

Министерство образования и науки РФ
ФГБОУ ВПО «Рыбинский государственный авиационный технический
университет имени П.А.Соловьева»

УТВЕРЖДАЮ

Проректор по науке и
инновациям

_____ Т.Д. Кожина

РАБОЧАЯ ПРОГРАММА

по дисциплине

Методы и алгоритмы анализа данных

для специальности

05.13.01 Системный анализ, управление и обработка информации
(в промышленности)

Виды занятий	Количество часов	Количество зачетных единиц
Лекции	30	0,8
Практические занятия	60	1,7
Самостоятельная работа	162	4,5
Всего:	180	7,0
Форма контроля	Зачет	

Рабочую программу составил:

Михайлов Н.Л..

Рабочая программа рассмотрена на заседании кафедры МПО ЭВС
«__» _____ 2011 г., протокол № _____

Заведующий кафедрой,

к.ф.-м.н., профессор

Шаров В.Г.

Рыбинск,

2011

Настоящая программа составлена в соответствии с действующим паспортом и утвержденным вузовским планом специальности научных работников 05.13.01 Системный анализ, управление и обработка информации (в промышленности).

Цель и задачи дисциплины

Подготовить аспиранта к исследовательской работе, связанной с прикладным анализом данных и принятием обоснованных решений на основе обработки и анализа данных.

1. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Введение

Предмет и задачи курса. Структура и содержание дисциплины, ее связь с другими дисциплинами учебного плана.

Способы представления и модели порождения экспериментальных данных. Измерительная информация и ее структура. Детерминированные и вероятностные модели сигналов. Представление сигналов во временной и частотной областях. Марковские модели. Дискретизация непрерывных моделей. Ошибки измерений и их свойства. Классификация задач обработки экспериментальных данных.

1.1. Основы теории статистических решений

Понятие о генеральной совокупности. Связь вероятностной модели с опытом: оценка и проверка. Описательная статистика. Плотности вероятности. Типовые распределения вероятностей. Эмпирические плотности вероятности. Метод моментов для оценки параметров плотностей распределения. Статистические гипотезы. Критерии: z – критерий, t – критерий, F – критерий, χ^2 – критерий, критерий Фишера, критерий Неймана – Пирсона, критерий максимума правдоподобия, критерий максимума апостериорной вероятности. Ранговая статистика. Критерий Колмогорова-Смирнова.

1.2. Корреляционный анализ

Одномерный корреляционный анализ. Многомерный статистический анализ. Корреляционный анализ. Понятие о дискриминантном и факторном анализе. Метод главных компонент. Канонические корреляции.

1.3. Регрессионные модели. Метод наименьших квадратов.

Определение параметров регрессионных моделей как задача статистического оценивания. Классификация оценок. Алгоритмы определения параметров по методу наименьших квадратов для линейных регрессионных моделей. Свойства оценок наименьших квадратов. Ортогональность и сингулярные разложения. Рекуррентный алгоритм метода наименьших квадратов. Определение параметров нелинейной регрессии. Пакеты прикладных программ регрессионного анализа данных.

1.4. Интерполяция и сглаживание экспериментальных данных методами сплайн-функций

Задачи интерполяции и сглаживания. Полиномиальная интерполяция и ее ограничения. Интерполяция и сглаживание кубическими сплайнами. Базисные сплайны (B-сплайны). Интерполяция B-сплайнами. Среднеквадратическая сплайн-аппроксимация с использованием B-сплайнов. Алгоритмы аппроксимации по полной выборке и рекуррентные алгоритмы аппроксимации. Пакеты прикладных программ по аппроксимации сплайнами.

1.5. Временные ряды. Методы анализа и прогнозирования временных рядов: модели стационарных и нестационарных временных рядов, идентификация модели временного ряда, алгоритмы сглаживания временных рядов. Оценивание длины периода и периодической составляющей.

1.6. Оценивание состояния процессов. Алгоритмы оптимальной фильтрации

Задачи экстраполяции, фильтрации и сглаживания сигналов как задачи оценивания состояния динамической системы. Дискретный фильтр Калмана и его свойства. Фильтрация стационарных сигналов. Фильтрация

сигналов для случая нелинейных моделей. Линеаризованный и расширенный фильтры Калмана. Субоптимальные фильтры. Расходимость процесса фильтрации. Методы предотвращения расходимости. Адаптивная фильтрация. Робастный фильтр Калмана.

1.7. Кластерный анализ. Иерархические и неиерархические методы кластерного анализа. Методы определения расстояния между кластерами. Алгоритм k-средних. Проверка качества кластеризации. Модифицированные алгоритмы кластерного анализа: BIRCH, CLARA, CURE, CHAMELEON, ROCK, WaveCluster.

1.8. Факторный анализ. Метод главных компонент. Методы детерминированного и стохастического факторного анализа

1.9. Алгоритмы спектрального анализа сигналов

Спектральное представление сигналов. Дискретизация по времени и частоте. Маскировка частот. Дискретное преобразование Фурье и его свойства. Алгоритмы быстрого преобразования Фурье и их классификация. Утечка мощности и сглаживающие окна. Оценка спектра мощности сигнала. Wavelet-анализ.

Заключение

Перспективы развития методов, алгоритмов и программного обеспечения обработки экспериментальных данных.

2. ПЕРЕЧЕНЬ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

(Лабораторных работ, семинарских занятий)

2.1. Оценка показателей качества объекта по результатам эксперимента.

2.2. Разработка регрессионной модели объекта по результатам эксперимента.

2.3. Анализ временных рядов. Определение наличия тренда и периодической составляющей

2.4. Алгоритмы спектрального анализа. Быстрое преобразование Фурье. Алгоритм Кули-Тьюки.

2.5. Кластерный анализ. Алгоритм k-средних.

2.6. Кластерный анализ. Реализация алгоритмов BIRCH, CHAMELEON, ROCK, WaveCluster.

2.7. Интерполяция и сглаживание экспериментальных данных методами сплайн-функций. Алгоритмы аппроксимации по полной выборке и рекуррентные алгоритмы аппроксимации.

2.8. Факторный анализ. Выделение факторов методом главных компонент.

2.9. Исследование частотно-временной динамики нестационарных процессов на основе Wavelet-анализ.

2.10. Предварительная обработка экспериментальных данных. Алгоритмы обнаружения и устранения грубых промахов.

2.11. Методы идентификации закона распределения экспериментальных данных.

3. ПЕРЕЧЕНЬ ТЕМ КУРСОВЫХ РАБОТ

(Курсового проекта, расчетно-графической работы)

Не предусмотрено

4. СПИСОК ЛИТЕРАТУРЫ И ПЕРЕЧЕНЬ ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ

Основной

4.1. Задорина Н.А. Обработка экспериментальных данных на ЭВМ. Учебное пособие. – Рыбинск: РГАТА, 2008.

4.2. Глинский В.В. Статистический анализ: Учебное пособие. – М.: ИНФРА-М, 2002.

4.3. Вишняков В.А. Статистическая обработка информации: Пособие. – Рыбинск, РГАТУ, 2011.

Дополнительный

4.4. Макарова Н.В., Трофимова В.Я. Статистика в EXCEL. Учебное пособие. – М.: Финансы и статистика, 2003.

4.5. Бендат Дж., Пирсол А. Прикладной анализ случайных данных. – М.: Мир, 1989.

4.6. Тихонов А. Н., Уфимцев М. В. Статистическая обработка результатов экспериментов: Учебное пособие. – М.: МГУ, 1988.

4.7. Отнес Р., Эноксон Л. Прикладной анализ временных рядов. – М.: Мир, 1982.

4.8. Шураков В. В., Дайнтбегов Д. М., Мизрохи С. В., Ясеновский С.В. Автоматизированное рабочее место для статистической обработки данных. – М.: Финансы и статистика, 1990.

4.9. Тюрин Ю.Н. Статистический анализ данных на компьютере. – М.: ИНФРА-М, 1998.

5. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ИЗУЧЕНИЮ ДИСЦИПЛИНЫ

Дисциплина является базовой для выполнения аналитической работы с результатами испытаний разрабатываемых методов, алгоритмов и экспериментальными данными. Теоретический курс изучается аспирантом в процессе работы на лекциях. При этом аспирант конспектирует излагаемый преподавателем материал, отвечает на вопросы, которые ставит преподаватель в процессе чтения лекции. Перед лекцией рекомендуется просмотреть материалы предыдущих лекций по данной дисциплине. Кроме того, изучение курса связано и с самостоятельной работой аспиранта с рекомендуемой литературой.

СПИСОК ЭКЗАМЕНАЦИОННЫХ (ЗАЧЕТНЫХ) ВОПРОСОВ

- 6.1. Классификация задач обработки данных
- 6.2. Точечные оценки. Доверительные оценки. Свойства оценок
- 6.3. Виды ошибок
- 6.4. Процедуры проверки статистических гипотез
- 6.5. Методы получения оценок
- 6.6. Метод наименьших квадратов
- 6.7. Метод моментов
- 6.8. Метод максимального правдоподобия
- 6.9. Метод максимума апостериорной плотности вероятности
- 6.10. Понятие статистической связи. Корреляция. Корреляционное поле
- 6.11. Коэффициент корреляции. Корреляционные отношения.
- 6.12. Многомерный корреляционный анализ. Ковариационные и корреляционные матрицы
- 6.13. Z-критерий для оценки значимости коэффициента корреляции
- 6.14. Постановка задачи фильтрации Калмана
- 6.15. Структура фильтра Калмана
- 6.16. Свойства оценок, полученных с помощью фильтров Калмана
- 6.17. Синтез фильтра Калмана для оценки постоянной величины
- 6.18. Синтез фильтра Калмана для оценки низкочастотных данных
- 6.19. Синтез фильтра Калмана для оценки низкочастотных квадратур узкополосного случайного процесса.
- 6.20. Анализ эффективности фильтров Калмана.
- 6.21. Анализ устойчивости фильтров Калмана
- 6.22. Методы повышения устойчивости фильтров Калмана
- 6.23. Постановка задачи регрессионного анализа
- 6.24. Решение задачи регрессионного анализа на конкретных данных.
Подбор прямой, правила анализа остатков
- 6.25. Процедура регрессионного анализа без повторных данных

- 6.26. Процедура регрессионного анализа с повторными данными
- 6.27. Планирование экспериментов для повышения точности регрессионной модели
- 6.28. Особенности многомерного регрессионного анализа данных
- 6.29. F-критерии и t-критерии для оценки значимости регрессии и оценки адекватности регрессионной модели
- 6.30. χ^2 -критерий
- 6.31. Спектральный анализ в Excel
- 6.32. Спектральный анализ в MathCAD
- 6.33. Интерпретация результатов спектральной обработки данных
- 6.34. Алгоритмы сглаживания временных рядов
- 6.35. Иерархические методы кластерного анализа.
- 6.36. Методы определения расстояния между кластерами
- 6.37. Алгоритм k-средних
- 6.38. Факторный анализ
- 6.39. Метод главных компонент
- 6.40. Методы детерминированного и стохастического факторного анализа

7. КОНТРОЛЬНЫЕ ВОПРОСЫ (ЗАДАЧИ ИЛИ ТЕСТЫ САМОПРОВЕРКИ)

1. В чем отличие интервальных оценок от точечных?
2. В чем суть метода наименьших квадратов?
3. Преобразуйте следующие статистические модели так, чтобы можно было применить МНК
 - 3.1. $y_i = b_0 + b_1\sqrt{x_i}$
 - 3.2. $y_i = b_0 + \frac{b_1}{x_i} + \frac{b_2}{x_i^2}$
 - 3.3. $y_i = b_1(1 + x_i)^{\frac{3}{2}}$
 - 3.4. $y_i = b_0(x_i + b_1)$
 - 3.5. $y_i = \sin(b_1x_i) + b_0$
4. Сформулируйте задачу регрессионного анализа

5. Решение задачи регрессионного анализа на конкретных данных. Подбор прямой, правила анализа остатков

Имеются следующие данные о годовых ставках месячных доходов по акциям за шестимесячный период:

Акция	Доходы по месяцам, %					
	А	5,4	5,3	4,9	4,9	5,4
В	6,3	6,2	6,1	5,8	5,7	5,7

Есть основания предполагать, что доходы Y по акции В зависят от доходов X по акциям А . Необходимо составить уравнение регрессии Y по X

6. В чем особенности многомерного регрессионного анализа данных?

7. Для приведенного в таблице набора данных выполнить объединение данных в 3 кластера, используя иерархический метод объединения. Для вычисления расстояния между кластерами использовать метод ближнего соседа. Процесс объединения отобразить в виде дендрограммы.

Таблица Набор данных

№	признак X	признак Y
1	27	19
2	11	46
3	25	15
4	36	27
5	35	25
6	10	43
7	11	44
8	36	24
9	26	14
10	26	14