательностей с подходом), множество которых обозначим T_B. В этом случае p_α является отношением частоты встречаемости слова α на периоде к длине периода.

Пусть $A = (X, Y, Q, h, f)$ - конечный сильносвязный автомат Мили, где X и Y - входной и выходной алфавиты, Q - множество состояний; $h: Q \times X \to Q$ - функция переходов; $f: Q \times X \to Y$ - функция выходов. Зафиксируем два набора слов: $\{\alpha_i \in X^*, i = 1,2,\ldots,t\}$ и $\{\beta_j \in Y^*, j = 1,2,\ldots,k\}$, $t \geq 0, k \geq 1$. Предположим, что автомат A, начиная работать из состояния q_0, перерабатывает последовательность $\chi = (x^{(1)}, x^{(2)}, \ldots)$ в последовательность $\gamma = (y^{(1)}, y^{(2)}, \ldots)$. Последовательности χ поставим в соответствие вектор

$$\mathbf{z}_{(A,q_0)}(\chi) = (p_{\alpha_1}(\chi), \ldots, p_{\alpha_t}(\chi), p_{\beta_1}(\gamma), \ldots, p_{\beta_k}(\gamma)), \quad (14)$$

если все величины, стоящие в правой части, определены.

Для автомата A это соответствие задает отображение

$$Z_{(A,q_0)}: \mathrm{T}_X \to [0,1]^{t+k} \subset R^{t+k}. \quad (15)$$

Через R_A обозначим замыкание множества $Z_{(A,q_0)}(\mathrm{T}_X)$. Корректность принятого обозначения следует из сильной связности автомата.

Пусть l - максимум из длин слов множеств $\{\alpha_i \in X^*, i = 1,2,\ldots,t\}$ и $\{\beta_j \in Y^*, j = 1,2,\ldots,k\}$. Определим автомат $A^{(l)} = (X, Y, Q^{(l)}, h^{(l)}, f^{(l)})$, положив $Q^{(l)} = \{((q^{(1)}, x^{(1)}), (q^{(2)}, x^{(2)}), \ldots, (q^{(l-1)}, x^{(l-1)}), q^{(l)})$, где $h(q^{(i)}, x^{(i)}) = q^{(i+1)}$, $i = 1,2,..,l\text{-}1$; $q^{(j)} \in Q$, $j = 1,2,..,l$, $x^{(j)} \in X$, $j = 1,2,..,l\text{-}1\}$ - множество состояний;

$h^{(l)}: Q^{(l)} \times X \to Q^{(l)}$ - функция переходов; $h^{(l)}\left(\left(\left(q^{(1)}, x^{(1)}\right), \left(q^{(2)}, x^{(2)}\right), ..., \left(q^{(l-1)}, x^{(l-1)}\right), q^{(l)}\right), x\right) =$ $\left(\left(q^{(2)}, x^{(2)}\right), ..., \left(q^{(l-1)}, x^{(l-1)}\right), \left(q^{(l)}, x\right), h\left(q^{(l)}, x\right)\right)$;

$f^{(l)}: Q^{(l)} \times X \to Y$ - функция выходов;

$f^{(l)}\left(\left(\left(q^{(1)}, x^{(1)}\right), \left(q^{(2)}, x^{(2)}\right), ..., \left(q^{(l-1)}, x^{(l-1)}\right), q^{(l)}\right), x\right) = f\left(q^{(l)}, x\right)$.

Пусть $C_l(A)$ - множество всех циклов в графе переходов автомата $A^{(l)}$, дуги $\left(q, h^{(l)}(q, x)\right)$ которого помечены парой символов $\left(x, f^{(l)}(q, x)\right)$. С каждым циклом из $C_l(A)$ свяжем циклические последовательности первых и вторых координат меток дуг цикла. Эти последовательности будем называть входной и выходной разметками цикла соответственно. Для $c \in C_l(A)$ введем обозначения: $l(c)$ - длина цикла, $\nu_\alpha(c)$ - относительная частота встречаемости слова α во входной разметке цикла c, $\nu_\beta(c)$ - относительная частота встречаемости слова β в выходной разметке цикла c, $\mathbf{z}(c) = \left(\nu_{\alpha_1}(c), ..., \nu_{\alpha_k}(c), \nu_{\beta_1}(c), ..., \nu_{\beta_t}(c)\right)$ - вектор относительных частот.

Под расстоянием ρ между двумя точками в R^n будем понимать максимум модулей разностей координат: $\rho(u, v) = \max_{1 \le i \le n} \left| u^{(i)} - v^{(i)} \right|$, $Conv\, E$ обозначает выпуклую оболочку множества E. Доказаны следующие теоремы.

Теорема 3.1. Справедливо равенство

$$R_A = Conv\left\{ \mathbf{z}(c), c \in C_l(A) \right\}. \qquad (16)$$

Теорема 3.2. Пусть автомат A перерабатывает последовательность $\chi^{(N)} = \left(x^{(1)}, x^{(2)}, ..., x^{(N)}\right)$ в последовательность $\gamma^{(N)} = \left(y^{(1)}, y^{(2)}, ..., y^{(N)}\right)$, $N = 1, 2, ...$ Пусть $\mathbf{z}^{(N)} = \left(p^{(N)}_{\alpha_1}, ..., p^{(N)}_{\alpha_t}, p^{(N)}_{\beta_1}, ..., p^{(N)}_{\beta_k}\right)$ - вектор относительных частот встречаемости слов $\alpha_1, ..., \alpha_t$ (для $\chi^{(N)}$) и $\beta_1, ..., \beta_k$ (для $\gamma^{(N)}$), l - максимум из длин этих слов, D - диаметр графа переходов автомата A. Тогда

$$\rho\left(\mathbf{z}^{(N)}, R_A\right) \le \frac{D + 2(l - 1)}{N + D + l - 1}. \qquad (17)$$

На основе этих результатов в **разделе 3.4** предложена следующая процедура проверки гипотезы о том, что неизвестный (выбранный из некоторого класса неэквивалентных автоматов, содержащего A_0) автомат A, входная и выходная последовательности которого наблюдаются, совпадает с известным автоматом A_0.

1. Выбирается набор слов $\alpha_1, ..., \alpha_t$ (для входной) и $\beta_1, ..., \beta_k$ (для выходной последовательности) и строится многогранник $R_0 \subset \left[0, 1\right]^{t+k}$ автомата A_0.

2. Подсчитываются относительные частоты встречаемости этих слов в имеющихся последовательностях. Вычисляется расстояние ρ между R_0 и вектором относительных частот.

3. Если $\rho > \frac{D + 2(l - 1)}{N + D + l - 1}$, то имеющаяся выходная последовательность не могла быть получена из входной при помощи автомата A_0. Если $\rho \le \frac{D + 2(l - 1)}{N + D + l - 1}$, то

наблюдаемые частоты встречаемости выделенных слов не противоречат проверяемой гипотезе.

В последнем случае целесообразно либо перейти к другому отрезку имеющихся последовательностей, либо изменить набор слов, частоты которых анализируются.

Описанная процедура справедлива для произвольных множеств слов $\{\alpha_1,...,\alpha_t\}$ и $\{\beta_1,...,\beta_k\}$. В частности, множество слов $\{\alpha_1,...,\alpha_t\}$ может быть пустым. В этом случае процедура будет использовать соотношения частот слов только в выходной последовательности автомата. Если алфавиты автомата двоичны, $k=t=1$, $\alpha_1=\beta_1=1$, то многогранник автомата является плоским многоугольником в квадрате $[0,1]\times[0,1]$.

Приведены оценки трудоемкости вычислений по этой процедуре.

Первый, предварительный этап вычислений состоит в построении многогранника, причем самым сложным является перебор всех циклов в графе. Объем такого перебора можно оценить сверху числом всех подграфов $2^{(|Q|\times|X|)^t}$. Для некоторых автоматов в главе 4 предложены аналитические методы построения многогранников.

На втором этапе вычислений используются наблюдения над имеющимися последовательностями. Показано, что трудоемкость проверки неравенства (17) в описанной процедуре в двумерном случае не превосходит $O(Log\,|Q|)$.

Предел, определяющий величину $p_\alpha(\omega)$, существует в случае, когда последовательность ω является периодической. В **разделах 3.5 и 3.6** рассматриваются более широкие классы последовательностей. Последовательность $\omega\in\Omega_B$ называется чезаровской, если предел $p_\alpha(\omega)$ существует для произвольного $\alpha\in B^*$. Чезаровских последовательностей «очень много» в следующем смысле.

Утверждение 3.1. Пусть $(w_0,w_1,...)$ - представление действительного числа $x\in(0,1)$ в виде $|B|$-ичной дроби. Мера Лебега множества тех $x\in(0,1)$, для которых последовательность $(w_0,w_1,...)$ является чезаровской, равна 1.

Если во входной последовательности конечного автомата существуют пределы относительных частот встречаемости каждого возможного слова, то интуитивно то же можно ожидать и от выходной последовательности. Однако это справедливо не для всех автоматов. Автомат назовем *чезарово-наследственным*, если для произвольного начального состояния чезаровские последовательности во входном алфавите он перерабатывает в чезаровские последовательности в выходном алфавите.

Показано, что автомат с конечным запоминанием является чезарово-наследственным (*Утверждение 3.2).*

Утверждение 3.5. Пусть $A=(\{0,1\},\{0,1\},Q,h,f)$ - сильносвязный автомат, $k=t=1$, $\alpha_1=\beta_1=1$. Если автомат A является чезарово-наследственным, то его многоугольник не содержит параллельных оси ординат сторон.

Следующие результаты (*Утверждения 3.6, 3.8, 3.9*) характеризуют чезарово-наследственность регистров сдвига и их обобщений.

Двоичный регистр сдвига A_f является чезарово-наследственным автоматом для произвольной функции f.

Если $m=2^t$, $t\ge0$, то при любой функции выходов f автомат $A_f^{(m)}$ является чезарово-наследственным. Если $m\ne2^t$, то найдется функция выходов f, для которой автомат $A_f^{(m)}$ не является чезарово-наследственным.

Если для f выполнено условие $f(0,0,...,0) \neq f(0,0,...,0,1)$, то $A_f^{\oplus}$ не чезарово-наследственный автомат.

Отметим, что доказательство отсутствия чезарово-наследственности проведено конструктивно, с предъявлением таких чезаровских входных последовательностей, для которых выходная последовательность не является чезаровской.

В **разделе 3.7** определены и изучены автоматы, сохраняющие значковые статистические свойства входной последовательности.

Пусть $A = (\{0,1\}, \{0,1\}, Q, h, f)$ – конечный автомат Мили с двоичными входным и выходным алфавитами. Начальное состояние, входную и выходную последовательности автомата обозначим $q \in Q$, $(x^{(1)}, x^{(2)}, ...)$, $(y^{(1)}, y^{(2)}, ...)$. Автомат A назван сохраняющим значковые статистические свойства входной последовательности, если равенство

$$\lim_{t\to\infty} \frac{1}{t} \sum_{j=1}^{t} x^{(j)} = \lim_{t\to\infty} \frac{1}{t} \sum_{j=1}^{t} y^{(j)} \tag{18}$$

выполнено для всех $q \in Q$ и всех бесконечных двоичных периодических (возможно, с подходом) последовательностей $(x^{(1)}, x^{(2)}, ...)$. Это условие равносильно тому, что для $k = t = 1$, $\alpha_1 = \beta_1 = 1$ многоугольник автомата совпадает с диагональю квадрата: $R_A = [(0,0),(1,1)]$.

В разделе исследуется, какие автоматы из классов A_f, $A_f^{(m)}$, $A_f^{\oplus}$ обладают этим свойством.

Множество функций $f(x_1, x_2, ..., x_n) \in F_n$, для которых автомат A_f сохраняет значковые статистические свойства входной последовательности, обозначим M_n.

Оказывается (*Утверждение 3.11*), при $f \notin M_n$, $n = 2,3,...$ автомат A_f ухудшает значковые статистические свойства некоторых входных последовательностей, в следующем смысле**.** Существует последовательность $(x^{(1)}, x^{(2)}, ...)$ со свойством

$$\lim_{t\to\infty} \frac{1}{t} \sum_{j=1}^{t} x^{(j)} = \frac{1}{2}, \tag{19}$$

для которой

$$\left| \lim_{t\to\infty} \frac{1}{t} \sum_{j=1}^{t} f\left(x^{(j)}, ..., x^{(j+n-1)}\right) - \frac{1}{2} \right| \geq \frac{1}{2} 4^{-n}. \tag{20}$$

Утверждение 3.12. Пусть $C(G_n)$ - множество простых циклов графа G_n. Регистр A_f сохраняет значковые свойства входной последовательности тогда и только тогда, когда f сохраняет вес каждого цикла $c \in C(G_n)$, т.е. выполнено равенство

$$\|f/c\| = \sum_{(x_1,x_2,...,x_n)\in c} f(x_1, x_2, ..., x_n) = \sum_{(x_1,x_2,...,x_n)\in c} x_1 = \|c\|, \tag{21}$$

где суммирование производится по всем двоичным векторам – вершинам цикла c.

Утверждение 3.13. Пусть $f \in M_n$, на вход регистра A_f с начальным состоянием $\alpha^{(0)} \in V_n$ поступает двоичная последовательность $x^{(1)}, x^{(2)}, ..., x^{(N)}$, снимается выходная последовательность $y^{(1)}, y^{(2)}, ..., y^{(N)}$, $N \geq 1$. Тогда справедливо неравенство

$$\left| \sum_{i=1}^{N} x^{(i)} - \sum_{i=1}^{N} y^{(i)} \right| \leq n. \tag{22}$$

Показано (*Утверждение 3.14*), что если $f \in M_n$, то

1) $f(0,0,...,0)=0$, $f(1,1,...,1)=1$.

2) $f(0,1,0,1...)\oplus f(1,0,1,0,...)=1$.

3) f – равновероятна.

Доказано (*Утверждение 3.15*), что функции $f(x_1,x_2,...,x_n)$, $f(x_n,x_{n-1},...,x_1)$ и $\overline{f}(\overline{x_1},\overline{x_2},...,\overline{x_n})$, где «$\overline{\ }$» - значок инверсии, принадлежат или не принадлежат M_n одновременно.

Утверждение 3.16. Пусть $f\in M_n$ и f отлична от координатных функций. Тогда f является нелинейной по всем аргументам, от которых она существенно зависит.

С помощью соотношения ортогональности для графа де Брейна в *Утверждениях 3.17 и 3.18* получено следующее псевдобулево представление для вектора $\mathbf{f}=(f(0,0,...,0),f(0,0,...,1),...,f(1,1,...,1))^T$ табличного задания функции $f\in M_n$.

Пусть $\mathbf{B}_n$ - $2^{n+1}\times 2^n$ матрица инцидентности графа G_n, $f_0\in M_n$. Для $f\in M_n$ существует единственный 2^{n-1} – мерный вектор $\mathbf{d}$ с условием $\mathbf{d}(0,0,...,0)=0$, такой, что

$$\mathbf{f}=\mathbf{B}_{n-1}\mathbf{d}+\mathbf{f}_0. \qquad (23)$$

Если $f_0=x_1\in F_n$, то вектор $\mathbf{d}$ является целочисленным и выполнены неравенства $0\le\mathbf{d}(a_1,a_2,...,a_{n-1})\le\sum_{i=1}^{n-1}a_i$.

В серии утверждений (*3.20-3.25*) оценивается мощность рассматриваемого класса функций. Показано, что $|M_n|\le 2^{2^{n-1}-1}$, $n=1,2,...$. С помощью конструктивных построений установлены две нижние границы мощности множества M_n, которые приводят к соотношению

$$\lim_{n\to\infty}\frac{\log_2\log_2|M_n|}{n}=1. \qquad (24)$$

Следующие два утверждения характеризуют сохранение значковых свойств обобщениями регистра сдвига.

Утверждение 3.26. Пусть $t=0,1,2,...$. При $m=4t+1$ и $m=4t+3$ автомат $A_f^{(m)}$ не сохраняет значковые свойства входа. При $m=4t+2$ автомат $A_f^{(m)}$ сохраняет значковые свойства входа только для $f(q)=q \bmod 2$.

С помощью компьютерных вычислений получена таблица количеств ОРС, сохраняющих значковые свойства входа для первых нескольких m, кратных 4: (8, 5), (12, 8), (16, 22), (20, 8), (24, 62), (28, 24), (32, 428), (36, 32), (40, 381), (44, 8), (48, 4056), (52, 8), (56, 3538), (60, 176), (64, 133184).

Утверждение 3.27. Автомат $A_f^{\oplus}$, $f\in F_n$, не сохраняет значковые свойства входа.

Глава 3 написана на основании публикаций автора [1, 8, 9, 12, 13, 14, 15, 17, 19, 21, 24, 39, 44, 47, 48].

В **четвертой** главе подход к идентификации автоматов на основе их многогранников развит для нескольких классов автоматов, близких к регистрам сдвига. Приводятся как аналитические результаты, так и результаты компьютерных вычислений при небольших размерах накопителя.

В **разделе 4.1** в качестве исследуемого автомата выступает автомат A_f, рассматриваются значковые свойства входной и выходной последовательностей, поэтому многогранники регистров A_f являются плоскими многоугольниками и определяются только булевой функцией $f\in F_n$.

Теорема 4.1. Справедливо равенство

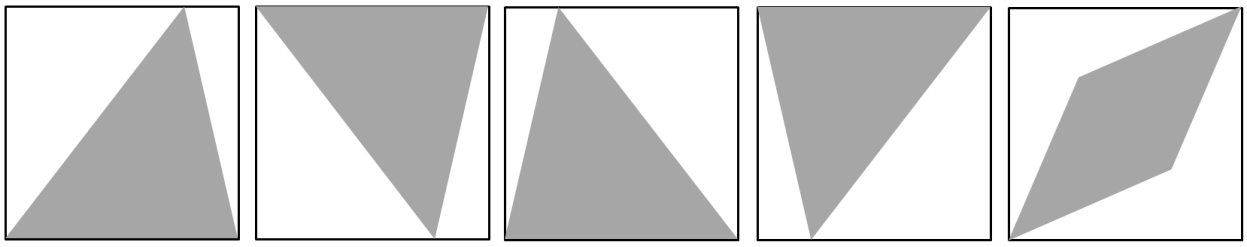
$$R_{A_f} = Conv\left\{\left(\frac{\|c\|}{l(c)}, \frac{\|f(c)\|}{l(c)}\right), c \in C(G_n)\right\}, \quad (25)$$

где c - цикл из множества $C(G_n)$ всех циклов графа G_n, $l(c)$ - его длина, $\|c\| = \sum_{(x_1,x_2,...,x_n)\in c} x_1$ - число единиц в его двоичной записи, $\|f/c\| = \sum_{(x_1,x_2,...,x_n)\in c} f(x_1,x_2,...,x_n)$ - вес функции f на вершинах цикла c, где обе последних суммы берутся по всем вершинам цикла c.

Теорема 4.2. Предположим, что автомат A_f, начиная работать из некоторого начального состояния, перерабатывает двоичную последовательность $\chi^{(N)} = (x^{(0)}, x^{(1)}, ..., x^{(N-1)})$ в двоичную последовательность $y^{(N)} = (y^{(0)}, y^{(1)}, ..., y^{(N-1)})$, $N \geq 1$. Справедливо неравенство

$$\rho\left(\left(\frac{1}{N}\sum_{i=0}^{N-1} x^{(i)}, \frac{1}{N}\sum_{i=0}^{N-1} y^{(i)}\right), R_{A_f}\right) \leq \frac{n}{N+n}. \quad (26)$$

Многоугольники функций будем изображать в квадрате $0 \leq z_1, z_2 \leq 1$, ось $(0\ z_1)$ горизонтальна, ось $(0\ z_2)$ вертикальна.



$f(x_1,x_2,...,x_n)$ $\overline{f(x_1,x_2,...,x_n)}$ $f(\overline{x}_1,\overline{x}_2,...,\overline{x}_n)$ $\overline{f(\overline{x}_1,\overline{x}_2,...,\overline{x}_n)}$ f - самодвойственна

Рис.2. Отношения двойственности булевых функций и симметрия многоугольников.

Многоугольники функций $f(x_1,x_2,...,x_n)$ и $\overline{f(x_1,\overline{x_2},...,x_n)}$ симметричны относительно горизонтальной оси $z_1 = 1/2$. Многоугольники функций $f(x_1,x_2,...,x_n)$ и $f(\overline{x}_1,\overline{x}_2,...,\overline{x}_n)$ симметричны относительно вертикальной оси $z_2 = 1/2$. Многоугольник самодвойственной функции центральносимметричен относительно $(1/2, 1/2)$ (см. Рис.2). Все точки графика вероятностной функции автомата A_f принадлежат его многоугольнику (Рис.3).

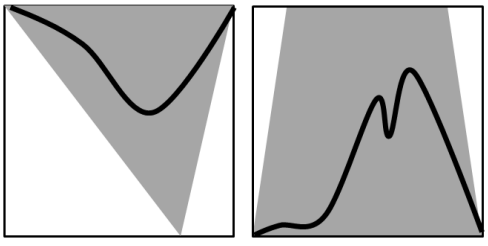


Рис.3. Схематический вид многоугольников R_{A_f} и графиков вероятностных функций автоматов A_f.

Если $\sum_{(x_1,...,x_n)} f(x_1,...,x_n) > 2^n - \frac{1}{n}\sum_{d/n}\phi(\frac{n}{d})2^d$, где $\phi(n)$ - функция Эйлера, то в многоугольнике R_{A_f} имеется точка с ординатой 1. Среди функций веса k, $0 \le k \le 2^n - \frac{1}{n}\sum_{d/n}\phi(\frac{n}{d})2^d$, существуют такие, многоугольники которых целиком лежат ниже прямой $z_2 = 1$.

Многоугольники функций веса 1 являются треугольниками, их вид описан в *Утверждении 4.8*.

Справедливо включение

$$R_{A_f} \subseteq Conv\left\{(0, f(0,0,...,0)), \left(1/n, \overline{f(0,0,...,0)}\right), \left(1 - 1/n, \overline{f(1,1,...,1)}\right), (1, f(1,1,...,1))\right\}. \quad (27)$$

Максимальные по включению многоугольники описываются в *Утверждениях 4.9 и 4.10*. Автомат A_f обладает максимальным многоугольником тогда и только тогда, когда значение $f(x_1, x_2, ..., x_n)$ на каждом из векторов веса 1 равно $\overline{f(0,0,...,0)}$, на каждом из векторов веса $n - 1$ равно $\overline{f(1,1,...,1)}$.

Через $v(f)$ обозначим число вершин многоугольника R_{A_f}, $f \in F_n$.

Утверждение 4.11. Для $n \ge 3$ справедливо:

если $f(0,0,...,0) = f(1,1,...,1)$, то $v(f) \le 1 + \sum_{k=1}^{2^n}\phi(k) < \frac{4^n}{2}$,

если $f(0,0,...,0) \ne f(1,1,...,1)$, то $v(f) \le 2\sum_{k=1}^{2^n}\phi(k) < 4^n$.

Утверждения 4.12-4.16 описывают многоугоугольники некоторых линейных функций и функций, инвариантных относительно циклического сдвига аргументов (Рис.4).

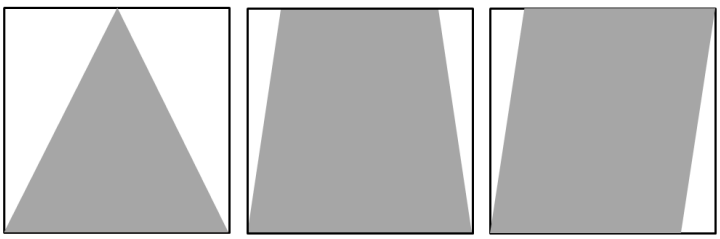


Рис.4. Многоугольники функций $x_1 \oplus x_n$, $\sum_{i=1}^{n} x_i$ при четном n, $\sum_{i=1}^{n} x_i$ при нечетном n.

Если многочлен $a(x) = \sum_{i=0}^{n-1} a_i x^i \in GF(2)[x]$ является примитивным степени $n - 1$, $n \ge 4$, то число вершин многоугольника булевой функции $\sum_{i=1}^{n} a_{i-1} x_i$ не менее шести.

Если $f(x_1, x_2, ..., x_n)$ инвариантна относительно циклического сдвига аргументов, т.е. $f(x_1, x_2, ..., x_n) = f(x_2, x_3, ..., x_n, x_1)$, то число вершин многоугольника R_{A_f} не превосходит четырех.

Все четыре функции одного переменного имеют различные многоугольники. Функциям двух переменных соответствуют 12 различных многоугольников. Имеется 68

различных многоугольников функций трех переменных. Имеется 1520 различных многоугольников функций четырех переменных, среди которых 4 отрезка, 106 треугольников, 576 четырехугольников, 662 пятиугольника, 164 шестиугольника и 8 семиугольников. Максимальное число функций от четырех переменных соответствует треугольнику Conv {(0,0), (1/2,1), (1,0)} и симметричному ему (по 1989 каждому).

На Рис.5 представлены данные компьютерного моделирования с автоматом A_f, $f(x_1,x_2,x_3,x_4)=\begin{cases}1, \textit{если } x_1+x_2+x_3+x_4\geq 2,\\ 0, \textit{в противном случае.}\end{cases}$ Многоугольник этого автомата есть $Conv\{(0,0),(1/4,0),(1/2,1),(1,1)\}$. Автомат обрабатывал случайную последовательность длиной 10000 знаков. Подсчитывались совместные относительные частоты встречаемости единиц в отрезках длины N во входной и выходной последовательностях. Соответствующие им точки отмечались на рисунках. Представленные данные иллюстрируют *Теорему 4.2* для различных значений N.

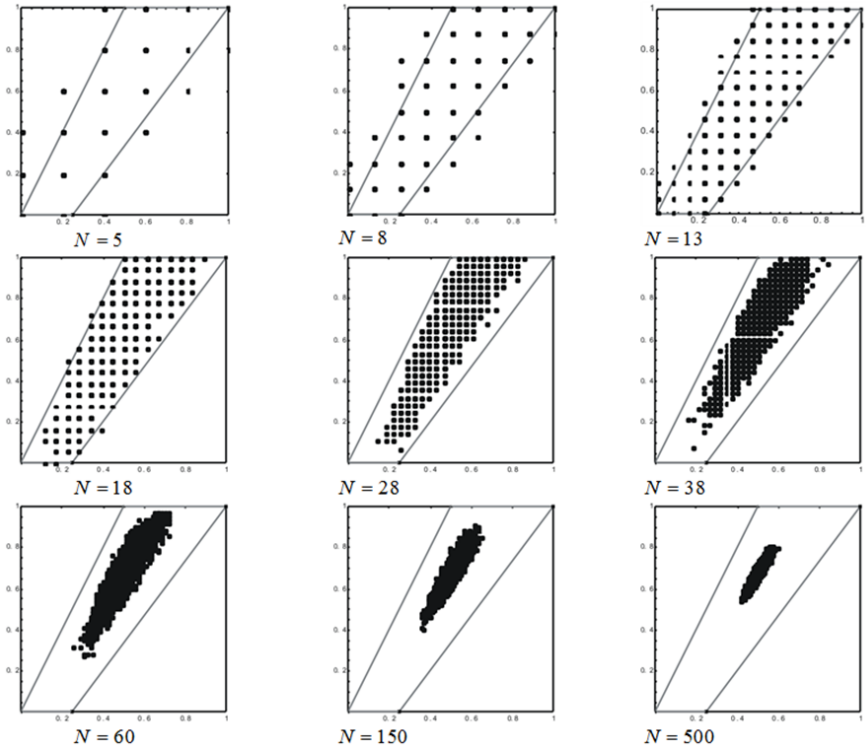


Рис.5. Результаты компьютерного моделирования для автомата A_f.

В **разделе 4.2** изучаются значковые статистические свойства входной и выходной последовательностей двоичных регистров сдвига с внутренним суммированием $A_f^{\oplus}$, $f(x_1,x_2,...,x_n)\in F_n$, $n=1,2,\ldots$. Предложен способ построения многоугольников таких автоматов и процедура проверки совпадения наблюдаемого автомата с эталонным, основанная на использовании аппарата многоугольников. Результаты здесь аналогичны *Теоремам 4.1 и 4.2*, с учетом того, что вместо графа G_n рассматривается граф переходов автомата $A_f^{\oplus}$.

Доказан ряд утверждений о строении многоугольников автоматов $A_f^{\oplus}$. Многоугольники $R_{A^{\oplus}_{f(x_1,x_2,\ldots,x_n)}}$ и $R_{A^{\oplus}_{\overline{f(x_1,x_2,\ldots,x_n)}}}$ симметричны относительно горизонтальной оси $z_2=1/2$. На основе установленного в *Утверждении 1.28* изоморфизма графов показано, что многоугольники регистров сдвига с внутренним суммированием, функции выходов

которых отличаются инверсией последней переменной, совпадают, $R_{A^{\oplus}_{f(x_1,x_2,...,x_n)}} = R_{A^{\oplus}_{f(x_1,x_2,...,\overline{x_n})}}$. В отличие от случая, исследованного в предыдущем разделе, автомат $A^{\oplus}_f$ для некоторых функций f не является чезарово-наследственным, и его многоугольник может совпадать со всем квадратом $[0,1]\times[0,1]$. При $n=4$ среди всех $2^{16}=65536$ регистров $A^{\oplus}_f$ имеется 332, многоугольник которых является полным квадратом. Получена (*Утверждение 4.21*) дваждыэкспоненциальная оценка количества регистров сдвига с внутренним суммированием, обладающих таким свойством.

В **разделе 4.3** рассмотрен случай, когда входная последовательность автомата A_f неизвестна, известны частоты единиц и биграмм «11» в выходной последовательности. Предложен способ построения многоугольников $R^b_{A_f}$, характеризующих частоты знаков и биграмм «11». Приведена процедура идентификации автомата, основанная на подсчете частот единиц и биграмм в выходной последовательности $y^{(0)}, y^{(1)}, ..., y^{(N-1)}$, использующая неравенство

$$\rho\left(\left(\frac{1}{N}\sum_{i=0}^{N-1} y^{(i)}, \frac{1}{N}\sum_{i=0}^{N-1} y^{(i)} y^{(i+1)}\right), R^b_{A_f}\right) \le \frac{n+2}{N+n+1}. \qquad (28)$$

Доказан ряд утверждений о строении многоугольников. Показано, что многоугольник $R^b_{A_f}$ расположен в треугольнике $Conv\{(0,0),(1,1),(1/2,0)\}$. Получена (*Утверждение 4.25*) нижняя оценка количества регистров, многоугольник которых максимален по включению. При $n=4$ среди всех многоугольников имеется ровно 529 различных. Имеется 696 отрезков, среди которых 12 различных, 44590 треугольников, среди которых 82 различных, 18176 четырехугольников, среди которых 282 различных, 1968 пятиугольников, среди которых 135 различных, 104 шестиугольника, среди которых 16 различных.

В **разделе 4.4** рассмотрены двоичные регистры сдвига с обратной связью, задаваемой произвольной булевой функцией. Через $A_{g,f}$ обозначим узел, состоящий из двоичного регистра с накопителем длины n, булевой функцией обратной связи $g(x_1,...,x_n)$ и булевой функцией выходов $f(x_1,...,x_n)$. Множество состояний автомата $A_{g,f}$ есть множество всех n-мерных двоичных векторов, входной и выходной алфавиты – множества $\{0,1\}$. Автомат $A_{g,f}$ функционирует следующим образом: под действием входного символа $a\in\{0,1\}$ осуществляется переход $(a_1,...,a_n)\to(a\oplus g(a_1,...,a_n),a_1,...,a_{n-1})$.

Рассматриваются значковые статистики входной и выходной последовательностей $A_{g,f}$. Проводятся построения, аналогичные описанным выше. Получена (*Утверждение 4.37*) дваждыэкспоненциальная оценка количества регистров, многоугольник которых максимален по включению и совпадает с квадратом $[0,1]\times[0,1]$. При $n=3$ для 65536 регистров сдвига имеется 5979 различных многоугольников. Среди различных многоугольников имеется 620 отрезков, 8908 треугольников, 29794 четырехугольника, 20946 пятиугольников, 4872 шестиугольника, 392 семиугольника, 4 восьмиугольника. Максимальное количество регистров соответствует треугольнику $Conv\{(0,0),(0,1),(1,1/2)\}$ и симметричному ему - по 710 каждому.

В **разделе 4.5** на примере двух классов автоматов предложен подход к оценке эффективности метода идентификации автоматов с помощью многогранников, использующий статистическую модель их распределения. Рассмотрим классы $K_n=\{A_f, f\in F_n\}$ и $K^{\oplus}_n=\{A^{\oplus}_f, f\in F_n\}$. Для точки $\mathbf{z}=(z_1,z_2)$ квадрата $[0,1]\times[0,1]$

определим функции $p_n(\mathbf{z})$ и $p_n^{\oplus}(\mathbf{z})$ как вероятности принадлежности этой точки многоугольнику автомата при его случайном равновероятном выборе из рассматриваемого класса:

$$p_n(\mathbf{z}) = \frac{\left| A \in K_n, \text{для которых } \mathbf{z} \in R_A \right|}{\left| K_n \right|}, \quad p_n^{\oplus}(\mathbf{z}) = \frac{\left| A^{\oplus} \in K_n^{\oplus}, \text{для которых } \mathbf{z} \in R_{A^{\oplus}} \right|}{\left| K_n^{\oplus} \right|}. \quad (29)$$

Величины $p_n(\mathbf{z})$ и $p_n^{\oplus}(\mathbf{z})$ в определенной степени характеризуют «информативность» пары чисел $\mathbf{z} = \frac{1}{N}\left(\sum_{i=0}^{N-1} x^{(i)}, \sum_{i=0}^{N-1} y^{(i)} \right)$, полученной по наблюдениям над входными и выходными последовательностями. Если, в частности, значения исследуемых величин близки к 1, то предложенная процедура не дает никакой полезной для идентификации информации. Если значения близки к $1/2$, то точка z принадлежит многоугольнику примерно каждого второго автомата из рассматриваемого класса, и можно полагать, что данное наблюдение отсекает примерно половину возможных вариантов. Малые значения исследуемых величин позволяют говорить о высокой информативности данного наблюдения.

Вычисления проводились для случая $n = 4$, значения функций подсчитывались в точках вида $\left(\frac{i}{10}, \frac{j}{10} \right)$, $0 \le i, j \le 10$. На Рис.6 приведены изображения построенных поверхностей $p_4(\mathbf{z})$ и $p_4^{\oplus}(\mathbf{z})$ после сглаживания. Ось абсцисс соответствует относительным частотам символа «1» во входной последовательности, ось ординат – в выходной, по оси аппликат отложены полученные значения p_4 и $p_4^{\oplus}$ соответственно.

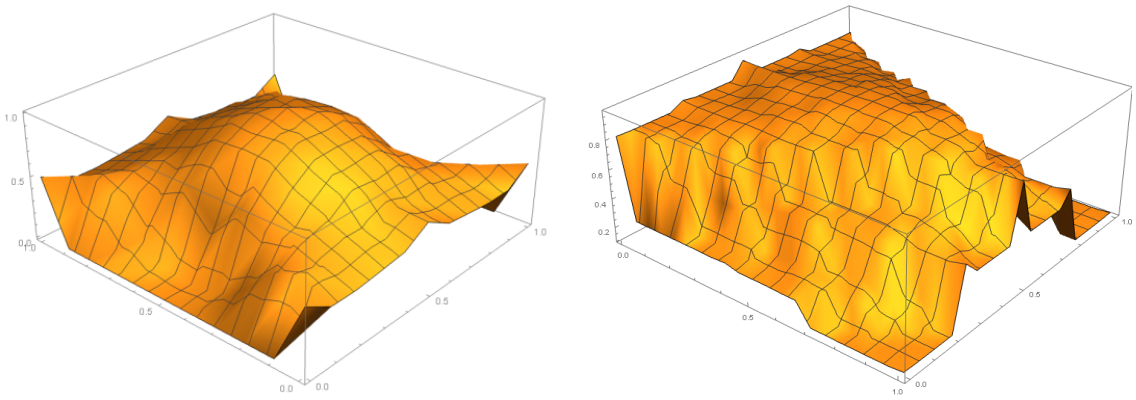


Рис.6. Изображения поверхностей $p_4(\mathbf{z})$ и $p_4^{\oplus}(\mathbf{z})$.

Анализ построенных поверхностей позволяет сделать следующие выводы.

Поверхности $p_4(\mathbf{z})$ и $p_4^{\oplus}(\mathbf{z})$ являются достаточно гладкими, но заметно различаются. Поверхность $p_4^{\oplus}(\mathbf{z})$ обладает симметрией относительно плоскости $z_2 = 1/2$, поверхность $p_4(\mathbf{z})$ обладает, помимо этого, еще и симметрией относительно плоскости $z_1 = 1/2$. Эти симметрии соответствуют доказанным в Главе 4 симметриям многоугольников.

Информативность наблюдения в значительной мере зависит от того, в какую область квадрата $[0,1] \times [0,1]$ попала точка $\frac{1}{N}\left(\sum_{i=0}^{N-1} x^{(i)}, \sum_{i=0}^{N-1} y^{(i)} \right)$. В обоих случаях наименее

информативны точки вблизи $\left(1/2, 1/2\right)$, в которых значения функций $p_4(\mathbf{z})$ и $p_4^{\oplus}(\mathbf{z})$ близки к максимальным.

Глава написана на основании публикаций автора [1, 5, 12, 13, 19, 24, 37, 44, 47, 48].

В **пятой** главе проводится сопоставление методов идентификации конечных автоматов и методов распознавания языка сообщений.

В **разделе 5.1** рассматривается задача распознавания языка сообщения и ее разновидности, особенности распознавания языка текстовых и речевых сообщений, печатных и рукописных текстов, отдельно выделено направление распознавания языка искаженного текста. Проанализирована актуальность различных разновидностей задач распознавания языка, приведены примеры аппаратных решений и патентов в этой области, принадлежащих отечественным и зарубежным фирмам.

В **разделе 5.2** с единой точки зрения проанализированы основные современные подходы к распознаванию языка текстов, речи, печатных и рукописных текстов. Приведены экспериментальные оценки точности распознавания языка в различных условиях. Анализируется связь задачи распознавания языка и задачи идентификации вероятностных автоматов. В отличие от задачи идентификации автоматов, задача распознавания языка неформализуема и поэтому все существующие методы ее решения являются эвристическими, и могут подтверждаться только практическими вычислениями на обучающих и тестовых корпусах. Однако между этими задачами можно проследить аналогии. Когда рассматривается фрагмент текста на неизвестном языке, замысел автора неизвестен, и смысл этого фрагмента текста неясен. Тем не менее, наблюдаемая последовательность символов подчиняется законам словообразования и словоупотребления истинного языка, которые индуцируют статистические закономерности в распределениях комбинаций символов алфавита в тексте. Зная заранее вычисленные статистические закономерности, например, частоты N-грамм, для каждого из проверяемых языков, можно сравнить их с теми, которые наблюдаются в рассматриваемом фрагменте текста, и выбрать язык, для которого это соответствие выполнено лучше всего. Этот подход сопоставим с подходом к идентификации автоматов со случайным входом: начальное состояние автомата неизвестно, но известно, что частоты событий в выходной последовательности автомата подчиняются определенным статистическим соотношениям. Зная заранее вычисленные статистические закономерности, например, математические ожидания частот N-грамм, для выходной последовательности каждого автомата из проверяемого класса, можно сравнить их с теми, которые наблюдаются в выходной последовательности, и указать автомат, для которого это соответствие выполнено лучше всего. При проведении аналогии между задачами идентификации автоматов и языков необходимо учитывать, что в реальных условиях задача распознавания языка имеет множество трудно формализуемых деталей.

В **разделе 5.3** генераторы искусственных текстов рассматриваются как автоматы со случайным входом. Для нескольких генераторов текстов (генератор на основе цепей Маркова, системы Nitrogen и Halogen) построен моделирующий их выход конечный детерминированный автомат со случайным начальным состоянием и случайным входом. Для генератора на основе цепей Маркова со словарем объема n такой автомат имеет n состояний, с использованием метода Биркгофа разложения стохастической матрицы во взвешенную сумму подстановочных получена оценка мощности входного алфавита $n^2 - n + 1$.

Рассмотрены системы генерации мультиязычных текстов. Отмечается, что при анализе текстов, являющихся результатом работы таких систем, задача распознавания языка идентична задаче идентификации конечного автомата в постановке, исследуемой в первых главах диссертации.

В **разделе 5.4** исследуется задача распознавания языка искаженного текста на примерах закрытой задачи для английского, испанского, польского и французского языков с различными типами случайных искажений (редукция алфавита, замена, вставка, выброс символа). Описаны обучающие и тестовые корпуса, способы искажений, применяемые для распознавания языка методы и результаты проведенных экспериментов. Отдельно рассмотрен случай, когда текст подвергается случайным искажениям типа «замена символа». Тексты приводились к латинскому алфавиту без пробелов и знаков препинания, в одном регистре. Задавалось значение $p \in (0,1)$ вероятности замены символа. Решение о замене того или иного символа в тексте принималось последовательно, начиная с первого символа. Замены осуществлялись независимо друг от друга. Замена символа производилась с вероятностью p. Если решение о замене символа принималось, то он заменялся на произвольный другой символ латинского алфавита по равновероятной схеме. Построен дискретный канал без памяти, который моделирует описанные искажения. Вычислена его пропускная способность $C(p)$ как функция от p.

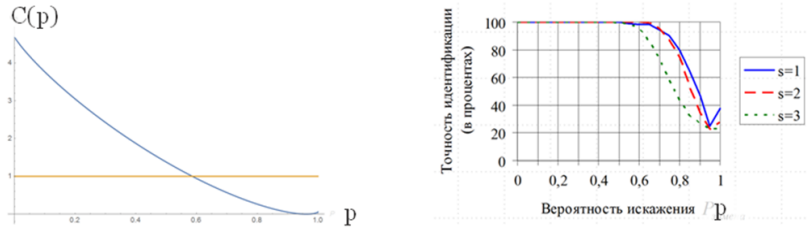


Рис.7. Пропускная способность канала и результаты экспериментов по распознаванию языка.

Построена процедура распознавания языка рассматриваемых текстов, использующая s-граммные статистики на символах, $s = 1, 2, 3$. График пропускной способности $C(p)$ и результаты распознавания языка отрезков текстов длиной 1000 символов с помощью построенной процедуры приведены на Рис.7. Согласно основному теоретико-информационному неравенству, исходный текст можно восстановить из результата его искажения после прохождения через канал связи, при условии $H < C$, где H - энтропия языка исходного текста, а C - величина пропускной способности канала связи. Для текстов большинства европейских языков считается, что $H \approx 1$. Для построенного канала неравенство $1 < C$ приводит к неравенству $p < 0.59$. Это означает, что при вероятностях искажений $p < 0.59$ теоретически возможно восстановление исходного текста по искаженному. При больших значениях p однозначное восстановление исходного текста теоретически невозможно. Результаты экспериментов показывают, что язык текста можно определять с высокой степенью надежности даже в этой ситуации.

Глава написана на основании публикаций автора [4, 17, 18, 22, 23, 25, 27-29, 36, 38, 40, 43].

Работа завершается перечислением перспективных направлений исследований.

Основные результаты работы состоят в следующем.

Разработанные в диссертации методы позволяют решать задачи распознавания и идентификации автоматов по статистическим свойствам наблюдаемых выходной и входной последовательностей в различных предположениях о природе входной последовательности. Разработанные методы не используют перебор множества состояний автомата.

1. Введено и изучено отношение статистической эквивалентности функций выходов конечных автоматов со случайным полиномиальным входом. Получены оценки количества классов эквивалентности и их мощностей. Предложен метод распознавания класса эквивалентности функции выходов автомата. Для двоичного регистра сдвига и значковой статистики выходной последовательности задача распознавания функции выходов сведена к задаче целочисленной минимизации модуля линейной формы при линейных ограничениях. Доказана логарифмическая по числу состояний оценка ее размерности.

2. Разработан метод распознавания функции выходов конечного вероятностного автомата по значениям вероятностей биграмм в его выходной последовательности на основе решения системы квадратичных уравнений, где вектором неизвестных является вектор табличного задания функции выходов, а количество уравнений – это количество известных значений вероятностей биграмм. Рангом соответствующей квадратичной формы является число ненулевых элементов в спектре неориентированной степени графа переходов автомата. Показано, что для автоматов с симметрической матрицей переходов рассматриваемая система допускает линеаризацию. Получены выражения для рангов квадратичных форм для регистров сдвига со случайным входом и двумя вариантами случайного управления движением. Разработанные методы позволили получить спектры и новую верхнюю границу числа независимости неориентированных степеней графа де Брейна, а также вычислить ряд комбинаторных характеристик обобщенных графов де Брейна, в том числе количества остовных деревьев и эйлеровых циклов.

3. Введены и исследованы многогранники автоматов, описывающие совместные статистические свойства входной и выходной последовательностей. Получено представление этих многогранников как выпуклой оболочки множества точек в действительном кубе подходящей размерности. Получены оценки трудоемкости построения многогранников. Предложен метод идентификации автоматов по статистическим свойствам входной и выходной последовательностей, основанный на использовании введенных многогранников. Проведенные построения подтверждены результатами компьютерных экспериментов для следующих классов автоматов: регистры сдвига, регистры сдвига с обратной связью, регистры сдвига с внутренним суммированием для небольших размеров накопителя.

4. Предложен новый принцип классификации дискретных функций, основанный на использовании многоугольников (многогранников) автоматов с заданной функцией выходов. Введено определение автомата, сохраняющего значковые статистические свойства двоичной входной последовательности. В терминах системы псевдобулевых неравенств получен критерий сохранения значковых статистических свойств входной последовательности регистрами сдвига. Доказано, что функция выходов такого регистра является либо координатной, либо нелинейной по всем своим существенным переменным. Получена дважды экспоненциальная по длине накопителя оценка количества таких регистров. Доказано, что в классе регистров сдвига с внутренним суммированием нет автоматов, сохраняющих значковые статистические свойства входной последовательности.

5. Определены и исследованы чезарово-наследственные автоматы, т.е. такие, которые бесконечные последовательности со свойством устойчивости относительных частот встречаемости произвольных слов в растущих начальных отрезках перерабатывают в выходные последовательности с таким же свойством. Получены достаточные условия для того, чтобы автомат обладал и не обладал свойством чезарово-наследственности. Доказано, что в классе обобщенных регистров сдвига чезарово-наследственными являются только классические регистры сдвига. Сформулированы условия, при которых регистр сдвига с внутренним суммированием не является чезарово-наследственным.

6. Экспериментально показано, что n-граммные методы распознавания языка позволяют распознавать язык искаженного текста даже при таких уровнях искажений (вероятность замены символа $p > 0.59$), когда в силу основного теоретико-информационного неравенства восстановление текста невозможно.

7. С помощью компьютерных вычислений построены многоугольники для трех классов автоматов в семействе двоичных регистров сдвига с накопителем размера 4 со всеми возможными функциями выходов. Проведенные построения подтверждают справедливость полученных теоретических результатов. Для регистров сдвига и регистров сдвига с внутренним суммированием построены поверхности, характеризующие множества многоугольников автоматов этих классов «в среднем». Построенные поверхности обладают симметриями, соответствующими доказанным свойствам многоугольников, и демонстрируют существенные различия усредненных характеристик многоугольников автоматов указанных классов.

Публикации по теме диссертации

Монография

1. Мельников С.Ю. Идентификация конечных автоматов на основе метода многогранников. –М. –Ижевск: Ин-т компьютерных иссл., 2013. – 136 с. ISBN 978-5-4344-0108-1.

Статьи, опубликованные в изданиях, включенных в базы Web of Science и Scopus

2. Melnikov S.Yu. The Spectra of Undirected de Bruijn Graphs and an Upper Bound for Their Independence Numbers // Discrete Mathematics and Applications 5(6). — 1995. — Pp. 535–540.
   Мельников С.Ю. Спектры неориентированных графов де Брейна и верхняя граница числа независимости для таких графов // Дискр. матем., том 7, вып.4, 1995, С.140-144.
3. Melnikov S.Yu., Samouylov K.E. The Recognition of the Output Function of a Finite Automaton with Random Input // In: Vishnevskiy V., Kozyrev D. (eds) Distributed Computer and Communication Networks DCCN 2018. Communications in Computer and Information Science, Springer. — 2018. — V. 919. — Pp. 525–531.
4. Iskhakova A., Kruglova S., Melnikov S., Sidorov E. The Approach to Minimize the Impostor Method Errors in the Author Identification Open Problem // Proceedings of the R. Piotrowski's Readings in Language Engineering and Applied Linguistics. S.Petersburg, Russia. — November 27, 2019. — CEUR Workshop Proceedings, V. 2552. — Pp. 60–72.
5. Мельников С.Ю., Самуйлов К.Е. Статистические свойства двоичных неавтономных регистров сдвига с внутренним суммированием // Информатика и ее применение. № 2, 2020, С. 80–85.
6. Melnikov S.Yu., Samouylov K.E. Probabilistic functions and statistical equivalence of binary shift registers with random Markov input // Proceedings of the Workshop on information technology and scientific computing in the framework of the X International Conference Information and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems (ITTMM 2020). Moscow, CEUR Workshop Proceedings. — 2020. — V.2639. — Pp. 93–99.
7. Melnikov S.Yu., Samouylov K.E. On recognition of the shift register output function with a random input for a Markov model of an input sequence // Proceedings of the Workshop on information technology and scientific computing in the framework of the X International Conference Information and Telecommunication Technologies and Mathematical Modeling of High-Tech Systems (ITTMM 2020). Moscow, CEUR Workshop Proceedings. — 2020. — V. 2639. — Pp. 100–107.
8. Melnikov S.Yu., Samouylov K.E. Polyhedra of Finite State Machines and their Use in the Identication Problem // In: Galinina O., Andreev S., Balandin S., Koucheryavy Y. (eds) Internet of Things, Smart Spaces, and Next Generation Networks and Systems. NEW2AN 2020,

ruSMART 2020. Lecture Notes in Computer Science, Springer, Cham. — 2020. — V. 12526. — Pp. 110–121.
9. Melnikov S.Yu., Samouylov K.E. Cesaro Sequences and Cesaro Hereditary Automata //In: Galinina O., Andreev S., Balandin S., Koucheryavy Y. (eds) Internet of Things, Smart Spaces, and Next Generation Networks and Systems. NEW2AN 2020, ruSMART 2020. Lecture Notes in Computer Science, Springer, Cham. — 2020. — V. 12526. — Pp. 259–269.
10. Beschastnyi1 V., Ostrikova D., Melnikov S., Gaidamaka Y. Modelling Multi-connectivity in 5G NR Systems with Mixed Unicast and Multicast Traffic // DCCN 2020. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). — 2020. — V. 12563. — Pp. 52–63.
11. Adamu A., Shorgin V., Melnikov S., Gaidamaka Y. Flexible Random Early Detection Algorithm for Queue Management in Routers Traffic // DCCN 2020. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). — 2020. — V. 12563. — Pp. 196–208.
12. Melnikov S.Yu., Samouylov K.E. Polygons characterizing the joint statistical properties of the input and output sequences of the binary shift register// ICFNDS'20: The 4th International Conference on Future Networks and Distributed Systems, November 2020. — Article No.: 10. — Pp. 1–6.

Статьи, опубликованные в рецензируемых изданиях из Перечня РУДН/ВАК

13. Мельников С.Ю. Многогранники, характеризующие статистические свойства конечных автоматов // Труды по дискретной математике, том 7. — Издательство физико-математической литературы, 2003. — С. 126–137.
14. Мельников С.Ю. О переработке конечными автоматами чезаровских последовательностей // Вестн. Моск. гос. ун-та леса Лесной вестник, № 1(32). — 2004. — С. 169–174.
15. Мельников С.Ю. Регистры сдвига с чезаровским входом // Вестн. Моск. гос. ун-та леса Лесной вестник, № 4(35). — 2004. — С. 200–202.
16. Мельников С.Ю. Об использовании спектров графов автоматов в задаче определения функции выходов по вероятностям биграмм в выходной последовательности. // Вестн. Моск. гос. ун-та леса Лесной вестник, № 2(51). — 2007. — С. 153–158.
17. Кулай А.Ю., Леднов Д.А., Мельников С.Ю. О статистических методах идентификации языка искаженных текстовых и речевых сообщений // Известия ЮФУ. Технические науки, № 8, 2008. — С. 177–183.
18. Кулай А.Ю., Мельников С.Ю. О точности идентификации языка искаженного текста в зависимости от степени искажения // Концептуальный спектр изысканий в современном речеведении (Вестн. Моск. гос. лингвист. ун-та, вып. 575, сер. Языкознание). — М. : ИПК МГЛУ «Рема», 2009. — С. 200–209.
19. Мельников С.Ю. Многоугольники, характеризующие статистические свойства булевых функций в схеме регистра сдвига // Вестник Российского государственного гуманитарного университета, № 12. — 2010. — С. 137–159.
20. Мельников С.Ю. О задаче определения функции выходов автомата со случайным входом по статистике встречаемости слова в выходной последовательности. // Докл. Томск. ун-та сист. упр. и радиоэл. (ТУСУР), № 1(23). — 2011. — С. 107–123.
21. Мельников С.Ю. Неавтономные двоичные регистры сдвига, сохраняющие значковые статистические свойства входной последовательности // Докл. Томск. ун-та сист. упр. и радиоэл. (ТУСУР), № 2(36). — 2015. — С. 86–99.
22. Мельников С.Ю., Пересыпкин В.А. О применении вероятностных моделей языка для обнаружения ошибок в искаженных текстах // Вестник компьютерных и информационных технологий, № 5. — 2016, С.29 – 34.
23. Белозеров А.А., Вахлаков Д.В., Мельников С.Ю., Пересыпкин В.А., Сидоров Е.С. Технологические аспекты построения системы сбора и предобработки корпусов

новостных текстов для создания моделей языка // Известия ЮФУ. Технические науки, № 12. — 2016. — С. 29–42.
24. Мельников С.Ю. Статистические свойства неавтономных обобщенных двоичных регистров сдвига // Докл. Томск. ун-та сист. упр. и радиоэл. (ТУСУР), т. 20, № 1. — 2017. — С. 93–95.
25. Белозеров А.А., Вахлаков Д.В., Мельников С.Ю., Пересыпкин В.А., Скавинская Д.В. Использование эволюционных методов дискретной оптимизации для коррекции искаженных текстов // Вестник компьютерных и информационных технологий, № 12. — 2018. — С. 3–10.
26. Максимовский А.Ю., Мельников С.Ю. Спектральные и комбинаторные свойства редуцированных графов де Брейна // Вопросы кибербезопасности, № 4(28). — 2018. — С. 70–76.
27. Бирин Д.А., Мельников С.Ю., Пересыпкин В.А., Писарев И.А., Цопкало Н.Н. Об эффективности средств коррекции искаженных текстов в зависимости от характера искажений // Известия ЮФУ. Технические науки, № 8. — 2018. — С. 104–114.
28. Германович А.В., Мельников С.Ю., Пересыпкин В.А., Сидоров Е.С., Цопкало Н.Н. Информационные измерения языка. Программная система оценки читаемости искаженных текстов // Известия ЮФУ. Технические науки, № 8. — 2019. — С. 6–18.
29. Вахлаков Д.В., Мельников С.Ю., Пересыпкин В.А. Многоэтапный метод автоматической коррекции искаженных текстов // Известия ЮФУ. Технические науки, № 7. — 2020. — С. 35–45.

Патенты

30. Котов М.А., Леднов Д.А., Мельников С.Ю., Федюкин М.В., Широкова А.М. Система определения параметров линейчатых спектров вокализованных звуков. Патент на полезную модель RUS 78470, 11.06.2008.
31. Котов М.А., Леднов Д.А., Мельников С.Ю., Федюкин М.В., Широкова А.М. Способ определения параметров линейчатых спектров вокализованных звуков и система для его реализации. Патент на изобретение № 2364957. – М. : Российское агентство по патентам и товарным знакам, 20.08.2009.

Свидетельство о государственной регистрации программы для ЭВМ

32. Свидетельство о государственной регистрации программы для ЭВМ № 2017615913. Анализ статистических свойств неавтономных двоичных регистров сдвига и обобщенных регистров сдвига. Правообладатель и автор: Мельников С.Ю. Заявка № 2017613213. Дата поступления 30 марта 2017 г. Дата государственной регистрации в Реестре программ для ЭВМ 26 мая 2017 г.

Публикации в других научных изданиях

33. Мельников С.Ю. О геометрической характеристике статистических свойств конечных сильносвязных автоматов Мили // Обозрение прикладной и промышленной математики, т. 11, вып. 4. — 2004. — С. 880–881.
34. Мельников С.Ю. Спектр неориентированных степеней двоичного графа де Брейна // Обозрение прикладной и промышленной математики, т. 13, в. 4. — 2006. — С. 682–683.
35. Мельников С.Ю. О классе двоичных функций, сохраняющих значковые статистические свойства последовательностей при преобразовании сдвигового типа // Обозрение прикладной и промышленной математики, т. 14, вып. 6. — 2007. — С. 1123–1124.
36. Kulay A.Y., Melnikov S.Y. Different approaches to the garbled text language recognition, using the data compression methods // Proc. XII Intern. Conference “Speech and Computer” 15–18 Oct. 2007, Moscow. — 2007. — Vol. 2. — Pp. 697–701.
37. Мельников С.Ю. О статистических характеристиках обработки двоичных последовательностей регистром сдвига с последовательным суммированием // Обозрение прикладной и промышленной математики, т. 16, вып. 4. — 2009. — С. 682–683.

38. Максимов А.В., Мельников С.Ю., Чавчавадзе Н.М. Тенденции развития методов автоматической идентификации языка речевых и текстовых сообщений // Обозрение прикладной и промышленной математики, т. 16, вып. 2. — 2009. — С. 365–367.
39. Мельников С.Ю. Двоичные обобщенные неавтономные регистры сдвига не наследуют чезаровские свойства входной последовательности // Обозрение прикладной и промышленной математики, т. 17, вып. 6. — 2010. — С. 910–911.
40. Кулай А.Ю., Мельников С.Ю. О моделировании результатов работы фонетического распознавателя с помощью вероятностного автомата, обрабатывающего тексты // Обозрение прикладной и промышленной математики, т. 18, вып. 2. — 2011. — С. 291–292.
41. Мельников С.Ю. О статистической эквивалентности двоичных функций в схеме регистра сдвига с бернуллиевским и марковским входом. // Обозрение прикладной и промышленной математики, т. 18, вып. 2. — 2011. — С. 305–306.
42. Максимовский А.Ю., Мельников С.Ю. О планарности одного подкласса обобщенных графов де Брейна // Обозрение прикладной и промышленной математики, т. 18, вып. 4. — 2011. — С. 647–648.
43. Кулай А.Ю., Мельников С.Ю. Технологические аспекты построения системы идентификации языка текстовых документов // Труды XIV Международной конференции «Речь и компьютер – 2011» (SPECOM'2011) (Казань, 27–30 сентября 2011 г.). — М., 2011. — С. 350–352.
44. Мельников С.Ю. Об идентификации регистров сдвига с троичным входом по значковым статистическим свойствам входной и выходной последовательностей // Обозрение прикладной и промышленной математики, т. 19, вып. 3. — 2012. — С. 412–413.
45. Максимовский А.Ю., Мельников С.Ю. О числе обобщенных в смысле Imase и Itoh регистров сдвига, устанавливаемых постоянным входом в фиксированное состояние // Обозрение прикладной и промышленной математики, т. 22, вып. 5. — 2015. — С. 590–591.
46. Melnikov S.Yu. Stationary Distribution of Random Walk on the Generalized de Bruijn Digraphs // II International Baltic Symposium on Applied and Industrial Mathematics (BISAIM 2016), Svetlogorsk, June 12–18, 2016, Review of Applied and Industrial Mathematics, 5(23) . — 2016. — Pp. 185–186.
47. Melnikov S.Yu., Samouylov K.E. Cesaro-heredity property in the shift register family // Материалы XXIII международной научной конференции DCCN 2020, ред. В. М. Вишневский, К. Е. Самуйлов. — Pp. 751–763.
48. Мельников С.Ю. Подход к оценке информативности метода многогранников в задаче верификации автомата по статистическим свойствам входной и выходной последовательностей // Обозрение прикладной и промышленной математики, т.27, вып. 2. — 2020. — С. 154–157.

Мельников С.Ю. (Россия)

Методы распознавания и идентификации конечных автоматов по статистическим характеристикам выходных и входных последовательностей

В диссертации для задач распознавания и идентификации конечных автоматов по совместным статистическим свойствам их входных и выходных последовательностей проанализированы два случая: когда вероятностная мера на пространстве входных последовательностей задана, и когда известны только частоты встречаемости определенных множеств слов в этих последовательностях. Для обоих случаев разработаны методы решения этих задач, не требующие перебора начальных состояний.

В случае, когда вероятностная мера известна, разработанные методы позволяют распознавать функцию выходов автомата с точностью до класса статистической эквивалентности. В случае, когда вероятностная мера неизвестна, используется предложенная характеризация совместных статистических свойств входной и выходной последовательностей с помощью многогранников автоматов.

Разработанные методы конкретизированы для следующих классов автоматов: регистры сдвига, обобщенные по Imase и Ito регистры сдвига, регистры с внутренним суммированием.

Melnikov S.Yu. (Russia)

Methods for recognition and identification of finite state machines by the statistical properties of their output and output sequences

Two cases were analyzed for the problem of recognizing and identifying finite automata by the joint statistical properties of their input and output sequences: when a probability measure on the space of input sequences is given and when only the frequencies of occurrence of certain sets of words in these sequences are known. In the case when the probabilistic measure is known, the developed methods allow us to recognize the output function of an automaton with an accuracy of the statistical equivalence class. In the case when the probability measure is unknown, the proposed characterization of the joint statistical properties of the input and output sequences with the help of automaton polyhedra is used.

Estimates of computational complexity are obtained and the efficiency of the developed algorithms for the following classes of automata is investigated: shift registers, generalized shift registers, registers with internal XOR.