

Министерство образования и науки Российской Федерации  
Государственное образовательное учреждение  
высшего профессионального образования  
САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ  
ЛЕСОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

---

*Кафедра Информационных систем и технологий*

ИНФОРМАЦИОННЫЕ СИСТЕМЫ  
И ТЕХНОЛОГИИ:  
ТЕОРИЯ И ПРАКТИКА

Сборник научных трудов

Выпуск 5

Санкт-Петербург  
2013

Рассмотрен и рекомендован к изданию  
Ученым советом лесотехнического факультета  
Санкт-Петербургского государственного лесотехнического университета  
имени С.М.Кирова  
« \_\_\_\_\_ » \_\_\_\_\_ 2013 г.

Редакционная коллегия

А.М. Заяц, кандидат технических наук, профессор (отв. редактор)  
М.А. Шубина, кандидат технических наук, доцент (отв. секретарь)  
И.В. Панфилов, доктор технических наук, профессор

Составитель

**М.А. Шубина**, кандидат технических наук, доцент (отв. секретарь)

Рецензент

Доктор технических наук, профессор **В.И. Николаев**  
(Санкт-Петербургский филиал ОАО  
«Российские телекоммуникационные сети»)

УДК 630

**Информационные системы и технологии: теория и практика: сб.**  
Научн. Тр. Вып.5./отв. ред. А.М.Заяц.-СПб.: СПбГЛТУ, 2013. – 55 с.

ISBN 978-5-9239-0163-4

Сборник подготовлен по материалам кафедры вуза, представленным на научно-технической конференции лесотехнического факультета СПбГЛТУ в январе 2012 года, и практических работ, выполненных ее сотрудниками

Темплан 2013 г. Изд.№  
ISBN 978-5-9239-0163-4

@Санкт-Петербургский государственный лесотехнический университет (СПбГЛТУ), 2013



*45 – ЛЕТИЮ СО ДНЯ ОСНОВАНИЯ  
КАФЕДРЫ ИНФОРМАЦИОННЫХ СИСТЕМ И ТЕХНОЛОГИЙ  
ПОСВЯЩАЕТСЯ*

**А.М.Заяц**, кандидат технических наук, профессор

**ИТОГИ НАУЧНО – ИССЛЕДОВАТЕЛЬСКОЙ РАБОТЫ  
ЗА 2012 ГОД**

Пять лет, в 210 летней истории нашего университета, срок небольшой, пять лет для кафедры - значимый период в ее развитии.

Значимость любого времени определяется тем следом в истории, который оставил коллектив своими делами.

Коллектив кафедры Информационных систем и технологий в 2012 году продолжил проведение исследований в рамках основного направления «Информационные системы и технологии: теория и практика», часть из которых ориентирована на разработку научно - методических трудов, направленных на применение современных информационно – образовательных технологий в учебном процессе таких как:

1. Использование компьютерных технологий и математических пакетов в решении оптимизационных задач лесопромышленного комплекса и управления лесным хозяйством;
2. Применение мультимедийных, Internet\Intranet и WEB - технологий в теоретической и практической подготовке студентов;
3. Разработка научных подходов в методике преподавания учебных дисциплин в рамках направления подготовки «Информационные системы и технологии».

В отчетом году под руководством преподавателей кафедры выполнялись 7 дипломных проектов (работ), 9 квалификационных работ бакалавров и 6 магистерских диссертаций. Итоги государственной аттестации представлены в таблице.

Направление, специальность		Итоговая аттестация	2011/2012_уч._год					
Шифр	Наименование		отл.		хор.		удов.	
			абс	%	абс	%	абс	%
230200.62	Информационные системы	ВКР	9	100	-	-	-	-
230201.65	Информационные системы и технологии	ДП (ДР)	3	44	2	28	2	29
30200.68	Информационные системы и технологии	МД	4	67	2	33	-	-
230200.62	Информационные системы	Итоговый экзамен	6	66	3	34	-	-
230200.68	Информационные системы и технологии		4	57	3	43		

В апреле 2012 года состоялся первый выпуск и успешная аккредитация магистров по направлению подготовки «Информационные системы и технологии» (Приказ Министерства образования и науки РФ № 618 от 25.05.2012 г.).

За отчетный период на кафедре изданы следующие учебно-методические материалы

<b>№ п/п</b>	<b>Название дисциплины</b>	<b>Название учебных, учебно-методических материалов</b>	<b>Изд-во, год</b>	<b>Авторы</b>
1	Методы и средства проектирования инф. систем и технологий, Проектирование информационных систем , Администрирование ИС	Проектирование информационных систем . Администрирование MS SQL Server. Мет. указ.	Изд-во СПбГЛТА 2012 г.	<b>Пресняков В.А.</b>
2	Методы и средства проектирования инф. систем и технологий	Проектирование информационных систем . Основы SQL . Мет. указ.	Изд-во СПбГЛТУ 2012 г.	<b>Пресняков В.А. Васильев Н.П.</b>
3	Научно-исследовательская практика магистров	Научно-исследовательская практика: Методические рекомендации по проведению практики магистров направления 230400 «Информационные системы и технологии». Мет. указ.	Изд-во СПбГЛТУ 2012 г.	<b>Зяяц А.М.</b>
4	Научно-педагогическая практика магистров	Научно-педагогическая практика: Методические рекомендации по проведению практики магистров направления 230400 «Информационные системы и технологии. Мет. указ.	Изд-во СПбГЛТУ 2012 г.	<b>Зяяц А.М.</b>
5	Мультимедиа технологии	Мультимедийные технологии. Учебное пособие	Изд-во СЛИ 2012 г.	<b>Жук Ю.А.</b>
6	Архитектура ЭВМ и систем	Архитектура ЭВМ и информационных систем. Структурная организация. Учебное пособие	Сдано в ИПО	<b>Панфилов И.В., Зяяц А.М.</b>
7	Архитектура современных ИС	Архитектура информационных систем. Функциональная организация. Учебное пособие	Сдано в ИПО	<b>Панфилов И.В., Зяяц А.М.</b>

На кафедре активно используются возможности информационной системы «Планы ВПО» (ММИС г. Шахты) и ее модули «Учебная нагрузка» и «План работы кафедры», позволяющие автоматизировать процесс распределения учебной нагрузки между преподавателями, формирование и заполнение индивидуальных планов преподавателей.

В отчетном году на кафедре проводились НИР фундаментального и прикладного характера в рамках госбюджетной работы «Информационные системы и технологии: теория и практика», инициативной НИР «Инновационный учебно – методический Web – ориентированный комплекс для профессиональной подготовки студентов лесных ВУЗов», хоздоговорной работы «Разработка алгоритмического обеспечения отказоустойчивого прибора комплексирования сигналов навигационно-измерительных датчиков скоростных судов (отказоустойчивого навигационного сервера)» сумма договора 750 тысяч рублей.

На кафедре организована научная работа студентов в рамках постоянно действующего кафедрального научного общества студентов (Руководитель ст. преподаватель Смелова Н.Б.). Общее количество студентов, участвующих в НИРС в 2012 г. составило **7 чел. (21%)**.

По результатам НИР в этом году получено шесть свидетельств о государственной регистрации Программ ЭВМ:

1. «Мультимедийное web–приложение мониторинга, прогнозирования и принятия решения по тушению лесных пожаров» - авторы Логачев А.А., Заяц А.М. Свидетельство о регистрации № 2012618909 от 02.10.2012

2. «Программа одноклассовой классификации на основе обобщенных робастных моделей засорения авторы» - авторы Жук Ю.А., Уткин Л.В. Свидетельство о регистрации: № 2012616682. Дата выдачи: 25.07.2012

3. «Система бинарной классификации при интервальных наблюдениях и генерации элементов обучающей выборки» - авторы Свидетельство о регистрации: № 2012616681. Дата выдачи: 25.07.2012

4. «Программа робастной классификации модифицированным бустингом на основе весового метода опорных векторов» - авторы Жук Ю.А., Уткин Л.В. Свидетельство о регистрации №2012660568 от 23.11.2012:

5. «Программа классификации на основе ограниченного бустинга с использованием метода ближайших соседей» - авторы Жук Ю.А., Уткин Л.В., Егоров А. А. Свидетельство о регистрации № 2012660584 от 23.11.2012

6. «Система контроля знаний студентов SerpTest v1.6» - авторы Хабаров С.П., Заяц А.М., Гоголевский А.С. № 2013610088 от 09.01.2013.



На Всероссийские (региональные), смотры-конкурсы представлялось пять работ, получено четыре награды.

1. Открытый конкурс на право получения грантов студентами и аспирантами вузов и академических институтов, расположенных на территории Санкт-Петербурга в 2011 г. Разработка методов и моделей классификации различных объектов лесных систем при ограниченной обучающей выборке (автор - Гоголевский А.С., руководитель проф. Уткин Л.В.) - размер гранта -50000 руб.,

2. Открытый конкурс на право получения грантов для студентов вузов, расположенных на территории Санкт-Петербурга, аспирантов вузов, отраслевых и академических институтов, расположенных на территории Санкт-Петербурга в 2012 г. Интеллектуальная система принятия решений для обнаружения аномального поведения на основе одноклассовой классификации (автор - Гоголевский А.С., руководитель проф. Уткин Л.В.) - размер гранта: 50000 руб.,

3. Диплом № 71/11 НР победителя конкурса на лучшую научную публикацию 2012г. среди студентов, аспирантов, молодых ученых и специалистов СПбГЛТУ «Мультимедийное web–приложение мониторинга, прогнозирования и принятия решения по тушению лесных пожаров». (автор Логачев А.А., руководитель проф. Заяц А.М.)

4. Диплом № 71/15 НР победителя конкурса на лучшую научную публикацию 2012г. среди студентов, аспирантов, молодых ученых и специалистов СПбГЛТУ «Использование бинарной классификации для идентификации биологических видов на основе данных генного маркирования». (автор Мальцев Е. А., руководитель доц. Жук Ю.А..

5. Заявка на участие в конкурсе работ на соискание премий Правительства РФ в области образования в 2012 году. Информационно – образовательная экспертно – консалтинговая платформа «Образование, лес, бизнес» (ILES v1.1) Руководитель: Заяц А.М. Исполнители: Хабаров С.П., Гоголевский А.С., студенты Логачев А.А., Тюрин А.С.

Преподаватели кафедры и студенты направления подготовки в отчетном периоде участвовали в следующих международных конференциях

1. Заяц А.М., Логачев А. А. «Мультимедийное web – приложение мониторинга, прогнозирования и принятия решения по тушению лесных пожаров», III Международная научно-практическая конференция „Системный анализ. Информатика. Управление”-Запорожье .:Классический приватный университет, Інститут інформаційних та соціальних технологій, 2012 г.

2. Заяц А.М., Тюрин А.С.«Веб приложение для взаимодействия с базой данных арендаторов лесных участков». III Международная научно-практическая конференция „Системный анализ. Информатика. Управление”-Запорожье .:Классический приватный университет, Інститут інформаційних та соціальних технологій, 2012 г.

По итогам НИР за 2012 год коллектив кафедры занял **6** место из **49** по результатам автоматизированной рейтинговой оценки кафедр ЛТУ.

В рамках плана мероприятий на кафедре была организована и проведена очередная научно – практическая конференция *«Информационные системы и технологии: теория и практика»*, материалы докладов которой представлены в данном сборнике.

**А.М.Заяц**, кандидат технических наук, профессор,  
**А.А.Логачев**, аспирант

### **ПОСТАНОВКА ЗАДАЧИ ВЫБОРА ВАРИАНТА ТУШЕНИЯ ЛЕСНЫХ ПОЖАРОВ И МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ДЛЯ ЕЕ РЕАЛИЗАЦИИ**

Пожары являются основным негативным фактором воздействия на лесные экосистемы. Согласно статистике общего повреждения площадей лесного фонда, на долю пожаров приходится около 70% погибших лесов. Из-за лесных пожаров происходит эрозия почв, снижение защитных и водоохранных свойств леса, нарушаются углеродные циклы, уничтожается лесная фауна, а также появляется угроза для населенных пунктов, сельскохозяйственных пастбищ и угодий.

Ежегодная сложная пожароопасная обстановка независимо от метеоусловий приводит к большому количеству пожаров на территории Российской Федерации, приносящих значительный ущерб.

В качестве мер по борьбе с лесными пожарами выделяют превентивные меры, мониторинг и непосредственно само тушение огня. Под превентивными мерами будем понимать комплекс мероприятий, направленных на снижение пожароопасности лесных участков, а также создание условий, способствующих более эффективной борьбе с лесными пожарами. Подобными мерами могут быть рытье противопожарных барьеров, строительство вышек, проведение санитарных рубок и очистка вырубок.

Превентивные меры тесно связаны с прогнозированием вероятности возникновения пожара, поскольку выполнить все возможные противопожарные действия не представляется возможным в связи с ограниченностью финансовых, технических и людских ресурсов. По этой причине необходимы методы прогнозирования вероятности возникновения пожаров и оценки влияния характеристик насаждения на его пожароопасность для принятия решения по проведению превентивных мер на каждом лесном участке.



Обычно быстрое развитие пожара не позволяет провести тщательный, а значит занимающий значительное время, анализ для принятия мер по его ликвидации. Решение в таких ситуациях необходимо принимать оперативно в условиях ограниченного времени. Случайный характер поведения пожара как процесса не всегда позволяет принять правильное решение по его тушению.

В этих условиях целесообразно применение интеллектуальных информационных технологий, базирующихся на математических моделях.

Здесь возможность автоматизированной работы с большим объемом данных, легкость в реализации вычислений и определении различных показателей являются крайне эффективными для обоснованного и оперативного принятия решения в чрезвычайных ситуациях.

Анализ показывает, что для прогнозирования возникновения лесных пожаров и построения математических моделей можно использовать различные подходы:

- построение прогноза на основе информации сети гидрологических станций;
- моделирование с использованием анализа данных антропогенных и природных явлений, а также метеоусловий в рассматриваемом регионе;
- разработка моделей на основе статистического анализа лесных пожаров в рассматриваемой местности, путем построения бинарной логистической регрессии, включающей параметры, связанные с характеристиками насаждений.

Самым простым с точки зрения исходных данных является прогнозирование на основе данных сети гидрологических станций. В основу этого подхода положена гипотеза о взаимосвязи вероятности возникновения лесного пожара и расхода воды в реке и возможности определения пожароопасности, путем сопоставления данных гидрографического стока участка с некоторым критическим значением. При этом следует учитывать случаи, когда перед пожарами расходы воды в реке не соответствовали критическим значениям или, например, во время засухи на некоторых реках может вообще отсутствовать сток. Поэтому для оценки пожароопасности участка необходимо рассматривать не только минимальный сток, но и кривые спада гидрографов стока. При длительном периоде засухи кривые спада становятся более крутыми нежели обычно, а вогнутая форма кривой может смениться на выпуклую. Обычно кривые спада гидрографов стока прерываются небольшими дождями или моросью. Внезапное прекращение уменьшение стока при засухе и его остановка на некотором минимальном значении, говорит о том, что сток иссяк и деревья получают воду только их глубоких родниковых стоков артезианских вод. Что является предпосылкой начала критического пожароопасного периода, который не всегда может быть прерван даже сезоном дождей.

Основным преимуществом использования гидрологических данных является то, что значения спада точны, их легко получить, и они существенны для всего водосбора. Гидрологические данные характеризуют информацию по водосбору в среднем, исключая случайные колебания. Ошибка в оценке спада гидрографа стока в один день не влияет на его определение в последующие дни.

Одним из недостатков данного подхода является возможность его применения только на территориях, где проводятся измерения расходов воды местного водосбора. Поэтому данный подход является эффективным только в качестве дополнительного использования совместно с другими.

Следующий подход основан на детерминировано-вероятностном методе прогнозирования лесных пожаров, учитывающий метеорологические условия, влияние человеческого фактора, а также вероятность возникновения возгорания по причине природных явлений. Вероятность возникновения возгорания определяется тремя составляющими: антропогенной, природной и вероятностью по метеоусловиям. Антропогенная и природная составляющие вычисляются как произведение соответствующих вероятностей: вероятность антропогенного явления на вероятность возникновения пожара из-за него; вероятность молниевых разрядов на вероятность возникновения пожара от молнии.

Вероятности антропогенной нагрузки, возникновения пожара вследствие антропогенной нагрузки, молниевых разрядов и возникновения пожара вследствие молнии определяются через частоту данных событий. То есть как отношение числа случаев возникновения антропогенной нагрузки и числа пожаров от антропогенной нагрузки, а также числа молниевых разрядов и числа пожаров от молниевых разрядов к числу дней в периоде рассматриваемого пожароопасного сезона и к общему числу пожаров соответственно.

Для нахождения вероятности возникновения лесного пожара по причине метеоусловий используется статистическая оценка предельных значений пожарной опасности. Оценка рассчитывается, как отношение индекса пожарной опасности в определенный день к критическому значению пожарной опасности. Индекс критического значения пожарной опасности находится путем сопоставления данных о пожарах за рассматриваемый период и значения индексов пожарной опасности в эти дни.

Исходными данными для прогнозирования являются ежедневные данные наблюдений метеослужб, индексы пожарной опасности, а также данные о пожарах в течение пожароопасного сезона за последний период. Пожары, возникшие по вине населения, сельхозпалов, железных дорог рассматриваются как антропогенные.

Преимуществами данного подхода являются высокая точность, комплексность и относительная универсальность использования в пределах регионов.

Недостатком является необходимость статистических данных по лесным пожарам, причинам их возникновения, а также по метеоусловиям.

Третьим из рассматриваемых является подход прогнозирования пожароопасности на основе анализа статистических данных по лесным пожарам. Отличие его от предыдущего состоит в том, что модель разрабатывается на основе статистических данных по участку, а также по данным его инвентаризации за обширный период времени. Используя данные инвентаризации лесов и статистики лесных пожаров, применяется бинарная логистическая регрессия, определяющая характеристики насаждения, оказывающие значительное влияние на вероятность возникновения пожара. Исходными параметрами данной модели являются высота насаждения, площадь поперечного сечения насаждения, средний диаметр и высота деревьев, а также доля лиственных пород.

Достоинствами такого подхода оказались простота и доступность конечных моделей для использования, а также известность исходных данных, являющихся характеристиками непосредственно насаждения.

Недостатком является достаточно жесткая привязанность использования моделей к той области, на анализе характеристик которой она и была создана, а также ее ограниченность по времени, поскольку с течением времени из-за изменения климатических и метеорологических условий точность модели может быть существенно снижена.

Для третьего подхода также требуется статистическая база лесных пожаров, а также данных об инвентаризации лесов. По этой причине применение второго и третьего подхода в условиях отсутствия или недостатка статистической информации не является должным образом эффективным.

Недостаточный объем статистических данных может быть компенсирован привлечением экспертов, оценивающих ситуацию с точки зрения своих знаний и опыта, совместное использование с другими методами прогнозирования, использование дополнительных неструктурированных данных.

Пусть варианты тушения пожара определяются совокупностью показателей принимающих различные значения во множестве вариантов.

Задачу выбора “наилучшего” с меньшим ущербом (риском) варианта из множества вариантов тушения пожара (построенного на основе их классификации) можно рассматривать как задачу принятия решений.

Рассмотрим эту задачу в формальной постановке. Пусть  $\omega_i$ - номера вариантов, один из которых можно выбрать при тушении пожара на лесном участке, а  $a_i$ - номера вариантов, один из которых необходимо выбрать, чтобы минимизировать риски. Множество  $A = \{a_1, \dots, a_m\}$  образует совокупность альтернатив, а множество  $\Omega = \{\omega_1, \dots, \omega_m\}$  образует набор возможных вариантов. Необходимо выбрать одну альтернативу из конечного множества  $A = \{a_1, \dots, a_m\}$   $m$  возможных альтернатив.

Предположим, что  $\pi(\omega_i)$  - вероятность того, что будет выбран  $i$ -й вариант. Эти вероятности образуют распределение  $\pi = (\pi_1, \dots, \pi_m)$ . При этом  $\sum_{i=1}^m \pi(\omega_i) = 1$  и для каждого  $a_i$  и  $\omega_j$  задана функция полезности  $u(a_i, \omega_j)$ , для которой введем обозначение, как  $u_{ij}$  и которая, характеризуется минимум рисков.

Представленные исходные данные достаточны для постановки задачи принятия решения. При этом наиболее применяемым критерием при указанных исходных данных является максимизация ожидаемой полезности  $E_{\pi} u_i$ , которая определяется как

$$E_{\pi} u_i = \sum_{j=1}^m u_{ij} \pi(\omega_j), \quad i=1, \dots, m.$$

Здесь  $u_i = (u_{i1}, \dots, u_{im})$  - вектор полезностей, соответствующих  $i$ -му действию. Другими словами, необходимо выбрать действие  $a_k$ , такое что  $E_{\pi} u_k = \max_{j=1, \dots, m} E_{\pi} u_j$ .

Во многих ситуациях принятия решения отсутствие достаточной статистики приводит к тому, что вероятности состояний  $\pi_i$  точно не известны. Поэтому задача не может быть решена классическими методами.

Предположим, что вероятности  $\pi_i$  сами по себе являются случайными величинами с функцией распределения  $p(\pi_1, \dots, \pi_m)$ . Найдем ожидаемую полезность с учетом случайности  $\pi_i$ .

Так как вектор вероятностей  $\pi$  является случайным с функцией распределения  $p(\pi_1, \dots, \pi_m)$ , то необходимо найти математическое ожидание  $E_p(E_{\pi} u_i) = E_p(u_i | \pi)$  ожидаемой полезности  $E_{\pi} u_i$  в соответствии с функцией распределения  $p$ . Найдем ожидаемую полезность  $E_p(u_i | \pi)$  по формуле полной вероятности

$$\begin{aligned} E_p(u_i | \pi) &= \int_0^1 \dots \int_0^1 (E_{\pi} u_i) p(\pi_1, \dots, \pi_m) d\pi = \\ &= \int_0^1 \dots \int_0^1 \sum_{j=1}^m \pi_j u_{ij} \cdot p(\pi_1, \dots, \pi_m) d\pi = \sum_{j=1}^m u_{ij} \int_0^1 \dots \int_0^1 \pi_j \cdot p(\pi_1, \dots, \pi_m) d\pi. \end{aligned}$$

Интеграл в последнем выражении есть ни что иное, как математическое ожидание  $E_p \pi_i$  случайной величины  $\pi_i$ . Таким образом, получаем

$$E_p(E_{\pi} u_i) = \sum_{j=1}^m u_{ij} \cdot E_p \pi_j.$$

Следующая задача найти  $E_p \pi_i$ , где  $i=1, \dots, m$ .

Одной из наиболее адекватных моделей случайных вероятностей на конечном наборе событий является модель распределений Дирихле, а при неполной статистической информации – обобщенная модель Дирихле, позволяющая получить оценку при минимальном наборе исходных дан-

ных. В таких случаях не только события, но и их вероятности предполагаются как случайные величины.

Главной особенностью распределения Дирихле является то, что при появлении новых событий и модификации вероятностей, апостериорное распределение остается в том же классе распределений Дирихле, как и априорное распределение. Так при получении новой статистической информации вероятности остаются случайными, но меняется их математическое ожидание. Для учета ограниченности статистических данных, можно использовать обобщенную модель распределений Дирихле, определяемую, как множество всех распределений Дирихле. Так как вместо одного распределения в обобщенной модели используется множество всех распределений Дирихле, то произвольная оценка вероятностей состояний, например, их математическое ожидание или дисперсия, может быть получена как интервал с нижней и верхней границей так, что любая точка этого интервала соответствует одному или нескольким распределениям из множества.

При накоплении статистических данных верхняя и нижняя вероятности событий сходятся, то есть увеличение числа наблюдений снижает неточность.

**В. А. Богатырев** доктор технических наук, профессор,  
**С.М.Алексанков, Д.В. Демидов, С.А. Башкова ,**  
**В.Ф.Беззубов, А.В. Полякова, И.Ю. Голубев,** аспиранты

### **НАДЕЖНОСТЬ ДУБЛИРОВАННОГО ВЫЧИСЛИТЕЛЬНОГО КОМПЛЕКСА ПРИ ОГРАНИЧЕНИИ ДОПУСТИМЫХ ЗАДЕРЖЕК ОБСЛУЖИВАНИЯ**

Высокие требования к надежности компьютерных систем, работающих в контуре управления, приводят к внесению избыточности в структуру последних. Поэтому в подобных системах в качестве базовых вычислительных средств используются *дублированные вычислительные комплексы* (далее – ДВК) [1-7]. В системах компьютерного управления ДВК функционируют в режиме параллельной работы при сравнении результатов в контрольных точках, либо в режиме разделения нагрузки, что позволяет повысить производительность, но может привести к потере результатов вычислений, что связано с возможностью возникновения сбоев в блоках комплекса. В данной статье будет рассмотрен первый вариант.

ДВК могут использоваться в условиях с различными требованиями к *максимальному времени ожидания ответа* от него. В таком случае *максимальным временем ожидания ответа от ДВК* будет максимальное время,

за которое должна быть обнаружена и предотвращена утечка без разрушительных последствий. Здесь мы рассмотрим, как ограничения на это время влияют на надёжность дублированных вычислительных комплексов.

Введём следующие термины: *максимальное время выполнения одной задачи*, *время восстановления ДВК*, *время окончания выполнения задачи*, *максимальное время повторного выполнения задачи*. Поясним их также на примере контроля утечки газа. *Максимальное время выполнения одной задачи* будет максимальным временем, за которое ДВК может обнаружить и предотвратить утечку. Для этого нужно потратить время, которое будет и будет *временем восстановления ДВК*. После восстановления нужно завершить начатую задачу или начать её заново. Для этого также тратится время - *время окончания выполнения задачи* или *максимальное время повторного выполнения задачи* соответственно [8].

Предположим следующую ситуацию: во время задачи, комплекс отказал таким образом, что:

- 1) в его памяти остались последние контрольные точки с промежуточными результатами выполнения этой задачи.
- 2) в памяти нет контрольных точек.

Тогда, в первом случае, после восстановления комплекса, можно продолжить выполнять задачу с момента, когда была сохранена последняя контрольная точка. Во втором же случае, задачу придётся выполнять сначала. Исходя из этих случаев отказа и ограничений *на максимальное время ожидания ответа от ДВК*, можно определить три варианта этих ограничений и особенности организации работы и восстановления ДВК после отказов для каждого из этих ограничений.

Рассмотрим эти варианты ограничений:

#### **1) Ограничения средней жёсткости на максимальное время ожидания ответа.**

При работе в условиях с таким ограничением, *максимальное время ожидания ответа от комплекса* не превышает сумму *максимального времени выполнения одной задачи* и *времени восстановления ДВК* после отказа без полной потери промежуточных результатов выполняемой задачи.

То есть, ограничение на время, за которое, например, должна быть обнаружена и предотвращена утечка газа (*максимальное время ожидания ответа от ДВК*) может позволить начать задачу обеспечения безопасности, затем, в случае отказа без потери промежуточных результатов, восстановить комплекс и закончить выполнение задачи, только без повторного выполнения (рис. 1).

Если же произошёл отказ с полной потерей промежуточных результатов, комплексу будет необходимо, после восстановления, начинать задачу (обеспечения безопасности) заново, из-за чего *максимальное время*

ожидания ответа от ДВК будет превышено, что недопустимо в данном варианте ограничений и может повлечь за собой разрушительные последствия (рис. 2).

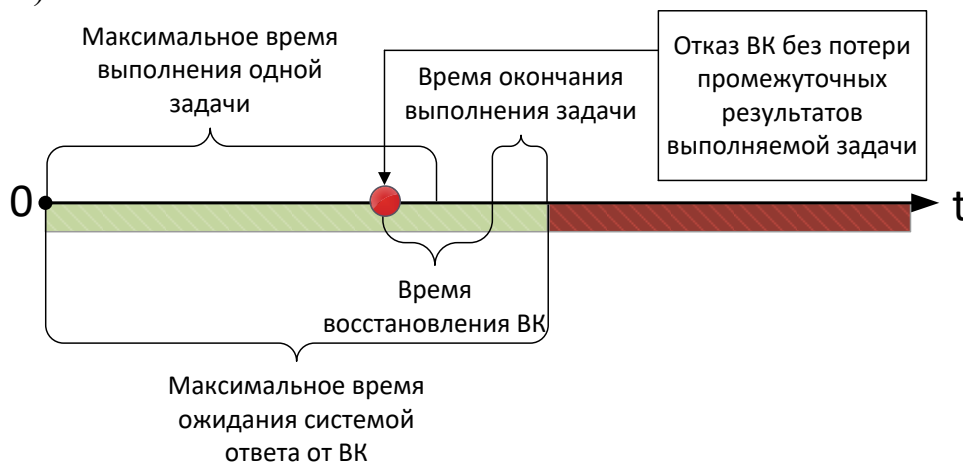


Рис. 1. Временная диаграмма работы ДВК в условиях с ограничением средней жёсткости на максимальное время ожидания ответа после отказа без потери промежуточных результатов выполняемой задачи

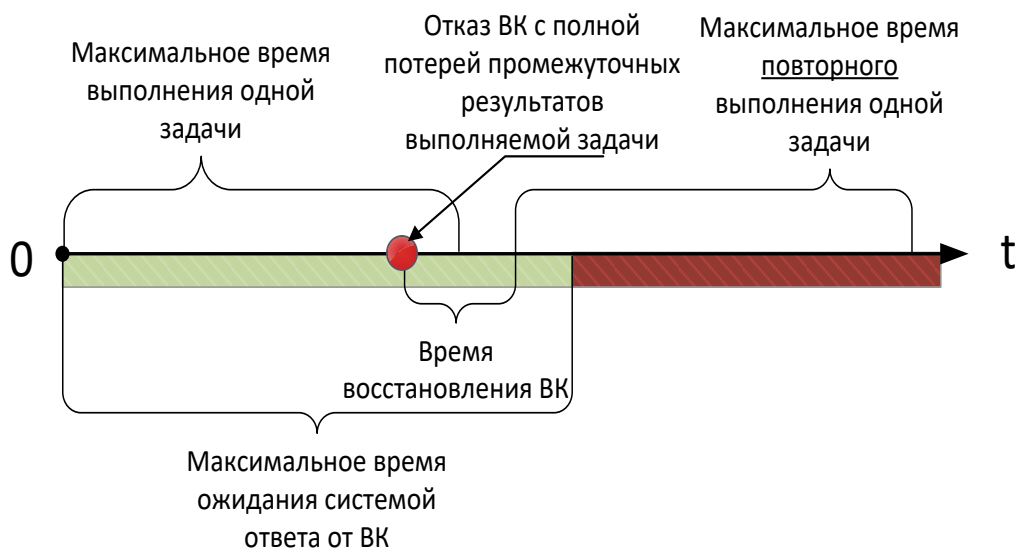


Рис. 2. Временная диаграмма работы ДВК в условиях с ограничением средней жёсткости на максимальное время ожидания после отказа с полной потерей промежуточных результатов выполняемой задачи

## 2) Жёсткое ограничение на максимальное время ожидания ответа от ДВК.

При работе в условиях с таким ограничением, *максимальное время ожидания ответа от комплекса не превышает максимальное время выполнения одной задачи.*

Здесь ограничение на время, за которое, например, должна быть обнаружена и предотвращена утечка газа (*максимальное время ожидания ответа от ДВК*) может позволяет начать и закончить задачу обнаружения

и предотвращения утечки (потратив *максимальное время выполнения одной задачи*), однако если произойдёт какой-либо отказ, то из-за затрат *времени на восстановление, максимальное время ожидания ответа от ДВК* будет превышено (рис. 3).

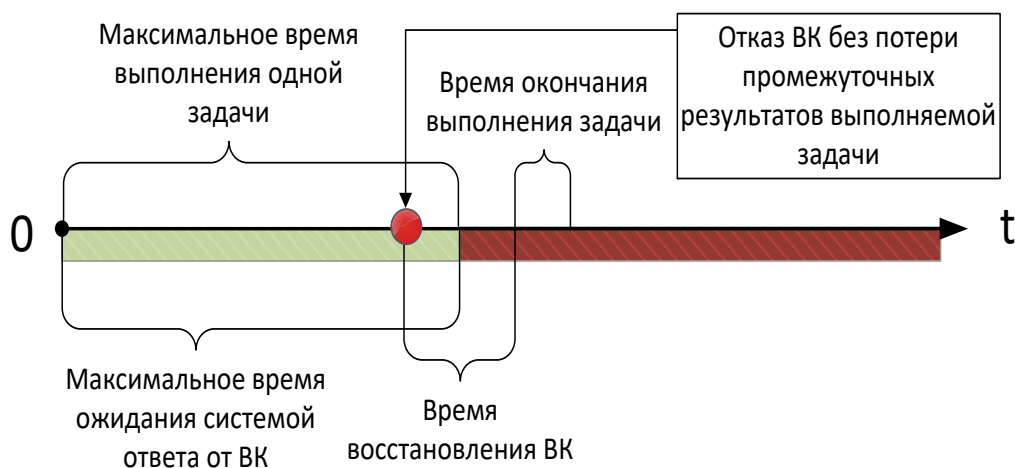


Рис. 3. Временная диаграмма работы ДВК в условиях с жёстко ограниченным временем ожидания ответа после отказа

### 3) Слабое ограничение на максимальное время ожидания ответа от ДВК

При работе в условиях с таким ограничением, *максимальное время ожидания ответа от комплекса* превышает сумму *максимального времени выполнения одной задачи, времени восстановления комплекса* после любого отказа и *максимального времени повторного выполнения одной задачи*.

То есть, в данном случае ограничение на время, за которое, например, должна быть обнаружена и предотвращена утечка газа (*максимальное время ожидания ответа от ДВК*) позволяет начать задачу обнаружения и предотвращения утечки и, затем, в случае отказа с потерей промежуточных результатов, восстановить комплекс, начать повторно и закончить задачу (рис. 4).



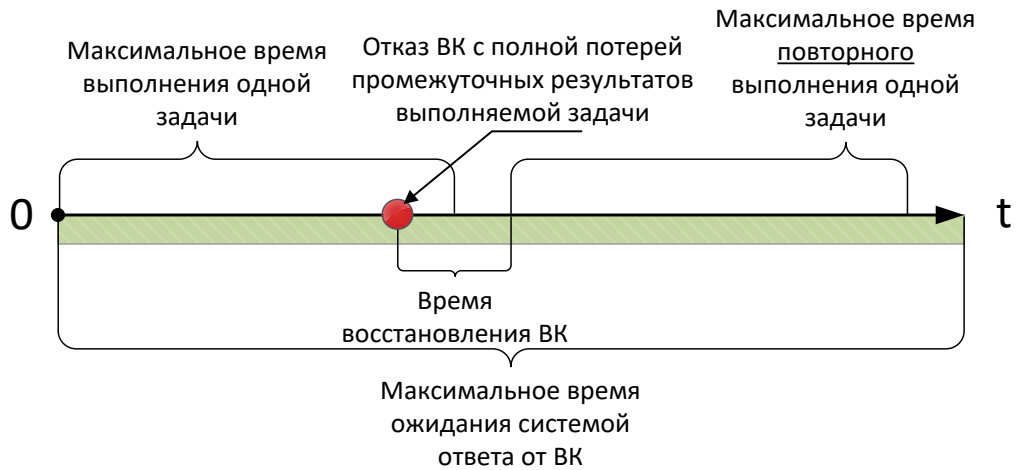


Рис. 4. Временная диаграмма для работы ДВК в условиях со слабо ограниченным временем ожидания ответа после отказа с полной потерей промежуточных результатов выполняемой задачи.

Рассмотрим структуру ДВК (рис. 5). Он состоит из двух ЭВМ и устройства связи (адаптера). Каждая ЭВМ состоит из центрального процессора и модуля памяти.

Адаптер выполняет следующие функции:

- Сравнение результатов обеих ЭВМ;
- Обеспечение прямого доступа к памяти между обеими ЭВМ;
- Синхронизация восстановленного блока памяти с рабочим (восстановление содержимого памяти).

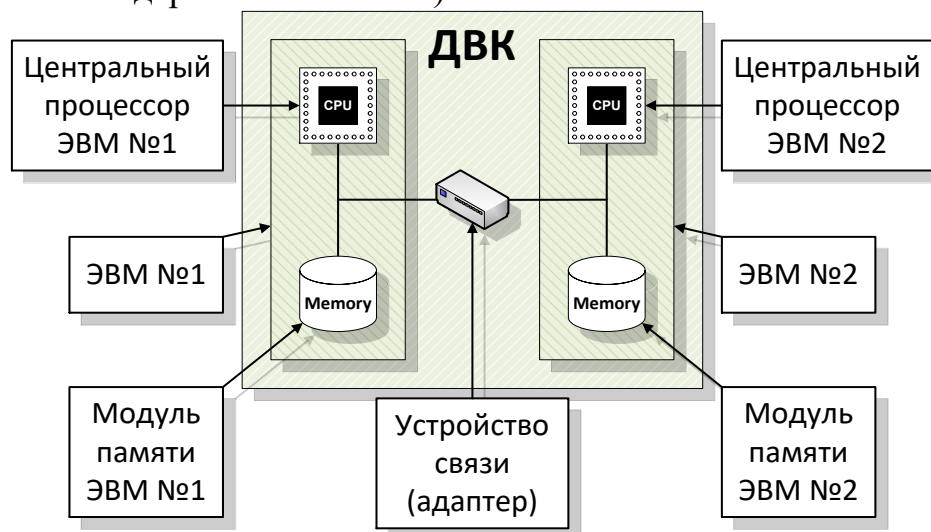


Рис. 5. Схема ДВК

Для каждого из трёх рассмотренных вариантов ограничений на максимальное время ожидания ответа существуют свои особенности организации работы и восстановления ДВК. Рассмотрим их:

**1) Особенности организации работы и восстановления ДВК после отказов в условиях с ограничением средней жесткости на время ожидания ответа.**

Так как при этом ограничении недопустима потеря промежуточных результатов, то отказ обоих модулей памяти либо последовательный отказ адаптера и затем модуля памяти (из-за чего в оставшемся модуле памяти не будет сохранена последняя контрольная точка), переводят комплекс в невосстанавливаемое состояние.

Также нужно упомянуть, что при отказе адаптера работать продолжает только одна ЭВМ, вторая же отключается, т.к. не синхронизируется (по контрольным точкам) с первой, из-за чего она не сможет продолжить выполнение начатой первой ЭВМ задачи.

**2) Особенности организации работы и восстановления ДВК после отказов в условиях с жестким ограничением на время ожидания ответа.**

В этом варианте ограничений недопустимы затраты времени на восстановление. Поэтому любой отказ комплекса (отказ двух процессоров, двух модулей памяти и т.д.), приведёт к переходу в невосстанавливаемое состояние.

Работа комплекса при отказе адаптера организуется аналогично варианту ограничений, рассмотренном выше.

**3) Особенности организации работы и восстановления ДВК после отказов в условиях со слабым ограничением на время ожидания ответа.**

В этом варианте допустимы как потеря промежуточных результатов, так и любое восстановление из нерабочего состояния. Это означает, что при таком ограничении нет невосстанавливаемых состояний.

Из-за того, что потеря промежуточных результатов допустима, при отказе адаптера остаётся работать только одна ЭВМ, а другая переходит в режим ожидания, готовая в случае отказа первой, начать выполнение задачи сначала.

Для оценки надёжности ДВК использовались марковские модели. В соответствии с рассмотренными особенностями организации работы и восстановления комплекса, были составлены графы для каждого из видов ограничений (рис. 6,7,8), в которых каждая вершина обозначает состояние ДВК и определяет вышедшие из строя блоки комплекса (память, процессор, адаптер).

Рёбра графа представляют переходы между состояниями, при отказах блоков комплекса или восстановлениях, с некоторой интенсивностью отказа или восстановления соответственно.

Жирной окантовкой выделены невосстанавливаемые состояния. Над стрелками написаны интенсивности: отказа процессора -  $\lambda_0$ , памяти -  $\lambda_1$ , адаптера -  $\lambda_2$  или восстановления -  $\mu$ .

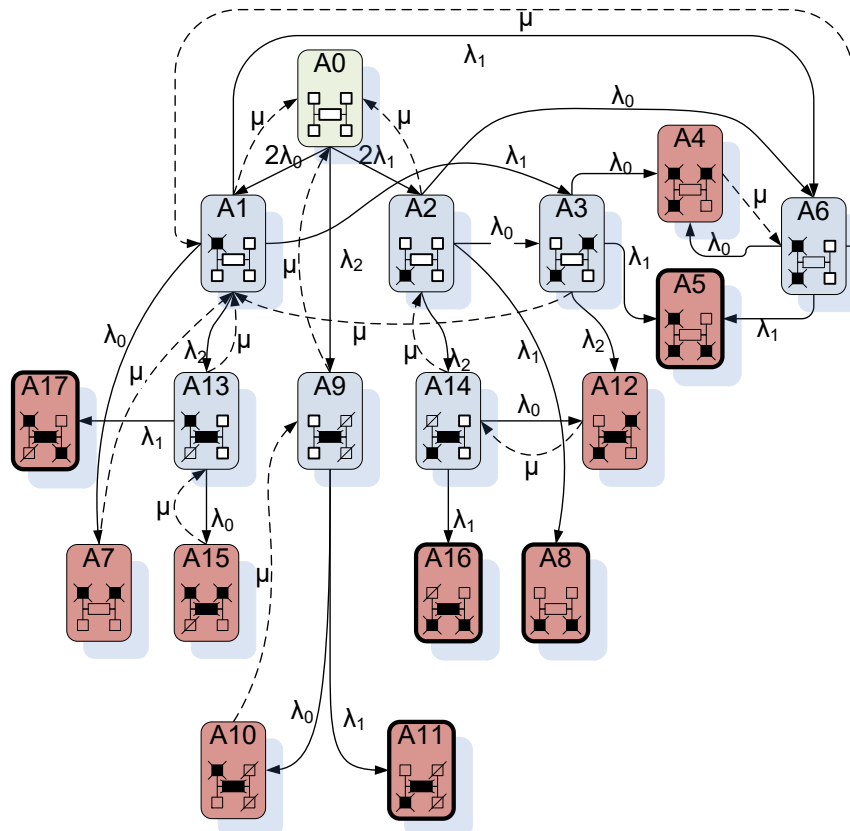


Рис. 6. Граф для ДВК, работающих в условиях с ограничением средней жесткости на максимальное время ожидания ответа

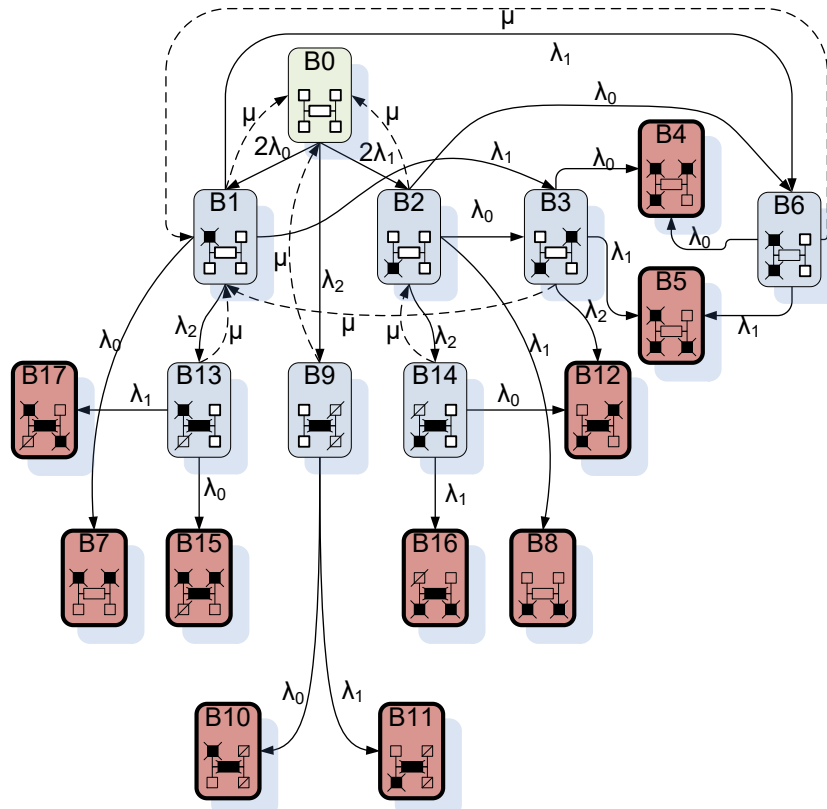


Рис. 7. Граф для ДВК, работающих в условиях с жёстким ограничением на максимальное время ожидания ответа.

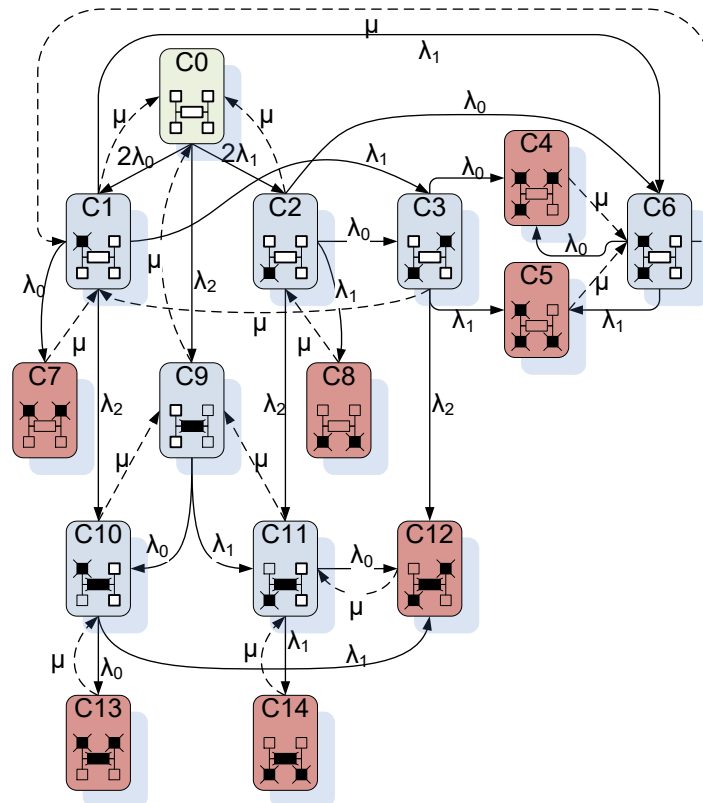


Рис. 8. Граф для ДВК, работающих в условиях со слабым ограничением на максимальное время ожидания ответа

Составив по графам матрицы интенсивностей переходов, можно с их помощью составить и решить дифференциальные уравнения и найти вероятности нахождения комплекса в  $i$ -том состоянии  $p_i(t)$  на произвольный момент времени  $t$ .

В качестве основного показателя надежности выбран *нестационарный коэффициент готовности* -  $K_a(t)$  – вероятность того, что в заданный момент времени  $t$  система находится в работоспособном состоянии. Математически  $K_a(t)$  можно выразить как сумму вероятностей всех работоспособных состояний системы:

$$S_w = \{i : \text{состояние}; i - \text{работоспособно}\}$$

$$K_a(t) = \sum_{i \in S_w} p_i(t)$$

$K_a(t)$  восстанавливаемых систем монотонно падает, приближаясь асимптотически к некоторому пороговому значению.  $K_a(t)$  невосстанавливаемых систем (или частично-восстанавливаемых, в которых имеется хотя бы одно финальное состояние их которого невозможно восстановиться) монотонно падает до нуля (либо асимптотически к нему приближается).

Для комплексов, работающих в каждом из трёх рассмотренных выше условий с ограничениями на максимальное время ожидания ответа нестационарный коэффициент готовности определяется как :

$$K_{a1}(t) = p_{A0}(t) + p_{A1}(t) + p_{A2}(t) + p_{A3}(t) + p_{A6}(t) + p_{A9}(t) + p_{A13}(t) + p_{A14}(t);$$

$$K_{a2}(t) = p_{B0}(t) + p_{B1}(t) + p_{B2}(t) + p_{B3}(t) + p_{B6}(t) + p_{B9}(t) + p_{B13}(t) + p_{B14}(t);$$

$$K_{a3}(t) = p_{C0}(t) + p_{C1}(t) + p_{C2}(t) + p_{C3}(t) + p_{C6}(t) + p_{C9}(t) + p_{C10}(t) + p_{C11}(t).$$

$K_{a1}$  – нестационарный коэффициент готовности для ДВК, работающих в условиях с ограничением *средней жёсткости* на время ожидания ответа;

$K_{a2}$  – нестационарный коэффициент готовности для ДВК, работающих в условиях с *жёстким* ограничением на время ожидания ответа;

$K_{a3}$  – нестационарный коэффициент готовности для ДВК, работающих в условиях со *слабым* ограничением на время ожидания ответа.

Для расчётов были взяты следующие интенсивности:

Процессора ( $\lambda_0$ ) – 0,0008 (1/ч);

Блока памяти ( $\lambda_1$ ) – 0,0005 (1/ч);

Адаптера ( $\lambda_2$ ) – 0,0003 (1/ч);

Интенсивность восстановления ( $\mu$ ) – 1 (1/ч).

Расчёты нестационарного коэффициента готовности ДВК, для рассматриваемых требований к максимальному времени ожидания ответа были проведены в программе Mathcad 15. Их результаты представлены на графике (рис. 9).

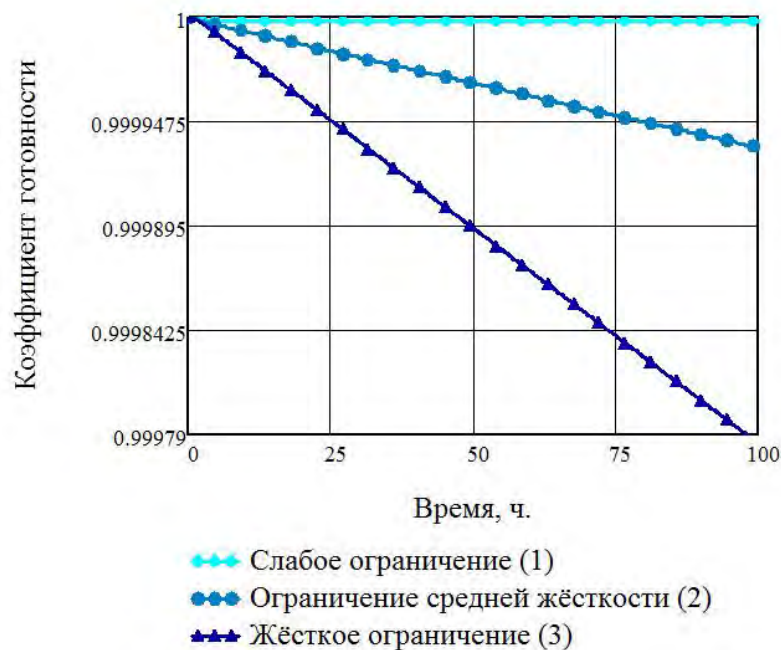


Рис. 9. График нестационарного коэффициента готовности

По графику видно, что нестационарный коэффициент готовности при работе условиях 3 падает быстрее всего по времени, так как здесь больше всего невосстанавливаемых состояний из-за того что каждое нерабочее состояние – невосстанавливаемое. При работе условиях 2 нестационарный коэффициент готовности падает чуть меньше, так как невосстанавливаемых состояний меньше, чем при работе в условиях 2 (невосстанавливаемые только состояния с потерей промежуточных результатов выполняемой задачи). При работе в условиях 1 падение происходит только в самом начале времени, потом нестационарный коэффициент готовности становится постоянным во времени, так как в этой системе нет невосстанавливаемых состояний и всегда можно перейти в рабочее состояние.

Таким образом, видно, что самая высокая надёжность у того ДВК, который работает в условиях 1 (со слабым ограничением времени ожидания ответа), так как у него самый высокий нестационарный коэффициент готовности по времени.

Также в следующей таблице (табл. 1), приведено время, за которое обеспечиваются заданные коэффициенты готовности, при заданных интенсивностях отказов и восстановления.

Из таблицы видно, что каждый из рассмотренных коэффициентов нестационарной готовности обеспечивается дольше всего для систем со слабо ограниченным временем ожидания. А коэффициент нестационарной готовности больше 0,99999 устанавливается после 1 часа 53 минут и поддерживается больше 300000 часов, это более 3 лет и 155 суток.

Нестационарный коэффициент готовности	Наработка (часы)		
	слабое ограничение на время ожидания ответа	ограничение <b>средней жёсткости</b> на время ожидания ответа	<b>жёсткое</b> ограничение временем ожидания ответа
>0,99999	30000	14,068	5,609
	> 3 г. 155 сут.	= 14 ч. 4 мин.	= 5 ч. 36 мин. 30 сек
>0,9999	30000	152,7	47,12
	> 3 г. 155 сут.	= 6 сут. 8 ч. 42 мин.	= 1 сут. 23 ч. 7 мин.
>0,999	30000	1542	462,5
	> 3 г. 155 сут.	= 64 сут. 6 ч.	= 19 сут. 6 ч. 30 мин.
>0,99	30000	15488	4636
	> 3 г. 155 сут.	= 64 г. 280 д.	= 193 сут.

Проведенные исследования могут быть использованы при выборе организации отказоустойчивого вычислительного процесса и прогнозировании надежности ДВК с различными вариантами соотношения времени восстановления и ограничений на выдачу результатов вычислений.

### Библиографический список

1. . Панфилов И.В., Половко А.М. Вычислительные системы. М., «Сов. радио» 1980. С. 304.
2. 25. Шубинский И.Б., Николаев В.И., Колганов С.К., Заяц А.М. Активная защита от отказов управляющих модульных вычислительных систем. СПб. «Наука» 1993. С. 285.
3. Богатырев В.А., Беззубов В.Ф., Котельникова Е.Ю., Богатырев.А.В., Землянухин А.А., Белоусов С.И., Румянцев А.С. Выбор структуры отказоустойчивого двухмашинного вычислительного комплекса.// Информационные системы и технологии: теория и практика. Сборник научных трудов. Выпуск 3. СПб. 2011. С. 14 – 20
4. Богатырев В.А., Башкова С.А., Беззубов В.Ф., Полякова А.В., Котельникова Е.Ю., Голубев И.Ю. Надежность дублированных вычислительных комплексов // Научно технический вестник информационных технологий механики и оптики, № 6(76) 2011г. С.74 – 78.
5. Богатырев В.А., Беззубов В.Ф., Голубев И.Ю. Сравнительный анализ структур отказоустойчивых дублированных вычислительных комплексов. // Информационно – измерительные и управляющие системы. 2011. № 2. С. 8 – 12.
6. Bogatyrev V.A. Exchange of Duplicated Computing Complexes in Fault tolerant Systems // Automatic Control and Computer Sciences. - 2011. - Vol. 46. - N 5. - P. 268–276. - ISSN 0146 -4116
7. Богатырев В.А., Демидов Д.В., Алексанков С.М. Оценка надежности дублированных комплексов с учетом контроля // Materiali VII mezinárodní vedecko-praktická konference Aktuální vymoženosti vědy -2011 Чехия, Прага 27.06.2011-05.07.2011. - Прага: Education and science, 2011. - С. 57-58. - 80 с.
8. Алексанков С.М. Построение модели дублированного вычислительного комплекса при ограниченном времени восстановления вычислительного процесса. // Аннотированный сборник научно-исследовательских выпускных квалификационных работ бакалавров и специалистов. НИУ ИТМО– СПб: НИУ ИТМО, 2011. – 110 с.

**Н.П. Васильев**, кандидат физико-математических наук, доцент  
**С.П.Хабаров**, кандидат технических наук, доцент

## **МЕТОД ИСКЛЮЧЕНИЯ ВЫБРОСОВ В ДАННЫХ, ПОЛУЧЕННЫХ ОТ ИЗМЕРИТЕЛЬНЫХ ПРИБОРОВ И ДАТЧИКОВ.**

### **Введение**

Качество современных информационно-управляющих систем во многом определяется достоверностью результатов, получаемых от измерительных приборов и датчиков. Как правило, измерения, поступающие в систему, содержат помехи двух типов:

- помехи низкого уровня (инструментальные и иные погрешности),
- помехи высокого уровня (выбросы и пропадания сигнала).

Помехи низкого уровня практически всегда присутствуют в системах и определяются статическими и точностными характеристиками датчиков и их измерительных элементов. Для исключения этого класса помех в системах используют различные методы фильтрации. Наибольшее распространение в управляющих системах получили различные модификации фильтров Калмана, в том числе и с учетом аддитивных возмущений, содержащих неизвестную постоянную составляющую [1].

Помехи высокого уровня (аномалии в измерениях) возникают в результате помех в канале связи, сбоев в работе приемной, передающей или измерительной аппаратуры. Они являются результатом изменения внешних воздействий на систему, таких как провалы напряжения питания, резкое изменение вибрационных или температурных параметров. Для исключения этого класса помех используют методы робастной локальной регрессии [2], медианную фильтрацию [3] и широкий набор методик статистического анализа случайных процессов.

Однако перечисленные выше методы напрямую хорошо работают, когда известен весь динамический ряд и для сглаживания или фильтрации некоторой  $x_i$  точки этого ряда может быть выбрано окно  $[x_{i-d}, x_{i+d}]$  произвольной длины  $d$ , к которому может быть применен тот или иной метод фильтрации.

В информационно-управляющих системах реального времени измерения проводятся в определенные моменты времени и последовательно передаются по каналам связи в обрабатывающий узел. Помехи, возникающие в линии связи, приемном и передающем устройствах вносят сложно компенсируемое искажение в принимаемый сигнал. Однако воздействие этих помех в большинстве случаев кратковременно (рис. 1).





Рис. 1. Результаты измерений.

В этих условиях, наиболее важной представляется задача разработки метода выявления подобных ситуаций. Это необходимо для предварительного просеивания исходных измерений, которые далее поступают на фильтр Калмана для компенсации помех низкого уровня.

### Постановка задачи

Рассматривается случайный процесс  $X(t)$ , который представляет собой результат измерения некоторого навигационного параметра. Предполагается, что структура этого процесса следующая:

$$X(t) = x(t) + \varepsilon(t).$$

В этом соотношении  $x(t)$  – истинное значение измеряемого параметра, которое является обычной (неслучайной) функцией времени, а величина  $\varepsilon(t)$  представляет случайную составляющую, которая моделирует искажения (в идеале можно считать белым шумом).

Измерения проводятся в определенные моменты времени:  $t_1, t_2, \dots, t_i, \dots$ . В результате, получается последовательность сечений исходного процесса:

$$X_i = x_i + \varepsilon_i \quad (i = 1, 2, \dots),$$

где  $\varepsilon_i$  – одинаково распределенные независимые нормальные случайные величины с нулевым математическим ожиданием и дисперсией  $\sigma^2$ .

Измерения проводятся последовательно, то есть в каждый  $i$ -ый момент времени процесс представлен выборочной совокупностью предыдущих  $i$  измерений.

Реальная ситуация такова, что на  $i$ -ом шаге возможно измерение, которое явно противоречит данной модели и поэтому должно быть выбраковано (признано выбросом) и в дальнейшем исключено из рассмотрения. Таким образом, требуется определить метод выявления подобных ситуаций для выполнения предварительного просеивания исходных выборочных измерений.

### Подход к исключению выбросов

Метод исключения выбросов основан на кусочно-линейной аппроксимации исходного процесса. Предполагается, что параметр  $x(t)$  является гладкой функцией, и в пределах малых промежутков времени  $\Delta t$  может быть достаточно точно представлен линейной функцией. Более точная формулировка предполагает выполнение для  $\Delta t$  следующего соотношения

(достаточно воспользоваться разложением Тейлора в окрестности центра  $t_c$  этого промежутка  $\Delta t$ ):

$$\left| \frac{1}{2} \frac{d^2 x(\bar{t})}{dt^2} \frac{\Delta t^2}{4} \right| \ll \sigma,$$

где  $\bar{t}$  – некоторая точка в промежутке  $\Delta t$ .

Если из каких-то дополнительных соображений известно ограничение на вторую производную измеряемого параметра:

$$\left| \frac{d^2 x(t)}{dt^2} \right| \leq D,$$

то получаем:

$$\Delta t \ll \sqrt{\frac{8\sigma}{D}}.$$

Это соотношение означает, что ошибка, которую вносит линеаризация параметра  $x(t)$  пренебрежимо мала по сравнению с шумом, искажающим измерение параметра.

В реальных ситуациях величина  $\sigma^2$  может быть определена исходя из характеристик измерительного тракта, предварительных экспериментов, описаний приборов. Величина второй производной фактически определяется инерционностью управляемого объекта и величиной управляющих сил, воздействующих на объект.

Далее предполагается, что в пределах этих малых промежутков времени измерения проводятся достаточно часто, например,  $n$  измерений.

Требуется классифицировать очередное  $i$ -ое измерение: является это измерение выбросом, или не является. Следует подчеркнуть, что известна только предыстория процесса, включающая последнее  $i$ -ое измерение, а не весь процесс в целом. В этом плане задача носит динамический характер. В этом особенность задачи. Если бы была известна вся выборка, то можно было бы воспользоваться многочисленными методами сглаживания [2,3].

В рамках сделанных предположений, в пределах окна, охватывающего последние  $n$  измерений, имеет место следующая линейная модель:

$$X_i = \beta_1 t_i + \beta_0 + \varepsilon_i \quad (i = 1, \dots, n).$$

Если вспомнить разложение Тейлора, о котором шла речь выше, то здесь  $\beta_1$  это производная параметра  $x(t)$  в центре  $t_c$  окна линеаризации  $\Delta t$ , а  $\beta_0$  это значение этого параметра в центре  $t_c$ . Как уже было сказано в пределах окна  $\Delta t$  помещается  $n$  измерений, которые без ограничения общности рассуждений, можно пронумеровать от 1 до  $n$ .

Для оценки коэффициентов представленной линейной модели воспользуемся методами линейного регрессионного анализа [5]. В дальнейшем оценки коэффициентов в отличие от их истинных значений  $\beta_0, \beta_1$  будем обозначать соответственно через  $b_0, b_1$ .

Имеют место следующие известные соотношения:

$$b_1 = \frac{\sum (t_i - \bar{t}) \sum (X_i - \bar{X})}{\sum (t_i - \bar{t})^2},$$

$$b_0 = \bar{X}.$$

В этом соотношении суммирование ведется по всем значениям индекса  $i$  от 1 до  $n$ , а диакритический знак черта, как принято, означает среднее.

В результате, имеем следующую оценку для измеряемого параметра:

$$\tilde{X}_i = b_1 t_i + b_0 (i = 1, \dots, n).$$

Предполагается, что все измерения, за исключением последнего, уже прошли процедуру отбора на предыдущих этапах. То есть каждое очередное измерение сопровождается оценкой его адекватности. Выброс, таким образом, возможен только для последнего измерения.

Для оценки адекватности последнего измерения построим доверительный интервал значений параметра  $\tilde{X}$ , предсказанных согласно построенной модели, с некоторой доверительной вероятностью  $1-\alpha$ . С этой целью вычислим дисперсию величины  $\tilde{X}$ :

$$V(\tilde{X}) = V(\bar{X}) + (t - \bar{t})^2 V(b_1) = \frac{\sigma^2}{n} + \frac{(t - \bar{t})^2 \sigma^2}{\sum (t_i - \bar{t})^2}.$$

Здесь величина  $\tilde{X}$  суть оценка измеряемого параметра, выполненная согласно построенной регрессионной модели в точке  $t$ . Если дисперсия искажающего шума  $\sigma^2$  является неизвестной, то можно воспользоваться её выборочным аналогом:

$$s^2 = \frac{1}{n-2} \sum (X_i - \tilde{X}_i)^2.$$

Тогда имеем:

$$V(\tilde{X}) = s^2 \left[ \frac{1}{n} + \frac{(t - \bar{t})^2}{\sum (t_i - \bar{t})^2} \right].$$

Следует обратить внимание на то, что величина дисперсии оценки зависит от точки  $t$ , в которой эта оценка выполняется. Очевидно, минимальное значение дисперсии будет для оценки в точке, совпадающей со средним значением  $\bar{t}$ , а максимальное на концах, то есть при крайних значениях рассматриваемого промежутка.

Интервал с доверительной вероятностью  $1-\alpha$  для оценки  $\tilde{X}$  строится следующим образом:

$$\tilde{X} \mp t \left( n-2, 1-\frac{1}{2}\alpha \right) \sqrt{V(\tilde{X})}.$$

Здесь используется распределение Стьюдента с  $n-2$  степенями свободы (поскольку такое распределение имеет величина  $s^2$  при предположении, что ошибки измерений подчинены нормальному закону).

В результате, для нашей линейной модели, которую на графике можно представить прямой (рис. 2), получается своеобразный коридор, наиболее узкий в центральной части рассматриваемого временного промежутка и расширяющийся к его концам. В рамках этого коридора с вероятностью  $1-\alpha$  будет находиться истинное значение измеряемого параметра.

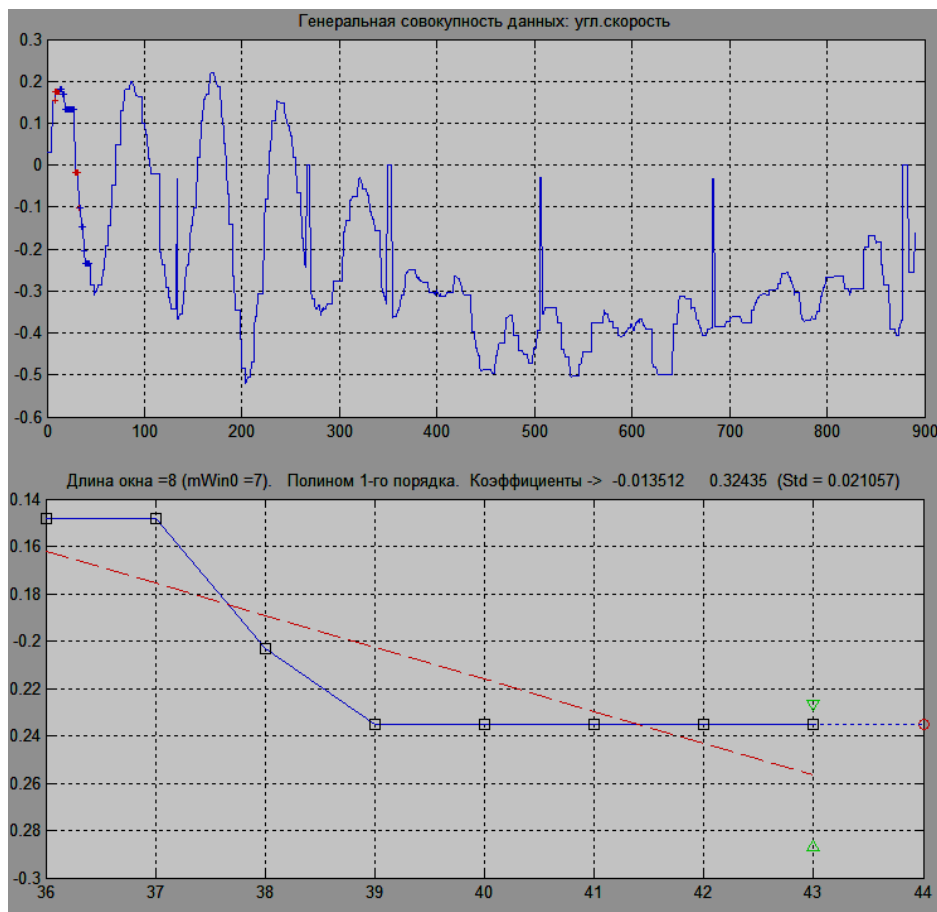


Рис. 2. Оценка текущего измерения на основе линейного регрессионного анализа

Полученное последнее измерение выбраковывается, если оно не попадает в заданный коридор, в противном случае это измерение считается приемлемым – адекватным принятой модели.

Расчеты можно проводить с учетом оцениваемого последнего измерения или без. Но поскольку это измерение может оказаться выбросом, который может существенно исказить регрессионную модель, то лучше его не учитывать.

Вместе с предложенной методом выбраковки измерений можно параллельно рассчитывать величину коэффициента детерминации, который часто используется в регрессионном анализе для оценки адекватности модели. Если этот коэффициент близок к единице, то модель адекватна, а значит и вновь поступивший результат измерения является приемлемым. Однако, для этого коэффициента нельзя дать четкие числовые ограничения и критерии.

## Заключение

Предложен метод предварительной выбраковки результатов измерений параметров в системах, управляемых в реальном масштабе времени. Метод пригоден для ситуаций, когда измерения проводятся с достаточной частотой, так что в рамках коротких временных интервалов, в которых приемлема линейная модель изменения измеряемого параметра, укладывается достаточно большое количество измерений для надежных статистических выводов.

Метод основан на линейном регрессионном анализе в указанных промежутках времени. Выбраковка выбросов может существенно повысить качество последующей фильтрации измерений (например, фильтрации Калмана).

В ходе реализации метода попутно оценивается скорость изменения измеряемого параметра. Скорость в данном методе, очевидно, представляет коэффициент  $b_1$  регрессионной модели, которая строится при каждом новом измерении. Скорость будет оценена таким образом в точке  $\bar{t}$ .

Если система содержит независимый измерительный тракт для скорости изменения этого же параметра, то данные оценки можно использовать для сопоставления и построения алгоритмов выявления отказов или сбоев используемых измерительных трактов, а также для обеспечения отказоустойчивой работы системы в целом.

## Библиографический список

1. Смагин В.И., Смагин С.В. Фильтрация в линейных дискретных нестационарных системах с неизвестными возмущениями // Вестник Томского государственного университета. Управление, вычислительная техника и информатика. 2011. №3 (16), с. 43-51.
2. Cleveland, William S. Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association, Volume 74 (368), 1979, с. 829–836.
3. Мироновский Л.А., Слаев В.А. Алгоритмы оценивания результата трех измерений. — СПб.: «Профессионал», 2010. — 192 с.: ил.
4. Иглин С.П. Теория вероятностей и математическая статистика на базе MATLAB. Учебное пособие. — Харьков НТУ «ХПИ», 2006. — 612 с.: ил.
5. Дрейпер Н., Смит Г. Прикладной регрессионный анализ: В 2-х кн. Кн.1.-М.: Финансы и статистика, 1986.- 366с.

**И.В. Ганичев** кандидат технических наук, доцент,  
**С.В. Киселева**, доцент,  
**Н.В. Лушкин** кандидат технических наук, доцент

## **ИНСТРУМЕНТАЛЬНЫЕ СРЕДСТВА ИНФОРМАТИЗАЦИИ ОБРАЗОВАНИЯ**

Прежде всего это три составные части программного обеспечения, в которые отличным образом укладывается концептуальные модели информационных процессов, систем и технологий, а также и их визуальные образы.

Тройка пакетов Mathcad, Maple и LaTeX представляют уникальные возможности отображения дидактических единиц и составных частей научной, информационной и исследовательской деятельности, начиная от базовых алгоритмов вплоть до предметных областей будущей деятельности выпускника ВУЗа. За ними стоит компьютерная (процедурная и символическая) математика и преобразования моделей объектов. Пакеты несут в себе заряд визуального представления математических формул и моделей, раскрытия глубинной сущности процедурных алгоритмов и технологической части проектирования информационных процессов, объектов и систем.

Примером первого сочетания подобного дидактического и пропедевтического входа со стороны представления чисел десятичной системы счисления является схема Горнера, раскрывающая структуру алгоритма двоичного кодирования (на основе перехода от диад к триадам и тетрадам кода). В результате разработки процедурной части на основе восприятия и разработки студентами интерактивного алгоритма игрового интеллектуального разложения целых чисел на элементарные компоненты по признакам «чет-нечет» изучается способы отладки тестирования кода программы. Здесь используется все достоинства визуализации некоммерческой версии пакета sMathStudio с аналогом «маткадовского» интерфейса.

Ядром символьных преобразований в Mathcad является подобная часть из пакета Maple.

Если пакеты Mathcad и Maple используются в исследовательской деятельности, то пакет LaTeX позволяет автоматизировать многие задачи набора текста для публикаций, включая набор текста на нескольких языках. Существует несколько стандартных стилевых файлов для статей, книг и т.д. Многие издательства и журналы предоставляют свои собственные стилевые файлы, что позволяет быстро оформить публикацию, соответствующую стандартам издания. Студенты могут оформить свои курсовые, дипломные и магистерские работы, используя пакет LaTeX.

Во многих компьютерных аналитических системах, например, Maple, Mathematica, Maxima возможен экспорт документов в формат \*.tex.

LaTeX относится только к языку разметки, он не является текстовым редактором. Для создания документа с его помощью, надо набрать .tex файл с помощью какого-нибудь текстового редактора, например, WinEdt, TEXCNTR.

Главная идея LaTeX состоит в том, что авторы должны думать о содержании, не беспокоясь о конечном визуальном облике. При подготовке своего документа, автор указывает логическую структуру текста (разбивая его на главы, разделы, таблицы, изображения), а LaTeX решает вопросы его отображения. Оформление определяется или заранее (стандартное), или разрабатывается для конкретного документа, при этом содержание отделяется от оформления. Из-за механизма программирования новых макросов, возможности системы практически не ограничены. Вот список некоторых возможностей LaTeX:

- алгоритмы расстановки переносов, определения пробелов между слов, балансировка текста в абзацах;

- автоматическая генерация содержания, списка иллюстраций, таблиц и т.д.;

- механизм работы с перекрёстными ссылками на формулы, таблицы, иллюстрации, их номер или страницу;

- механизм цитирования библиографических источников;

- размещение иллюстраций (иллюстрации, таблицы и подписи к ним автоматически размещаются на странице и нумеруются);

- оформление математических формул, возможность набирать многострочные формулы, большой выбор математических символов;

- оформление химических формул и структурных схем молекул;

- оформление графов, схем, диаграмм;

- оформление алгоритмов, исходных текстов программ с синтаксической подсветкой;

Документ LaTeX – это текстовый файл, содержащий специальные команды языка разметки. Документ делится на преамбулу и тело. Преамбула содержит информацию о классе документа, использованные пакеты макросов, определения макросов, автора, дату создания документа и другую информацию.

Например,

```
\documentclass[12pt]{article} %Документ будет печататься в 12
пунктов и принадлежит классу article. Знак % перед текстом является
комментарием
```

```
\usepackage[russian]{babel} % Пакет поддержки русского языка
```

```
\begin{document}
```

```
\begin{center}
```

```
{\LARGE\bf Управление процессом выщелачивания пульсациями
растворителя. }
```

```
\end{center}
```

```

\begin{center}
{\Large Н.В. Лушкин}
\end{center}
\begin{center}
{\it Санкт-Петербургская государственная лесотехническая академия
им.~С.~М.~Кирова Санкт-Петербург}
\end{center}
\bigskip
\bigskip Анализируя полученные математические зависимости и экс-
периментальные данные по кинетике .....
% текст статьи
$$ %формулы
\frac{\partial T_1}{\partial t}+V\frac{\partial T_1}{\partial
Z}=D_p\frac{\partial^2 T_1}{\partial Z^2}-\frac{(1-
\varepsilon)\lambda_1}{\varepsilon\rho_1
C_{p1}}\left(3\bar{M}_0\varphi_1^2\frac{\partial\varphi_1}{\partial
t}+\frac{\kappa}{\bar{M}_0\rho_s}
C_{ps}\right)\frac{\partial\bar{T}_s}{\partial t}\right),
\eqno(1) % номер формулы
$$
\bigskip
\begin{thebibliography}{99} % список литературы
\bibitem{Lush}{\it Лушкин Н.В.}\quad Система управления техно-
логическими комплексами выщелачивания бокситовых спёков.//Актуальные
проблемы современной математики. Ленинградский областной педагогиче-
ский институт. Т.~{\bf 1}.Санкт-Петербург -- 1996,с.12-23.
%\bibitem{proto}{\it Протодьяконов И.О., Люблинская И.Е., Рыжков
А.С.}\quad
%Гидродинамика и массообмен в дисперсных системах жидкость --
%твердое тело. -- Л.: Химия, 1987; 336~с.
\bibitem{акс}{\it Аксельруд Г.А.}\quad Теоретические основы хими-
ческой технологии, 1953, №~5, сс.~674-677.
\bibitem{Lush1}{\it Лушкин Н.В.,Абрамов В.Я.,Иванов В.А.}\quad
Управление распределёнными процессами в колонных аппаратах с плотно
движущимся слоем.//Изв.вуз.СССР,разд."Цветная металлургия". 1990.N 5
с.102-107.
\bibitem{Lush2}{\it Абрамов В.Я.,Лушкин Н.В.}\quad Модель про-
цесса выщелачивания полидисперсных бокситовых спёков.//Цветная ме-
таллургия. 1990.N 1 с.8-10.
\end{thebibliography} %конец списка литературы
\end{document} %конец документа

```



LaTeX – свободное программное обеспечение, доступен на условиях LaTeX Project Public License и доступен для многих операционных систем.

Чтобы использовать babel с русскими текстами, необходимо включить в преамбулу документа строки:

```
\usepackage[...]{inputenc}  
\usepackage[russian]{babel}
```

В квадратных скобках вместо ... нужно указать кодировку, в которой будет набираться русский текст, например: cp866, cp1251, koï8-r или utf8. Эти строки должны находиться в преамбуле до вызова каких-либо пакетов, использующие русские буквы. После этого в тексте можно набирать русские буквы в указанной кодировке.

**Ю. А.Жук**, кандидат педагогических наук, доцент

## **РОБАСТНАЯ МОДЕЛЬ ОБНАРУЖЕНИЯ АНОМАЛИЙ С ИСПОЛЬЗОВАНИЕМ МОДЕЛИ ЗАСОРЕНИЯ**

Одной из задач статистического машинного обучения является задача обнаружения аномальных наблюдений, которая заключается в идентификации новых или неизвестных системе данных. Идентификация аномалий используется в различных областях (обнаружение несанкционированного доступа, мошенничества, неисправностей, отказов в сенсорных сетях).

Особенностью задачи обнаружения аномальных наблюдений является отсутствие информации о классах, которым принадлежат объекты соответствующих наблюдений. С этой точки зрения задачу обнаружения аномальных наблюдений можно рассматривать как частный случай задачи кластеризации, когда основные данные образуют один кластер, а аномальные данные другой кластер или несколько других кластеров в зависимости от того, каким образом они сосредоточены.

Существует много способов решения задачи обнаружения аномальных наблюдений. Обзор методов идентификации выбросов (аномальных наблюдений) представлен в работах. Сложность решения задачи обнаружения аномалий заключается в том, что точное понятие аномалии различно для разных областей применения. Например, в медицине небольшие отклонения от нормы (например, колебания температуры тела), могут быть аномалией, в то время как аналогичные отклонения в акциях рынка (например, колебания в стоимости акций) могут рассматриваться как нормальные. Таким образом, применение разработанной методики в одной области не может быть всегда использовано в другой.

Задачу обнаружения аномальных наблюдений можно рассматривать и как одноклассовую классификацию (ОКК). Основная часть методов обнаружения аномальных наблюдений, основанная на идеях ОКК, решается методом оценивания некоторой бинарной функции  $f$ , которая задает границы области нормальных наблюдений. Отдельные точки данных, которые находятся за пределами этого региона, могут рассматриваться как аномалии. Данную функцию в дальнейшем будем называть разделяющей, поскольку подразумевается, что она разделяет нормальные и аномальные наблюдения. Основной проблемой построения такой функции является то, что в большинстве случаев она должна быть нелинейной.

Существует много моделей ОКК, использующих метод опорных векторов для обнаружения аномальных наблюдений. Одним из наиболее известных подходов к решению этой задачи является модель, предложенная в работе. Согласно этой модели элементы обучающей выборки в виде векторов  $m$  признаков при помощи спрямляющего отображения, переводятся в часть сферы с центром в точке  $0$ . Далее находится плоскость, которая отсекает ту часть сферы, где находится основная часть переведенных векторов как можно дальше от  $0$  и ближе к данным на сфере.

Предлагается модификация модели, предложенной в работе Тэкса и Дьюина. В соответствии с этим подходом, обучение заключается в поиске наименьшей гиперсферы, содержащей большинство элементов обучающей выборки. Слово "большинство" определяет нормальные наблюдения. Все оставшиеся элементы обучающей выборки, оказавшиеся вне гиперсферы, считаются аномальными или ошибочными. Данная модель устанавливает компромисс между числом аномальных данных в обучающей выборке, которые не входят в гиперсферу и размером гиперсферы, т.е. ее радиусом. Этот компромисс также достигается решением задачи квадратичного программирования. Влияние аномальных наблюдений задается специальным штрафным параметром.

Модель Тэкса и Дьюина является одной из наиболее эффективных моделей в случае, когда имеется большое количество наблюдений. Это условие не всегда выполняется. Предлагаемая в данной работе новая модель является модификацией модели Тэкса и Дьюина, которая учитывает возможность того, что объем обучающей выборки может быть небольшим и использование эмпирического распределения вероятностей элементов выборки может привести к существенным ошибкам. Основная идея, предлагаемая и реализованная в данной работе, заключается в использовании одной из модификаций известной робастной модели  $\epsilon$ -засорения.

Существует большое количество робастных моделей классификации. Большинство подходов предполагает, что каждая точка обучающей выборки может находиться в любом месте внутри шара с центром в этой точке. Радиус этого шара является параметром робастности. Чем больше радиус,

тем больше робастность. Обычно принимается решение с условием "наихудшей" ситуации, когда радиусы шаров максимальны.

Предлагаемая модель принципиально отличается от существующих робастных моделей. Поскольку имеющиеся наблюдения – это данность, любые предположения о том, что они могут быть другими не совсем верны. Поэтому предположение о том, что точки находятся внутри шаров совершенно необоснованно. Иной подход к построению робастных моделей рассматривает не неточности элементов обучающей выборки, а неточности вероятностей этих элементов. Равные вероятности точек в соответствии с эмпирическим функционалом риска – это не данность, а предположение. Здесь можно обоснованно говорить о неверности такого предположения и рассматривать возможные отклонения от этих вероятностей. Поэтому при наличии  $n$  наблюдений предлагается строить окрестности не вокруг самих точек, а вокруг вероятностей  $1/n$  этих точек при условии, что сумма вероятностей всех точек равна 1. Это приводит к модификации модели  $\epsilon$ -засорения, когда основное распределение вероятностей является равномерным, а засоряющее распределение может быть произвольным.

Такая модель образует множество распределений вероятностей на элементах обучающей выборки. Для решения задачи ОКК необходимо выбрать одно распределение из множества, которое максимизирует функционал риска, и одно распределение, которое минимизирует этот функционал. В первом случае используется известная минимаксная (пессимистическая) стратегия принятия решений, которая предполагает «страховку» против наихудшего случая. Противоположной стратегией является оптимистическая стратегия, в соответствии с которой выбирается "наилучшее" распределение, обеспечивающее минимальное значение функционала риска.

В работе предлагается решение задачи обнаружения аномальных наблюдений методом минимаксной и миниминной стратегий, которые сводятся к решению задач квадратичного программирования.

#### Модель обнаружения аномалий Тэкса и Дьюина

Задача обнаружения аномалий может быть сформулирована следующим образом. Предположим, что имеется обучающая выборка  $x_1, \dots, x_n \subset X$ , где  $n$  - число наблюдений,  $X$  - некоторое множество, например, компактное подмножество  $R^m$ . Тэкс и Дьюин в работе предложили искать шар, с минимальным объемом, содержащий все или большинство точек обучающей выборки. Поскольку объем определяется его радиусом  $r$  или квадратом радиуса, то объем шара минимизируется путем минимизации радиуса. С другой стороны, полученный шар должен содержать все элементы обучающей выборки  $x_i$ . Это приводит к следующей задаче оптимизации:

$$F(r, \mathbf{a}) = \min_{r, \mathbf{a}} r^2 \quad (1)$$

при ограничениях  $\|\mathbf{x}_i - \mathbf{a}\|^2 \leq r^2, \forall i = 1, \dots, n$ .

Здесь  $\mathbf{a}$  является центром шара. Ограничения означают, что каждая точка  $\mathbf{x}_i$  находится внутри шара. Однако, данное условие является слишком строгим и может привести к очень большому шару, поэтому допускается, что некоторые точки данных могут быть вне шара. Это достигается введением вспомогательных переменных  $\xi_i \geq 0$ . Отсюда получим следующую задачу оптимизации:

$$F(r, \mathbf{a}) = \min_{r, \mathbf{a}} \left( r^2 + C \sum_{i=1}^n \xi_i \right),$$

при ограничениях  $\|\mathbf{x}_i - \mathbf{a}\|^2 \leq r^2 + \xi_i, \xi_i \geq 0, \forall i = 1, \dots, n$ .

Здесь параметр  $C$  определяет компромисс между объемом шара и количеством ошибок (числом «выброшенных» или аномальных точек данных).

Тэксом и Дьюином было показано, что двойственная задача имеет вид

$$L(r, \mathbf{a}, \alpha_i, \gamma_i, \xi_i) = \sum_{i=1}^n \alpha_i \langle \mathbf{x}_i, \mathbf{x}_i \rangle - \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle \quad (2)$$

при ограничениях  $0 \leq \alpha_i \leq C, i = 1, \dots, n, \sum_{i=1}^n \alpha_i = 1$ .

Оптимальное значение  $\alpha_i$  соответствует максимальному значению  $L$ . Радиус шара  $r$  вычисляется как расстояние от центра сферы до опорных векторов с весом меньшим, чем  $C$ , т.е.

$$r^2 = (\mathbf{x}_k \cdot \mathbf{x}_k) - 2 \sum_{i=1}^n \alpha_i \langle \mathbf{x}_i, \mathbf{x}_k \rangle + \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle$$

для любого опорного вектора  $\mathbf{x}_k$  так что  $\alpha_k < C$ .

Новый объект  $\mathbf{z}$  относят к аномальным, если  $\|\mathbf{z} - \mathbf{a}\|^2 \geq r^2$ .

Робастные модели и множества распределений вероятностей

Использование робастных моделей в машинном обучении получило широкое распространение благодаря их способности уходить от слишком жестких предположений, принятых в стандартных моделях.

Одна из наиболее популярных робастных моделей классификации основана на предположении, что каждое наблюдение может быть аддитивно зашумлено, т.е.  $\mathbf{x}_i^* = \mathbf{x}_i + \Delta \mathbf{x}_i$ , где шум  $\Delta \mathbf{x}_i$  имеет некоторое определенное распределение вероятностей. Простейшее предположение заключается в том, что шум ограничен  $\|\Delta \mathbf{x}_i\| \leq \delta_i$  и имеет равномерное распределение. В соответствии с этой моделью, для каждого  $i$ -го наблюдения известно толь-

ко то, что оно находится внутри шара радиуса  $\delta_i$  с центром в точке  $x_i$ . Такая модель имеет достаточно понятную геометрическую интерпретацию.

Другой класс робастных моделей основан на уходе от слишком жестких предположений о вероятности каждого наблюдения.

Нами предлагается класс моделей, которые основаны на использовании робастной модели  $\varepsilon$ -засорения и обобщают ее. Пусть  $p = (p_1, \dots, p_n)$  – оценка некоторого распределения вероятностей. Модель  $\varepsilon$ -засорения образует множество распределений вероятностей  $M(\varepsilon) = \{(1-\varepsilon)p_i + \varepsilon q_i\}$  для каждого фиксированного  $\varepsilon \in (0,1)$  и  $p_i$ , где  $q_i$  – произвольные вероятности, на которые накладывается единственное ограничение  $q_1 + \dots + q_n = 1$ . Другими словами,  $q = (q_1, \dots, q_n)$  является произвольным распределением вероятностей из единичного симплекса. В соответствии с этой моделью, для  $0 < \varepsilon < 1$  множество  $M(\varepsilon)$  есть множество всех распределений вероятностей с элементами, ограниченными нижней границей  $(1-\varepsilon)p_i$  и верхней границей  $(1-\varepsilon)p_i + \varepsilon$ .

Вернемся к целевой задаче (1) и перепишем ее следующим образом:

$$F(r, \mathbf{a}) = r^2 + C \sum_{i=1}^n \max\left(0, \|\mathbf{x}_i - \mathbf{a}\|^2 - r^2\right).$$

Второй член можно рассматривать как ожидаемые потери  $R(r, \mathbf{a})$  при условии, что функция потерь равна

$$L(\mathbf{x}, \mathbf{a}, r) = \max\left(0, \|\mathbf{x} - \mathbf{a}\|^2 - r^2\right)$$

Предположение об эмпирическом распределении вероятностей означает, что каждая точка  $x_i$  имеет вероятность  $p_i = 1/n$ . Это слишком строгое предположение, когда число точек обучающей выборки мало. Для ослабления этого жесткого условия для вероятностей точек применим модель  $\varepsilon$ -засорения. В соответствии с моделью имеется некоторое неизвестное истинное распределение вероятностей в  $M(\varepsilon)$ , но мы его не знаем, а знаем только то, что оно принадлежит  $M(\varepsilon)$ . В ряде робастных моделей классификации предполагается, что каждая точка может находиться внутри шара. Отличием предлагаемой модели является то, что вероятность  $1/n$  каждой точки (но не сама точка) может находиться в единичном шаре, а точнее симплексе. В этом заключается основная идея построения робастной модели ОКК, рассмотренная ниже.

Для построения модели ОКК необходимо определить правило выбора оптимального распределения из  $M(\varepsilon)$ . Согласно минимаксной стратегии мы выбираем такое распределение вероятностей из  $M(\varepsilon)$ , что функционал риска  $R(r, \mathbf{a})$  достигает максимума для каждого фиксированного набора  $r, \mathbf{a}$ . Согласно миниминной стратегии выбирается распределение вероятностей из  $M(\varepsilon)$ , минимизирующее функционал риска  $R(r, \mathbf{a})$ .

Пусть  $h = (h_1, \dots, h_n)$  – распределение вероятностей, которое принадлежит множеству  $M(\varepsilon)$ . Максимальное значение ожидаемых потерь  $R(r, \mathbf{a})$  имеет вид:

$$\bar{R}(r, \mathbf{a}) = \max_{h \in M(\varepsilon)} R(r, \mathbf{a}).$$

Минимаксный ожидаемый риск в соответствии с минимаксной стратегией принимает вид:

$$\bar{R}(r_{\text{opt}}, \mathbf{a}_{\text{opt}}) = \min_{r, \mathbf{a}} \bar{R}(r, \mathbf{a}) = \min_{r, \mathbf{a}} \max_{h \in M(\varepsilon)} R(r, \mathbf{a}).$$

Верхняя граница ожидаемого риска может быть найдена при решении следующей задачи программирования:

$$\begin{aligned} \bar{R}(r, \mathbf{a}) &= \max_{h \in M(\varepsilon)} \sum_{i=1}^n h_i L(\mathbf{x}_i, \mathbf{a}, r) = \max_{h \in M(\varepsilon)} \sum_{i=1}^n \left( \frac{(1-\varepsilon)}{n} + \varepsilon q_i \right) L(\mathbf{x}_i, \mathbf{a}, r) \\ &= \frac{(1-\varepsilon)}{n} \sum_{i=1}^n L(\mathbf{x}_i, \mathbf{a}, r) + \varepsilon \cdot \max_q \sum_{i=1}^n q_i L(\mathbf{x}_i, \mathbf{a}, r), \end{aligned}$$

при ограничениях

$$0 \leq q_i \leq 1, \quad q_1 + \dots + q_n = 1. \quad (3)$$

Полученная задача оптимизации является линейной с переменными оптимизации  $q_1, \dots, q_n$ , но целевая функция зависит от  $\mathbf{a}$ ,  $r$ . Используем тот факт, что все точки  $q$  относятся к единичному симплексу в конечномерном пространстве. Оптимальное решение достигается в крайних точках симплекса и число этих крайних точек равно  $n$ . Крайние точки симплекса имеют вид  $(1, 0, \dots, 0)$ ,  $(0, 1, \dots, 0)$ ,  $\dots$ ,  $(0, 0, \dots, 1)$ . Это означает, что

$$\bar{R}(r, \mathbf{a}) = \frac{(1-\varepsilon)}{n} \sum_{i=1}^n L(\mathbf{x}_i, \mathbf{a}, r) + \varepsilon \cdot \max_{i=1, \dots, n} q_i L(\mathbf{x}_i, \mathbf{a}, r).$$

Целевая функция  $F(r, \mathbf{a})$  может быть переписана теперь как

$$F(r, \mathbf{a}) = r^2 + \frac{C(1-\varepsilon)}{n} \sum_{i=1}^n L(\mathbf{x}_i, \mathbf{a}, r) + C\varepsilon \cdot \max_{i=1, \dots, n} q_i L(\mathbf{x}_i, \mathbf{a}, r).$$

Введем переменную  $G = \max_{i=1, \dots, n} \xi_i$ . Это приводит к задаче квадратичного программирования

$$F(r, \mathbf{a}) = \min_{r, \mathbf{a}, \xi_i, G} \left( r^2 + \frac{C(1-\varepsilon)}{n} \sum_{i=1}^n \xi_i + C\varepsilon G \right)$$

при ограничениях  $\xi_i \geq \|\mathbf{x}_i - \mathbf{a}\|^2 - r^2$ ,  $\xi_i \geq 0$ ,  $i = 1, \dots, n$ ,  $G \geq \xi_i$ ,  $i = 1, \dots, n$ .

Следующей задачей является минимизация  $F(r, \mathbf{a})$  по параметрам  $\mathbf{a}$  и  $r$ . Вместо того, чтобы минимизировать исходную целевую функцию, можно максимизировать двойственную целевую функцию, Лагранжиан

$$L = r^2 + \frac{C(1-\varepsilon)}{n} \sum_{i=1}^n \xi_i + C\varepsilon G - \sum_{i=1}^n \lambda_i \xi_i -$$

$$-\sum_{i=1}^n \varphi_i \left( \xi_i + r^2 - \left( \|\mathbf{x}_i\|^2 - 2\langle \mathbf{a}, \mathbf{x}_i \rangle + \|\mathbf{a}\|^2 \right) \right) - \sum_{i=1}^n \eta_i (G - \xi_i).$$

Здесь  $\lambda_i \geq 0, \eta_i \geq 0, \varphi_i \geq 0, i = 1, \dots, n$  - множители Лагранжа. Приравнявая нулю частные производные функции  $L$ , получим

$$L = \sum_{i=1}^n \varphi_i \langle \mathbf{x}_i, \mathbf{x}_i \rangle - \sum_{i=1}^n \sum_{j=1}^n \varphi_i \varphi_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle. \quad (4)$$

при ограничениях  $0 \leq \varphi_i \leq C(1 - \varepsilon)n^{-1} + \eta_i,$

$$C\varepsilon = \sum_{i=1}^n \eta_i, \quad \sum_{i=1}^n \varphi_i = 1, \quad \eta_i \geq 0, \quad i = 1, \dots, n.$$

Запишем одно из условий Каруша-Куна-Такера  $\eta_i (G - \xi_i) = 0$ . Отсюда равенство  $G = \xi_i$  имеет место для одного значения  $i = k$ , где  $k = \arg \max_{i=1, \dots, n} \xi_i$ . Следовательно,  $\eta_i = 0, \forall i \neq k$ . Возвращаясь к ограничениям, получаем  $\eta_k = C\varepsilon$  и

$$0 \leq \varphi_i \leq C(1 - \varepsilon)n^{-1}, \quad i = 1, \dots, n, \quad i \neq k, \quad 0 \leq \varphi_i \leq (1 - \varepsilon)n^{-1} + C\varepsilon, \quad i = k.$$

Из вышесказанного следует, что задачу (4) можно разложить на  $n$  задач

$$L_k = \sum_{i=1}^n \varphi_i \langle \mathbf{x}_i, \mathbf{x}_i \rangle - \sum_{i=1}^n \sum_{j=1}^n \varphi_i \varphi_j \langle \mathbf{x}_i, \mathbf{x}_j \rangle,$$

при ограничениях  $0 \leq \varphi_i \leq C(1 - \varepsilon)n^{-1} + C\varepsilon \cdot \mathbf{1}_k(i), i = 1, \dots, n, \sum_{i=1}^n \varphi_i = 1.$

Здесь  $\mathbf{1}_k(i)$  - индикаторная функция. Оптимальное значение  $\varphi_i, i = 1, \dots, n$ , соответствует наименьшей целевой функции  $L_k, k = 1, \dots, n$ .

Миниминная стратегия может рассматриваться как прямая противоположность минимаксной стратегии. Согласно миниминной стратегии, минимизируем ожидаемые потери  $R$  по всем распределениям из множества  $M(\varepsilon)$ , а также по всем значениям  $r$  и  $\mathbf{a}$ . Стратегию можно назвать оптимистической, так как она выбирает «лучшее» распределение из множества  $M(\varepsilon)$ .

Аналогично минимаксной стратегии, можно записать

$$\underline{R}(r, \mathbf{a}) = \min_{h \in M(\varepsilon)} R(r, \mathbf{a}).$$

Нижняя граница ожидаемого риска может быть найдена решением следующей задачи программирования (см. вывод для минимаксной стратегии):

$$\underline{R}(r, \mathbf{a}) = \frac{(1 - \varepsilon)}{n} \sum_{i=1}^n L(\mathbf{x}_i, \mathbf{a}, r) + \varepsilon \cdot \min_q \sum_{i=1}^n q_i L(\mathbf{x}_i, \mathbf{a}, r),$$

при ограничениях (3).

Таким образом, имеет место

$$F(r, \mathbf{a}) = r^2 + C \min_{k=1, \dots, n} \left\{ \frac{(1 - \varepsilon)}{n} \sum_{i=1}^n L(\mathbf{x}_i, \mathbf{a}, r) + \varepsilon \cdot L(\mathbf{x}_k, \mathbf{a}, r) \right\}.$$

Следующая задача состоит в минимизации функции  $F(r, \mathbf{a})$  по параметрам  $r$  and  $\mathbf{a}$ . Можно доказать, что задача будет совпадать с той же задачей, полученной для минимаксной стратегии. Однако, оптимальные значения  $\varphi_i$ ,  $i=1, \dots, n$ , в этом случае соответствуют наименьшей целевой функции  $L_k$ ,  $k=1, \dots, n$ .

#### Результаты моделирования

В этом разделе предлагаемый метод иллюстрируется с использованием генерации случайных данных, составляющих обучающую выборку и выборку для тестирования. Метод программно реализован с использованием известного пакета R (<http://www.r-project.org/>). В качестве показателя качества классификации использована точность классификации  $ACC$ , которая определяется, как доля правильно классифицированных данных к общему числу всех данных. Обозначим точность классификации при использовании минимаксной стратегии  $ACC_{mx}$ , миниминной стратегии  $ACC_{mn}$  и для стандартного метода опорных векторов  $ACC_{st}$ .

В качестве примера, все анализируемые модели тестировались на известном наборе данных Iris из UCI Machine Learning Repository. Обучающая выборка содержала три класса (Iris Setosa, Iris Versicolour, Iris Virginica) по 50 экземпляров, с четырьмя признаками: длина чашелистиков в см, ширина чашелистиков в см, длина лепестков в см, ширина лепестков в см. Предполагалось, что данные из класса Iris Setosa являлись аномальными. Количество данных обучающей выборки для тестирования  $n=30$ . Для эксперимента, произвольно выбиралось  $n$  точек, так, что  $(1-\varepsilon_0)n$  точек было взято из множества нормальных примеров и  $\varepsilon_0 n$  точек из аномальных примеров. Здесь  $\varepsilon_0=50/150=0.333$ . Параметрами для моделирования являлись  $\nu=0,333$ ,  $\sigma=0.001$ . Можно отметить, что оптимальные значения параметра  $\varepsilon$  находятся в интервале от 0.3 to 0.4, где точность классификации имеет наибольшее значение. Это говорит о соответствии или близости оптимальных значения  $\varepsilon$  и параметра  $\varepsilon_0$ .

Кроме этого была исследована зависимость точности классификации от количества  $n$  данных обучающей выборки. В частности, если взять  $n=50$ , тогда  $ACC_{mx}=0.947$ ,  $ACC_{mn}=0.707$ ,  $ACC_{st}=0.773$ . Если взять  $n=80$ , тогда  $ACC_{mx}=0,833$ ,  $ACC_{mn}=0.62$ ,  $ACC_{st}=0.713$ . Из приведенных выше экспериментов следует, что существует некоторое оптимальная число примеров обучающей выборки обеспечивающей максимальную точность классификации.

В качестве еще одного примера рассмотрим множество данных Pima Indian Diabetes, которое также взято из UCI Machine Learning Repository. Множество Пима данных имеет восемь признаков ( $m=8$ ) с 768 обучающими примерами, 500 из которых отмечены как нормальные. Остальные 268 нормальных примеров рассматривались, как аномальные наблюдения. Затем предполагается, что  $\varepsilon_0=268/768=0.35$ . Для эксперимента, произволь-



но выбиралось  $n$  точек так, что  $(1-\varepsilon_0)n$  точек было взято из множества нормальных наблюдений, а  $\varepsilon_0 n$  точек из множества аномальных наблюдений. Если взять  $n=50$ , тогда  $ACC_{mx}=0.683$ ,  $ACC_{mn}=0.656$ ,  $ACC_{st}=0.66$ . При  $n=100$ , получаем  $ACC_{mx}=0.684$ ,  $ACC_{mn}=0.677$ ,  $ACC_{st}=0,667$ . Из результатов исследования видно, что минимаксная стратегия обеспечивает лучшие результаты по сравнению с миниминной и стандартной моделью. Однако, это преимущество случайно, поскольку трудно выбрать аномальные наблюдения из множества данных по Pima Indian Diabetes, в связи с их структурой.

Предложенная модель имеет целый ряд преимуществ. Во-первых, ее реализация достаточно проста, так как задача представляется в виде стандартной задачи квадратичного программирования с малым числом ограничений, что существенно ускоряет расчеты. Во-вторых, использование робастности позволяет учесть нарушение равномерности распределения вероятностей элементов обучающей выборки особенно, когда обучающая выборка сравнительно мала и доля аномальных наблюдений велика.

**М.О. Лебедев**, кандидат технических наук, доцент

## **ОПЫТ СОЗДАНИЯ ДИНАМИЧЕСКОЙ ИНФОРМАЦИОННО-ПОИСКОВОЙ И РЕГИСТРИРУЮЩЕЙ СИСТЕМЫ**

Была разработана динамическая информационно-поисковая и регистрирующая система (ДИПРС) для торгово-коммерческого предприятия как средство автоматизации подготовки прайс-листов на свою продукцию.

В большинстве случаев ДИПРС могут быть представлены в виде совокупности трех составляющих:

1. основной список;
2. информационный блок;
3. журналы.

Под основным списком будем понимать информацию, параметры которой (состав полей) не меняется (или меняется крайне редко). Каждая запись основного списка определяет основную единицу хранения в информационной системе (ИС). Например, список студентов ВУЗа, список учащихся школ, список пациентов больницы, список производимой продукции, список товаров и т.д. Основной список представляет собой высоко нормализованную (не ниже третьей нормальной формы) таблицу базы данных. Состав полей этой таблицы практически неизменяемый.

Под информационным блоком будем понимать некую совокупность параметров, описывающих каждый элемент основного списка. Состав параметров может изменяться при переходе от одного элемента основного списка к другому. Этот блок содержит актуальную информацию и не содержит историю изменения этой информации.

Под журналом будем понимать некую совокупность параметров, каждое изменение которой регистрируется как отдельная запись. Совокупность параметров связана с элементами основного списка. Состав полей каждого журнала определен, но может меняться в ходе развития информационной системы. Этот блок содержит историю (движение) параметров, описывающих каждый элемент основного списка.

Информационная составляющая системы была разработана на базе системы Access, программная часть, включающая два клиентских приложения и один СОМ-сервер, разработана в среде Delphi. Одной из основных задач при разработке системы было обеспечение возможности расширения номенклатуры товаров и их параметров, регистрируемых в базе данных, с автоматической адаптацией клиентских программ к новому составу данных (таблиц и полей в терминах теории БД). Разработанная система не предполагала наличия журналов. При этом состав параметров каждой категории (группы) товаров мог меняться в ходе эксплуатации системы. Система должна позволять изменение состава параметров для выборки товаров без ухудшения своей работоспособности. Решение поставленной задачи базировалось на разделении описания объектов и хранимых данных (каждый тип данных хранится в отдельной таблице).

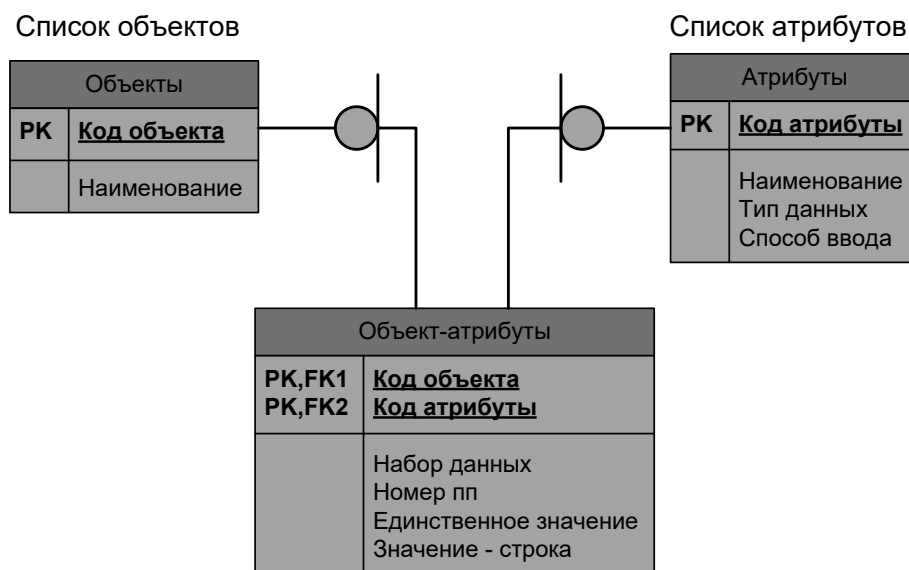


Рис. 1. Описание объектов

Обобщенная структур БД приложения показана на рисунке 2.

Желтым цветом на схеме выделены таблицы, представляющие собой основной список системы и таблицы информационного блока. Совокупность таблиц «Таблица\_<тип>» представляет собой хранилища данных различных типов, относящихся к различным категориям товара.

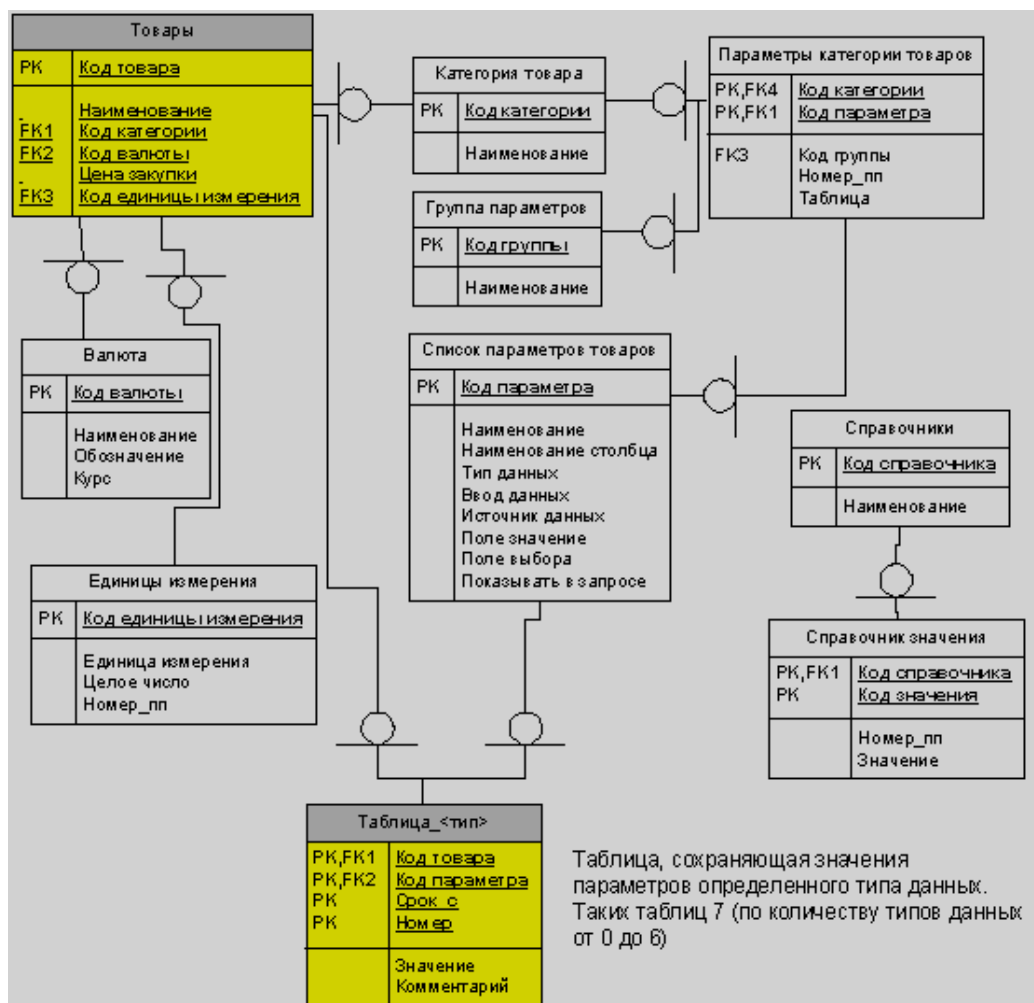


Рис.2. Обобщенная структура базы данных

Программное обеспечение ДИПРС включает в себя три основных компонента: СОМ-сервер, непосредственно работающий с БД, и два клиентских приложения. Структура программного обеспечения представлена на рис. 3.

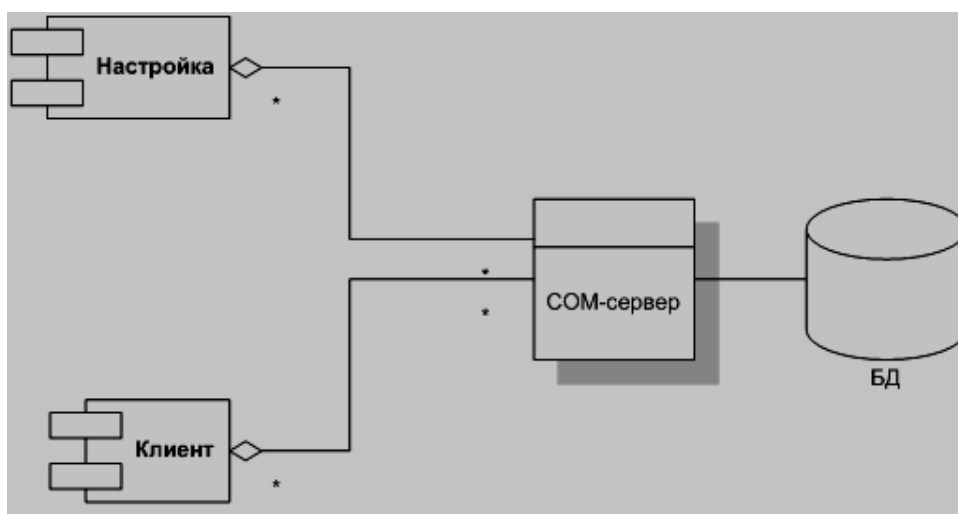


Рис. 3. Структура программного обеспечения.

Программа «Настройка» решает следующие задачи:

- формирования списка параметров групп товаров;
- подготовка справочников возможных значений;
- формирование списков товаров.

Программа «Клиент» решает следующие задачи

- формирование выборок различных товаров по задаваемым условиям;
- подготовка коммерческого предложения.

В ходе разработки системы были выявлены неопубликованные особенности системы Access, влияющие на скорость обмена данным между сервером и клиентом.

Результат опытной эксплуатации системы подтвердил правильность принципов формирования ДИПРС.

**И.В.Панфилов**, доктор технических наук, профессор

## **ОЦЕНКА БЫСТРОДЕЙСТВИЯ МУЛЬТИПРОЦЕССОРНОЙ ВЫЧИСЛИТЕЛЬНОЙ СИСТЕМЫ**

При разработке мультипроцессорной вычислительной системы (МПВС) основным показателем является быстродействие системы. С увеличением числа процессоров в МПВС возрастает доля непроизводительных (системных) затрат, которые в основном связаны с ожиданием процессоров в очереди к системным ресурсам.

Одним из факторов, снижающих эффективность МПВС, являются конфликты при обращении к памяти. Эти конфликты бывают двух типов – программные и аппаратные.

Программные конфликты возникают тогда, когда какой-либо из процессоров «пытается использовать» общую базу данных, которая в данный момент используется другим процессором, включившим механизм защиты для предотвращения доступа со стороны какого-либо другого устройства (программой «затвор»).

Аппаратные конфликты возникают при попытке двух или большего числа процессоров или одного устройства ввода-вывода и одного процессора обратиться к другому и тому же модулю памяти в течение одного цикла памяти.

Таким образом, введение дополнительных процессоров в МПВС вместе с соответствующим увеличением памяти и мощности устройств ввода-вывода не дает линейного возрастания общего быстродействия МПВС.

Непроизводительные затраты МПВС из нескольких процессоров обусловлены:

а) временем ожидания в очереди при обращении нескольких процессоров в течение цикла памяти к одним и тем же модулям оперативной памяти (коэффициент относительных потерь  $\eta_c$ );

б) временем ожидания в очереди при одновременном обращении двух и более процессоров к общим таблицам и программам операционной системы (коэффициент потерь  $\eta_0$ );

в) затраты времени на работу программ диспетчера операционной системы (коэффициент потерь  $\eta_n$ ).

Обозначим эффективное быстродействие одного процессора  $B_{эн}$ , тогда эффективное быстродействие МПВС [1,2] будет

$$W_э = MB_{эн}(1-\eta_c)(1-\eta_0)(1-\eta_n),$$

где  $M$  – число процессоров в системе.

Эффективное быстродействие процессора определим из выражения

$$B_{эн} = \{1/B_{нн} + n[P(\tau - k_0T_u) + k_0T_u]\}^{-1},$$

где  $B_{нн}$  – номинальное быстродействие процессора при выполнении арифметических и логических операций и равно  $1/\tau_{cm}$ ;

$n$  – среднее число обращений к памяти за данными, приходящимися на одну операцию;

$P$  – вероятность обращения к оперативной памяти при выполнении операции процессором;

$\tau$  – длительность цикла быстродействующей сверхоперативной памяти (БСОП) или КЭШ-памяти;

$k_0$  – относительный коэффициент, характеризующий эффективность обмена данными с оперативной памятью;

$T_u$  – длительность основного цикла оперативной памяти;

$\tau_{cm}$  – длительность короткой операции (операция сложения).

Коэффициент относительных потерь  $\eta_c$  из-за ожидания в очереди при обращении нескольких процессоров к одним и тем же модулям оперативной памяти в течение цикла обращения к ОП.

$$\eta_c = 1 - (B'_{эн} / B_{эн}),$$

где  $B'_{эн}$  – быстродействие процессора МПВС с учетом простоя в очереди на обслуживание к модулю памяти.

Эффективное быстродействие процессора с учетом простоя в очереди определим из выражения

$$B'_{эн} = \{1/B_{нн} + n[P(\tau - k_0T_\tau) + (1-l)k_0T_\tau]\}^{-1},$$

где  $l$  – число заявок в очереди на обслуживание модулей памяти, определяемое в соответствии с формулой Поллачека-Хинчина выражением

$$l = \frac{(\lambda/\mu)^2}{2(1-\lambda/\mu)} = \frac{\rho^2}{2(1-\rho)}, \quad (1)$$

Пусть заявки процессоров к модулям памяти образуют пуассоновский поток с параметром  $\lambda$ , равным

$$\lambda = [B_{эн} M (1 - \rho) + \lambda_0] / r,$$

где  $\lambda_0$  – интенсивность обращения системы обмена к ОП;  
 $r$  – число модулей памяти.

Интенсивность обращения системы обмена к ОП определяется быстроедействием ( $M$  процессоров) и частотой их обращения во внешнюю память в соответствии с выражением

$$\lambda_0 = B_{эн} h \cdot M / (H\nu),$$

где  $h$  – средний объем информации (в словах), принимаемый или выдаваемый за один запрос к системе обмена;  $H$  – среднее число операций процессора, приходящихся на один запрос к внешней памяти;  $\nu$  – количество слов в группе обмена между БСОП и ОП.

Пропускную способность каждого модуля памяти  $\mu$  найдем из выражения

$$\mu = 1 / (k_0 T_ч).$$

Коэффициент относительных потерь  $\eta_0$  из-за ожидания в очереди при одновременном обращении двух и более процессоров к общим базам данных операционной системы определяется из выражения

$$\eta_0 = 1 - \frac{B_{эн}}{B_{эн}} \sum_{i=1}^k \frac{l_i \theta_i}{S_i}, \quad (2)$$

где  $k$  – число системных таблиц и программ;  $l_i$  – средняя длина очереди к  $i$ -й общей системной программе;  $\theta_i$  – количество операций, необходимых для одноразовой обработки  $i$ -й системной программы;  $S_i$  – длительность среднего интервала (в операциях) между обращениями к  $i$ -й системной программе.

Величину  $l_i$  определим по формуле (1) при условии пуассоновского потока запросов к общим базам данных с параметром

$$\lambda_i = B_{эн}' M / S_i \quad (3)$$

и показательным законом обслуживания их с параметром  $\mu_i$  равным

$$\mu_i = B_{эн}' / \theta_i \quad (4)$$

Подставляя (3), (4) в (1) получаем

$$l_i = M^2 \theta_i^2 / [2S_i (S_i - M\theta_i)] \quad (5)$$

Используя (5) и (2), находим выражение для  $\eta_0'$ :

$$\eta_0' = 1 - \frac{B_{эн}'}{B_{эн}} \sum_{i=1}^k \frac{M^2 \theta_i^2}{2S_i^2 (S_i - M\theta_i)}$$

Коэффициент потерь  $\eta_n$  из-за работы программ диспетчера операционной системы определим из выражения

$$\eta_n = \frac{B_{эн}' \Delta}{B_{эн} \bar{\theta}},$$

где  $\bar{\theta}$  – средняя трудоемкость полного алгоритма (комплекса) вычислительного процесса;  $\Delta$  – затраты на обслуживание заявки в системе в операциях.

Затраты на обслуживание заявки  $\Delta$  найдем по приближенной формуле для оценки одноканальной системы с диспетчером

$$\Delta \approx \alpha + p_0 \frac{\lambda\beta(T + \alpha)}{1 - \lambda\beta} + e \left[ \frac{\lambda\beta(T + \gamma)}{1 - \lambda\beta} + \frac{\beta - \gamma}{p_0} \right],$$

а  $\bar{e}$  и  $p_0$  из выражений

$$p_0 = \frac{1 - \lambda(T + \gamma + \beta)}{1 - \lambda(\beta - \alpha + \gamma)},$$

$$\bar{e} = \frac{1 - p_0}{2p_0},$$

где  $\alpha$  – затраты на подготовку и проведенные обслуживание заявки в операциях;  $\gamma$  – затраты на обслуживание заявки, поступающей из очереди к диспетчеру;  $\beta$  – затраты на постановку заявки в очередь к диспетчеру в операциях;  $\lambda$  – параметр пуассоновского потока заявок к диспетчеру;  $M=1/T$  – параметр показательного закона обслуживания заявок диспетчером;  $\bar{e}$  – среднее число заявок в системе;  $p_0$  – вероятность отсутствия заявок в системе.

Учитывая, что каждое произведение  $\eta_0\eta_c, \eta_0\eta_n, \eta_c\eta_n, \eta_0\eta_0\eta_n$  значительно меньше  $\eta_i$ , то выражение быстродействия (производительность) мультипроцессорной вычислительной системы приближений можно представить в виде

$$W = B_{\text{ин}}M[1 - (\eta_c + \eta_0 + \eta_n)] \approx B_{\text{ин}}M(1 - \eta),$$

где  $\eta = \eta_c + \eta_0 + \eta_n = k(M)$ .

Непроизводительные затраты системы при работе диспетчерских алгоритмов и устранения конфликтных ситуаций не дают линейного возрастания быстродействия системы при увеличении числа процессоров.

### Библиографический список

1. Мультипроцессорные системы и параллельные вычисления: Пер. с англ./Под ред. Ф.Г.Энслоу. – М: Мир,1976. – 384 с.
2. Панфилов И.В., Половко А.М. Вычислительные системы /Под ред. А.М.Половко. – М.: Сов.радио, 1980, - 304 с.

**И.В.Панфилов**, доктор технических наук, профессор

## **КОНЦЕПЦИЯ ЗАЩИТЫ ИНФОРМАЦИИ ОТ КОМПЬЮТЕРНЫХ ПРЕСТУПЛЕНИЙ**

Анализ зарубежной и отечественной литературы показывает, что уровень мероприятий, предпринимаемых для защиты информации ниже уровня информатизации и автоматизации общества.

Уязвимость информации в автоматизированных комплексах обусловлена большей концентрацией вычислительных ресурсов, их территориальной распределенностью, долговременным хранением большого объема данных на магнитных носителях, одновременным доступом к ресурсам многих пользователей. В этих условиях необходимо принятие мер защиты информации, но при этом существуют трудности:

- на сегодняшний день нет единой теории защищенных систем;
- производители средств защиты в основном предлагают отдельные компоненты для решения частных задач, оставляя вопросы формирования систем защиты и совместимости этих средств на усмотрение потребителей;
- для обеспечения надежной защиты необходимо разрешить целый комплекс технических и организационных проблем и разработать соответствующую документацию.

Обеспечение информационной безопасности – достаточно серьезная задача, поэтому необходимо разработать концепцию безопасности информации, где определить национальные и корпоративные интересы, принцип обеспечения и пути поддержания безопасности информации, а также сформировать задачи по их реализации.

Концепция – официально принятая система взглядов на проблему информационной безопасности и пути ее решения с учетом современных тенденций. Она является методологической основой политики разработки практических мер по ее реализации. На базе сформулированных в концепции целей, задач и возможных путей их решения формируются конкретные планы обеспечения информационной безопасности.

Безопасность информации – это способность системы, ее обрабатывающей, обеспечить в заданный промежуток времени возможность выполнения заданных требований по величине вероятности наступления событий, выражающих в утечке, модификации или утраты данных, представляющих ту или иную ценность для их владельца. Российские предприятия вынуждены уделять больше внимания повышению компьютерной безопасности в своих автоматизированных системах обработки информации (в локальных сетях), поскольку для нашего общества характерны:



- экономические отношения различных форм собственности в условиях риска;
- повышенный рост криминализации общества в целом и в частности.

Анализ социальной обстановки в нашей стране свидетельствует о том, что в годы перестройки произошло резкое увеличение количества программно-диверсионных актов в отношении объектов и персонала государственных и частных предпринимательских структур. Организаторы актов диверсии в отношении тех или иных предприятий стремятся нанести им материальный ущерб, сорвать на длительный срок их нормальное функционирование.

Основная цель предприятия – сохранение и приумножение капиталов для получения прибыли, обеспечения его безопасности, т.е. сведения к минимуму уменьшения прибыли за счет непредвиденных убытков при наступлении нежелательных событий. Это достигается с помощью комплекса мероприятий по снижению всех видов риска и созданием эффективных систем защиты информации, защиты ресурсов предприятия.

Компьютерная безопасность предприятия осуществляется в двух направлениях:

- физическая безопасность;
- информационная безопасность.

Составляющие физической безопасности:

- системы контроля доступа на территорию объекта, содержащие систему охранной сигнализации и контрольно-пропускной пункт – служит для ограничения доступа лиц на территорию объекта, а также совместно со специальными аппаратными решениями составляют средство защиты от побочных электромагнитных излучений и наводок информации;
- средства охраны помещений, офисов, спецхранилищ и т.п.;
- средства контроля за доступом к информационным залам;
- организационные мероприятия (работа с людьми с целью формирования у них понимания необходимости проведения мероприятий по обеспечению безопасности информации);
- отбор надежных профессионально подготовленных людей при приеме на работу и т.п.

### **Информационная безопасность**

Назначение систем информационной безопасности сводится к защите от несанкционированного доступа и модификации информации, а также восстановлению информации после разрушения. Функции систем информационной безопасности: аутентификация, разграничение доступа, защита на сетевом уровне.

Аутентификация чаще всего выполняется с помощью паролей. Разработан сервер Kerberos, предназначенный для аутентификации пользова-

теля, входящего в сеть с любого узла. Целесообразна периодическая смена паролей, доступ к файлам пароля должен быть только у администратора сети и т.п.

Разграничения доступа должно обеспечиваться на нескольких уровнях. Так существует четырехуровневая модель. На внешнем уровне устанавливаются права доступа к корпоративной сети извне и выхода из нее изнутри. На сетевом, системном и прикладном уровнях регламентируются права доступа к сетевым информационным ресурсам, ресурсам операционных систем и пользовательскими данными. Другая модель устанавливает уровни входа в систему доступа к базам данных, доступа к приложениям. Права доступа часто представляются трехразрядным восьмеричным кодом АВС, в котором А – права владельца, В – права членов группы, С – права остальных пользователей, а три бита выражают права чтения, записи и использования.

Между общедоступными и секретными объектами в сети можно установить специальное программное обеспечение, называемое *брандмауэром*, которое либо запрещает выполнение определенных действий на сервере, либо фильтрует пакеты, разрешая проход только от оговоренных узлов.

Борьба с перехватом сообщений на сетевом уровне осуществляется методом криптографии. Криптография – это наука об обеспечении безопасности данных путем их шифрования.

Составляющие информационной безопасности:

- конфиденциальность – защита от несанкционированного получения информации;
- целостность – защита от несанкционированного изменения информации;
- доступность – защита от несанкционированного удержания информации и ресурсов.

Конфиденциальность и целостность информации – это задача создания ситуации, при которой исключается утечка, модификация и утрата информации для ее владельца.

Доступность информации должна обеспечиваться основными средствами автоматизации ее обработки. Их задача – предоставить доступ к информации, санкционированный ее владельцем или доверенным лицом, отвечающим за ее безопасность.

Оба направления безопасности информации взаимосвязаны. Например, эффективный входной контроль является одним из методов по ограничению доступа к компьютерам, а работа с персоналом оказывает значительное влияние на информационную безопасность предприятия, так как сотрудники предприятия часто являются источником утечки информации.

Обеспечение информационной безопасности любой системы включает в себя не только защиту информационных ресурсов, но и создание не-

обходимого уровня информативности как самой системы о процессах, происходящих внутри ее и во внешней среде.

**М.А.Шубина**, кандидат технических наук, доцент

## **ИСПОЛЬЗОВАНИЕ МЕТОДОВ КЛАССИФИКАЦИИ КОСМИЧЕСКИХ ИЗОБРАЖЕНИЙ ДЛЯ МОНИТОРИНГА ОСОБО ОХРАНЯЕМЫХ ТЕРРИТОРИЙ.**

Вопросы сохранения природы относятся к числу актуальных задач мирового сообщества. Наиболее эффективным способом сбережения природы является выделение территорий с особым статусом землепользования – особо охраняемых территорий ООПТ [1]. Для мониторинга экологической ситуации этих территорий эффективным способом является использование дистанционных методов. Одним из направлений дипломного проектирования на кафедре «информационных систем и технологий» является использование материалов дистанционного зондирования для оценки состояния ООПТ. Важнейшим этапом этого процесса является классификация изображений с помощью специальных программных систем, имеющих в интернете в свободном доступе, например, ENVI.

Алгоритмы классификации изображений использовались для оценки состояния особо охраняемых природных территорий на примерах изображений водно-болотных угодий Псковской области: государственного природного заповедника "Полистовский", целью создания которого является сохранение одной из крупнейших в Европе Полисто-Ловатской системы верховых болот, изучение естественного хода природных процессов и явлений, генетического фонда растительного и животного мира, отдельных видов и сообществ растений и животных, типичных и уникальных экологических систем; «Себежского национального парка» [2]. В данной работе приведены результаты анализа алгоритмов фильтрации и классификации программного комплекса ENVI 4.7 с участием студентов и магистров.

### **Основные способы компьютерной классификации объектов по снимкам.**

Общие подходы и способы классификации. По индексным и другим производным изображениям успешно выделяются определенные типы объектов, однако такие изображения не позволяют одновременно дешифровать все многообразие объектов, отображаемое на тематических (например, ландшафтных) картах. Эта задача *компьютерного дешифрирования* решается на основе более трудоемкого, но универсального подхода — *классификации*, под которой понимают автоматизированное подразделение всех пикселей снимка на группы, соответствующие разным объек-

там (классы), обычно по спектральным признакам, т.е. на основе различий в значениях спектральной яркости. Основопологающим в классификации является правильный выбор признаков. Недаром известный американский ученый Б.К.П.Хорн в своей книге «Зрение роботов» (1989) писал: «Никакие ухищрения в алгоритме решения не могут компенсировать неудачный выбор признаков. Когда возникают трудности с классификацией, не нужно искать экзотических математических хитростей, а лучше поискать подходящие признаки». Для успеха классификации в пространстве спектральных признаков каждому классу должна соответствовать своя область значений яркости, не пересекающаяся с областями других классов. Чтобы проверить, так ли это, строят двумерные графики пространства спектральных признаков и выбирают для классификации те сочетания спектральных зон, в которых изучаемые объекты различаются лучше всего.

*Способы классификации* зависят от того, как выделяют и ограничивают область значений яркости класса. Наиболее оптимальный (экономичный) способ классификации должен учитывать характер распределения значений яркости между классами и внутри каждого класса.

***Группы способов классификации по использованию распределения значений яркости*** делятся на непараметрические и параметрические. В группе *непараметрических способов* распределение значений яркости внутри класса не учитывается и не описывается никакими параметрами. Это распределение может быть неоднородным, класс может включать произвольные скопления сочетаний спектральных яркостей. Например, так распределяются яркости крыш домов, сделанных из материалов с разной отражательной способностью, но отнесенных в единый класс «крыши». Для отделения этого класса от других нужно точно описать границу занимаемой им области. Если область отстоит далеко от областей других классов, то ее можно ограничить простой геометрической формой, например прямоугольником (рис 1.а) [3].

Прямоугольник в двумерном пространстве спектральных признаков задается лишь минимальным и максимальным значением яркости класса в каждой спектральной зоне, поэтому классификация идет быстро. Если же область находится рядом с другими и ее границы имеют сложную геометрическую форму, то для классификации требуется много данных и машинного времени. Например, при многоугольной форме области нужно знать все значения яркости, задающие углы многоугольника (рис. 1. б).

У непараметрических способов два основных применения. Во-первых, их используют для крайне простых классификаций, когда надо различить всего 2 - 4 контрастных по яркости объекта (например, вода, суша, облака). Во-вторых, их применяют в наиболее сложных случаях, когда классы тесно соседствуют в пространстве спектральных признаков и внутренне неоднородны, из-за чего невозможно применение параметрических способов.

Вторая группа включает *параметрические способы* классификации, предполагающие, что распределение значений спектральной яркости внутри класса подчинено определенному закону. Тогда распределение моделируют с помощью стандартных расчетов на основе этого закона, что позволяет не тратить время на детальное определение границы области, занимаемой значениями яркости класса.

Почти всегда используют закон *нормального (гауссова) распределения*, типичного для яркостей природных объектов. Для него характерны симметричное распределение значений яркости вокруг среднего значения, непрерывность и преобладание малых отклонений от среднего.

Графически этому соответствует симметричная колоколообразная кривая (рис.1. в). Нормальное распределение значений яркости класса в каждой спектральной зоне рассчитывается по двум основным параметрам: *среднему значению  $\beta_i$*  и *стандартному отклонению  $\lambda_i$* .

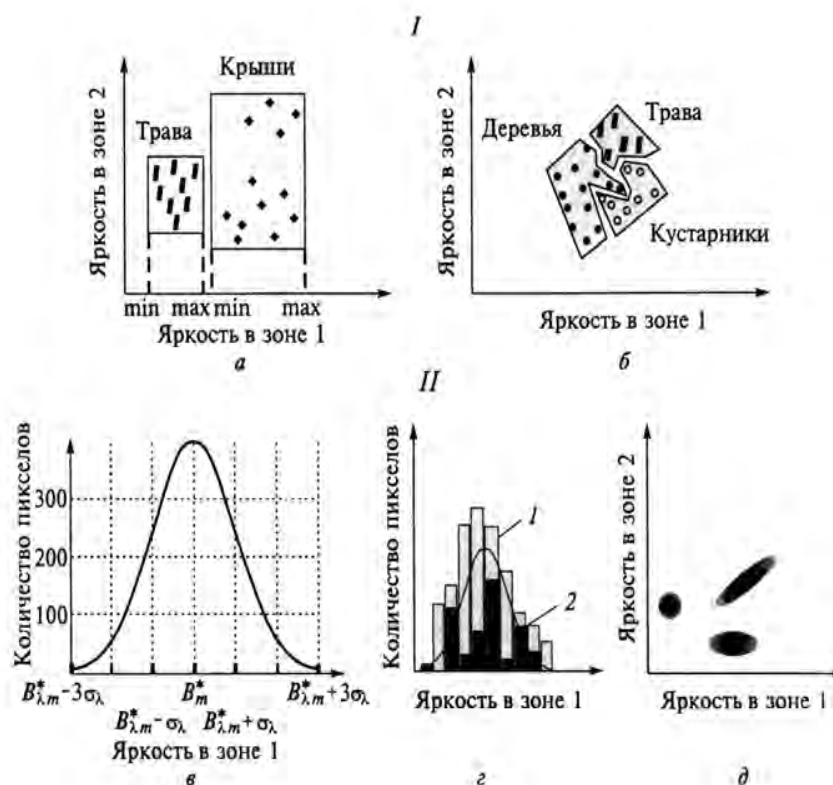


Рис 1. Способы классификации : I— непараметрический:  
 а — с простыми границами между классами; б — со сложными границами между классами;

II — параметрический:  $v$  — кривая нормального распределения значений яркости;  $z$  — оценка нормальности распределения по гистограмме значений яркости класса: 1 — близкое к нормальному; 2 — многомодальное, далекое от нормального;  $d$  — отображение классов с нормальным распределением яркостей в пространстве спектральных признаков.

Стандартное отклонение определяют по формуле

$$\sigma_{\lambda} = \sqrt{\frac{\sum_{i=1}^n (B_{\lambda i}^* - B_{\lambda m}^*)^2}{n-1}},$$

где  $n$  — количество значений яркости пикселей в пределах класса.

Стандартное отклонение интерпретируется как отстояние значений яркости от среднего значения яркости всего класса. Чем меньше  $\sigma$ ; тем лучше среднее значение представляет класс в целом. Из математической статистики известно, что для нормального распределения в пределах  $\pm 1\sigma$  от среднего значения находится более 68 % общего числа значений, а в пределах  $\pm 3 < \sigma$  — более 99 %. Т.о., интервал  $B_{\lambda m} \pm 3\sigma$  содержит практически все значения яркости для нормально распределенного класса. Чтобы оценить, близость распределения значений спектральной яркости класса к нормальному, строят гистограмму по значениям яркости пикселей в пределах класса и сравнивают ее с графиком кривой нормального распределения, имеющей те же  $M$ , и  $\sigma$  (рис. з).

Наиболее важно, чтобы гистограмма была одномодальной, т.е. имела только один четко выраженный пик по яркости. Если на ней есть несколько пиков, то класс следует разбивать на несколько подклассов по яркости. Симметричность и высота гистограммы по сравнению с теми же параметрами кривой нормального распределения имеют важное, но второстепенное значение. При параметрической классификации по многозональному снимку предполагают, что распределение значений яркости в пределах каждого класса в каждой спектральной зоне нормальное. При двумерном пространстве признаков область класса с нормальным распределением значений яркости отобразится как круг, если  $\sigma$  в двух зонах одинаковые, а при разных  $\sigma$   $x$  — как эллипс. Если зональные значения яркости класса меняются взаимосвязано, то эллипс имеет определенный наклон (рис. д).

Задачи лесопользования, решаемые с помощью алгоритмов классификации:

- Определение породного состава лесов по оптическим и радарным космическим снимкам.
- Разделение лесов на категории по возрасту, степени спелости, запасу древесной массы, биологической продуктивности.

- Изучение по космическим снимкам негативных процессов, воздействующих на лесные массивы: влияния вредителей и болезней, иссушения или переувлажнения лесов, приводящих к их деградации и гибели.

- Обнаружение массивов леса, полностью или частично поваленного ураганными ветрами, определение площадей ветровалов.

- Контроль лесовосстановительных работ, мониторинг процессов лесовосстановления и др.

- Выявление существующих вырубок и гарей, оперативный автоматизированный мониторинг появления новых участков, пройденных пожарами, и вырубок (в т.ч. несанкционированных), определение экономического и экологического ущерба.

- Автоматизированная экспресс-оценка по оптическим и радиолокационным снимкам фактических площадей рубок, появившихся со времени последнего лесоустройства.

- Изучение природных условий, способствующих или препятствующих активной лесохозяйственной деятельности (выявление плоских пониженных заболоченных участков, бессточных котловин, резких перегибов рельефа и т.п.) с применением ЦМР.

- Оперативное автоматизированное выявление очагов пожаров размером вплоть до десятков квадратных метров на базе традиционных алгоритмов.

Программные системы, позволяющие автоматизировать процесс классификации: ERDAS, ErMapper, ENVI, IDRISI и др.

Можно рассмотреть основные алгоритмы классификации на примере программного комплекса ENVI (Environment for Visualizing Images — среда для отображения снимков), выпускаемого американской корпорацией ИТТ Visual Information Solutions, который содержит достаточно полный набор функций для визуализации и обработки данных дистанционного зондирования Земли (ДЗЗ) и их интеграции в геоинформационные системы (ГИС).

Классификация начинается с обрисовки эталонных полигонов. С помощью Edit \ Edit/Create Regions открывается окно векторных инструментов Tools. Эти инструменты позволяют рисовать полигоны, присваивать им цвета и имена. Если одному классу соответствуют несколько эталонных полигонов, то они должны быть названы одинаково. Цвет полигона в дальнейшем будет присвоен соответствующему классу. После того как все эталонные полигоны созданы, необходимо сосчитать статистику.

Окно supervised Classification открывается с помощью команды Process \ Classification \ Supervised Classification. В нем можно выбрать каналы и способ классификации: Maximum Likelihood, Minimum Distance, Minimum Distance (Std. Dev.), Parallelepiped, Mahalanobis, о которых говорилось выше.

Способ Maximum Likelihood имеет четыре разновидности: Enhanced, Enhanced Neighbor, Standard and Standard Neighbor. Кнопка Setup открывает окно, где нужно установить параметры классификации. Если выбран способ Minimum Distance или Minimum Distance (Std. Dev.), то можно установить порог (Threshold) (в стандартных отклонениях) принадлежности пикселя классу. Если пиксель находится дальше заданного значения от центра класса, то он помечается как отклассифицированный. Если выбран способ Maximum Likelihood Enhanced Neighbor, то можно установить следующие параметры:

Prior Probability – в поле Training Area Information можно установить процентное содержания пикселей каждого класса в результирующем изображении,

Neighbor Filter – размер окна фильтра учета соседних пикселей.

Наиболее часто используются следующие алгоритмы классификации:

С обучением (в процессе проведения такой классификации значения яркости текущего пикселя сравниваются с эталонными и на основании этого сравнения пиксель относится к наиболее подходящему классу объектов):

1. Способ параллелепипеда, который использует простое решающее правило, и используется, когда значения спектральной яркости разных объектов практически не перекрываются, а классов объектов немного
2. Способ минимального расстояния, основанный на Евклидовой метрике, когда в класс распределяются пиксели, минимально различающиеся по спектральной яркости объектов.
3. Способ максимального правдоподобия, который предполагает, что в каждом канале обычно распределяются статистические данные для каждого класса, и вычисляет вероятность, что данный пиксель принадлежит определенному классу. Если не выбирается порог вероятности, все пиксели будут классифицированы. Каждый пиксель относиться в тот класс, который имеет самую высокую вероятность (то есть, максимальная вероятность).

В программном комплексе ENVI представлен широкий набор алгоритмов для проведения классификации с обучением (рис.2):

- Способ параллелепипедов (Parallelepiped) (рис.3)
- Алгоритм минимального спектрального расстояния (Евклидова метрика, Minimum distance)
- Алгоритм расстояния Махаланобиса (Mahalanobis Distance)
- Способ максимального правдоподобия (Maximum likelihood)
- Классификация с помощью метода спектрального угла (Spectral Angle Mapper)
- Двоичное кодирование (Binary Encoding)



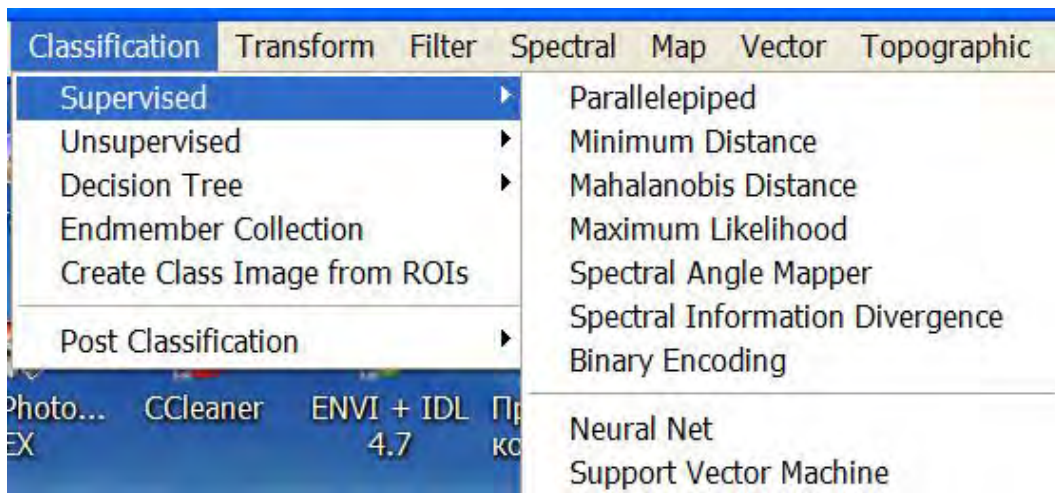


Рис.2. Алгоритмы классификации с обучением

Например, для способа параллелепипеда для всех классов выбирается порог отклонения от среднего и другие параметры (рис.3):

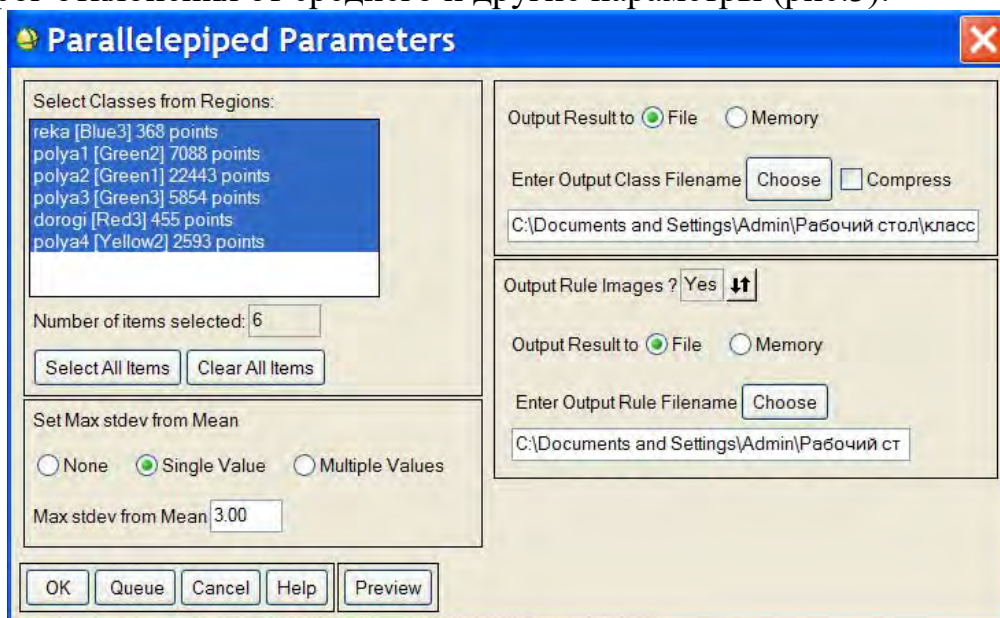


Рис.3. Окно системы ENVI для задания параметров при использовании способа параллелепипеда

Классификация без обучения (относится к цифровым автоматизированным методам обработки космических изображений), который позволяет выделять контуры с неконтрастной по спектральной яркости структурой (растительность, открытые почвы, вода, облака другие объекты.): K-Means, IsoDATA. Прежде чем приступить к классификации необходимо определить статистические характеристики изображения. Для этого выбирается пункт меню Process \ Calculate Statistics. Просмотреть и распечатать результаты посчитанной статистики можно через View \ Statistics \ Show Statistics. Можно сосчитать статистику по одному или нескольким регионам.

Чтобы запустить неконтролируемую классификацию нужно выбрать пункт меню Process \ Classification \ ISOCLASS Unsupervised Classification. Откроется окно Unsupervised Classification, где можно указать параметры процесса классификации: Autogenerate – начальное число классов,

Bands to use – каналы, которые будут участвовать в классификации,

Use classes – использование классов других классифицированных изображений,

Desired percent unchanged – процент сходимости итерационного процесса,

Maximum number of classes – максимальное число классов,

Minimum members in a class (%) – минимальное число членов класса,

Maximum standard deviation – максимальное стандартное отклонение, при котором класс будет делиться пополам при последующей итерации,

Min distance between class means – минимальное расстояние между классами, выраженное в стандартных отклонениях, при котором классы будут объединяться при последующей итерации.

После запуска процесса классификации открывается окно, в котором в режиме реального времени показывается процент сходимости при каждой итерации. Если процесс затянулся, его можно прервать, не дожидаясь, пока процент сходимости достигнет заданного значения.

Для оценки достоверности полученное в результате классификации изображение можно сравнить с другим отклассифицированным изображением, принятым за эталонное. Это могут быть результаты ручного дешифрирования, данные полевых исследований и т. п. С помощью View \ Statistics \ Confusion Matrix можно построить матрицу сравнения трех видов:

Raw Count – матрица числа пикселей, попавших в каждый класс обоих изображений,

User's Accuracy – матрица отношений числа пикселей одного изображения к числу пикселей другого, попавших в каждый класс,

Producer's Accuracy – матрица отношений правильно и неправильно отклассифицированных пикселей каждого класса.

Если требуется выделить только некоторые классы, то в окне Formula слоя Class Display нужно набрать формулу типа:

```
if (i1=1) OR (i1=5) OR (i1=22) then 1 else null.
```

Тогда требуемые классы (в примере 1, 5 и 22) будут выделены своими цветами, а остальные черным.

Если необходимо объединить несколько классов в один, это можно сделать с помощью формулы типа:

```
if (i1=1) OR (i1=5) OR (i1=22) then 1 else i1.
```

Тогда эти классы (например, 1, 5 и 22) будут показаны одним цветом (в примере цветом 1 классы).

Если пиксель в процессе классификации не отнесен ни к одному из классов, он показывается черным цветом. Эти пиксели можно отнести к одному из классов (в примере к 1 классу) с помощью формулы типа:

```
if (isnull(i1)) then 1 else i1.
```

После всех обработок и преобразований итоговое растровое изображение может быть переведено в векторную форму для его дальнейшего использования в ГИС-приложениях и создания цифровых карт. Создание векторного файла происходит в окне, открываемом с помощью Process \ Raster cells to vector polygons.

Были получены следующие результаты: 1. для особо охраняемой территории «Себежский национальный парк» (выполнил магистр Евдокимов П.А.), большую часть которого занимают рекреационные зоны и земли традиционного хозяйственного пользования, созданы векторные слои баз данных (территория покрытая водой, слабо выраженное заболачивание территории, сильно выраженное заболачивание территории), выявлены границы и изменения болотных систем по изображениям системы Landsat, полученными 4 мая 1986г., 2 мая 2000г. и 6 мая 2011, определены возможные варианты дальнейшего развития болотных урочищ и фаций, а также отмечено уменьшение обводнённости болотного массива в пределах 15 %.

2. Были установлены типы лесов на тестовом участке Ярославской области, типы земель тестового участка (выполнил Сучков И.В.). Подсчитано изменение площади вырубок 2-х периодов ( 1992 год и 2002 год). космического снимка тестового участка к физической карте. С помощью алгоритма обработки ISODATA были выявлены типы земель и лесов данного тестового участка. Также была создана 3-х мерная проекция данной территории. Разработана методика выявления вырубок на территории тестовых участков Ярославской и Ленинградской областей.

3. Для выделения участков пожаров и гарей на тестовых участках Бурятии близ Баргузинского заповедника был использован метод эталонной классификации [4]. Применялась бинарная классификация. Площадь лесов потерянных из-за пожара составила 17260 кв. м.

4. Определены возможности развития оврагов на тестовом участке Курской области (выполнила Коренькова С.М.). Выбраны три космических снимка разных годов. Первый снимок сделан 7 января 2002 года съёмочной системой DigitalGlobe и взят из программы Google Earth. Вторым снимком сделан 9 января 2010 года съёмочной системой GeoEye-1 и взят с сайта [www.sovzond.ru](http://www.sovzond.ru). Третий снимок сделан 9 января 2011 съёмочной системой RapIDEye и взят из программы Google Earth.

Для анализа процесса эрозии по изображениям тестовых участков была выбрана классификация без обучения (автономная, Unsupervised Classification). Один из основных алгоритмов для классификации без обучения является ISODATA (iterative Self-Organizing Data Analysis Technique - итеративный самоорганизующийся способ анализа данных). Данный спо-

соб применяют для более точной, многошаговой обработки снимков. Этот способ требует, от аналитика выбора числа групп (кластеров), которые будут выделены на данных и задания ряда ограничивающих параметров минимальное количество пикселей в классе, число итераций, порог сходимости классов.

С помощью выбранного алгоритма ISODATA на тестовых участках выделены границы эрозии, произведен анализ изменения границ с течением времени, что дает возможность прогнозирования дальнейшего роста оврага. Сравнивая результаты обработок двух снимков алгоритмом ISODATA определено, что границы участков подвержены овражной эрозии, с 2002 года по 2010 год размер оврага увеличился на 53 гектара.

5. Был выполнен анализ растительности двух участков Калининградской области (выполнила Федотова А.С.): леса и сельхозугодий и определено, что для них наиболее точным в определении классов для первого снимка оказался способ минимальной дистанции, позволяющий выделить поля с уже начавшимся цветением, а для второго, залесённого - метод параллелепипедов, наименее подходит способ максимального правдоподобия, т. к. он наиболее грубо определяет классы (разница в точности 5 и 15 %), для анализа состояния песчаных дюн ООПТ «Куршская коса» использовались разные методы - необучаемой классификации (ISODATA), позволивший определить расположение дюнных гряд и морского побережья, изменение глубины дна Куршского залива, расположение на изображении лесных массивов и обучаемой классификации (метод параллелепипедов), позволивший выделить границы дюнных гряд и морского побережья, и изменение их границ за 4 года.

### **Библиографический список**

1. Шубина М.А., Попова Т.А., Бычкова И.А. Применение дистанционных методов в изучении особо охраняемых природных территорий. В кн.: Методы дистанционного зондирования земли при решении природоресурсных задач. СПб, ВСЕГЕИ. 2004, с.110-119.

2. Шубина М.А., Попова Т.А., Бычкова И.А. Материалы дистанционного зондирования как средство познания природной среды. В кн.: «Национальный парк «Себежский»:научно-исследовательская работа, охрана, экологическое просвещение и развитие экологического туризма». Материалы науч.-практ.конф. Псков, 2006, с.6-13.

3. Кронберг П. Дистанционное изучение Земли. М.Мир, 1998, 468с.

4. Шубина М.А., Симоненков Е.В. Выделение областей лесных пожаров на Landsat – изображении с помощью программы открытого доступа ENVI. В сб. «Информационные системы и технологии: теория и практика» Сб. науч. Трудов. Вып.3 часть 2. СПб, СПбГЛТА, 2011. с. 37-45.

## ОГЛАВЛЕНИЕ

<b>А.М.Заяц.</b> Итоги научно-исследовательской деятельности за 2012 год.....	3
<b>А.М.Заяц, А.А.Логачев.</b> Постановка задачи выбора варианта тушения лесных пожаров и математическая модель для ее реализации.....	8
<b>В. А. Богатырев, С.М.Алексанков, Д.В. Демидов, С.А. Башкова, В.Ф.Беззубов, А.В. Полякова, И.Ю. Голубев.</b> Надежность дублированного вычислительного комплекса при ограничении допустимых задержек обслуживания.....	14
<b>Н.П. Васильев, С.П.Хабаров.</b> Метод исключения выбросов в данных, полученных от измерительных приборов и датчиков.....	24
<b>И.В. Ганичев, С.В. Киселева, , Н.В. Лушкин.</b> Инструментальные средства информатизации образования.....	30
<b>Ю. А.Жук.</b> Робастная модель обнаружения аномалий с использованием модели засорения.....	33
<b>М.О. Лебедев.</b> Опыт создания динамической информационно-поисковой и регистрирующей системы.....	41
<b>И.В.Панфилов.</b> Оценка быстродействия мультипроцессорной вычислительной системы.....	44
<b>И.В.Панфилов.</b> Концепция защиты информации от компьютерных преступлений.....	48
<b>М.А.Шубина.</b> Использование методов классификации космических изображений для мониторинга особо охраняемых территорий.	51