

Правительство Российской Федерации

**Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
«Национальный исследовательский университет
«Высшая школа экономики»**

Факультет Компьютерных наук
Департамент больших данных и информационного поиска
Базовая кафедра Яндекс

УТВЕРЖДАЮ
Академический руководитель
образовательной программы
«Науки о данных»
по направлению 01.04.02
«Прикладная математика и информатика»
С.О. Кузнецов

«__» _____ 2015 г.

**Программа дисциплины «Вероятностно-статистические методы в теории
принятия решений»**

для направления 01.04.02 "Прикладная математика и информатика" подготовки
магистра
для магистерской программы "Науки о данных"

Автор программы:

Пузыревский И.В. (ipuzrevskiy@hse.ru)

Одобрена на заседании базовой кафедры Яндекс «__» _____ 2015 г.

Заведующий кафедрой _____ М.А. Бабенко

Рекомендована Академическим советом образовательной программы
«Науки о данных» «__» _____ 2015 г.

Менеджер базовой кафедры Яндекс _____ И.И. Алескерова

Москва, 2015

*Настоящая программа не может быть использована другими подразделениями
университета и другими вузами без разрешения подразделения разработчика программы.*



Пояснительная записка

Автор программы

Пузыревский И.В.

Аннотация

Дисциплина «Вероятностно-статистические методы в теории принятия решений» предназначена для подготовки магистров 01.04.02 – Прикладная математика и информатика.

За последние двадцать лет существенно возросла потребность в решении ряда практических задач таких как автоматическое обнаружение неисправностей (разладок, сбоев, аномалий и т.п.), обслуживание оборудования на основе автоматического контроля его состояния, обеспечение безопасности сложных технических и информационных систем (самолетов, судов, ракет, ядерных электростанций, различных интернет сервисов, и т.д.), автоматический контроль качества выпускаемой продукции, предсказание естественных катастрофических явлений (землетресения, цунами, и т.д.), мониторинг в биомедицине и финансовой сфере.

Основная черта вышеперечисленных задач состоит в том, что по сути все они сводятся к выявлению момента резкого изменения (разладки) некоторых характеристик рассматриваемого объекта на основе статистических данных о других характеристиках этого объекта и/или детектированию наблюдений, которые по тем или иным статистическим свойствам значительно отличаются от большинства наблюдений из исследуемой генеральной совокупности.

С развитием информатики появилась возможность построения автоматизированных информационных систем для статистической обработки огромного объема реальных данных с целью вынесения тех или иных суждений об истинных характеристиках процесса. Для создания таких систем с привлечением программных средств требуется прежде всего разработка соответствующих фундаментальных математических методов обработки поступающей и поступившей информации исходя из естественных критериев оптимальности. Именно поэтому в данном курсе:

1. Рассматривается общая теория оптимальной остановки стохастических процессов, позволяющая строить оптимальные методы обработки поступающей и поступившей информации, в том числе и методы скорейшего обнаружения разладки.
2. Рассматриваются особенности использования методов скорейшего обнаружения разладки и детектирования аномалий для решения практически задач в различных областях науки и техники.

Предлагаемый курс позволит студентам с одной стороны ознакомиться с основными теоретическими свойствами алгоритмов скорейшего обнаружения разладки, а с другой стороны познакомит их с особенностями практической реализации и примерами применения современных алгоритмов скорейшего обнаружения разладки и детектирования аномалий.

Программа курса предусматривает лекции (30 часов) и практические занятия (30 часов).



Учебные задачи курса

Целью данного курса является ознакомление слушателей с общей теорией оптимальной остановки стохастических процессов, научить их строить оптимальные методы обработки поступающей и поступившей информации, в том числе и методы скорейшего обнаружения разладки. Кроме того, слушатели должны получить понятие об использовании методов скорейшего обнаружения разладки и детектирования аномалий для решения практически задач в различных областях науки и техники.

Тематический план дисциплины «Вероятностно-статистические методы в теории принятия решений»

№	Название темы	Всего часов по дисциплине	Аудиторные часы		Самостоятельная работа
			Лекции	Сем. и практика	
1	Основы теории скорейшего обнаружения разладки	38	6	6	20
2	Практическая реализация стандартных статистик для скорейшего обнаружения разладки	38	6	6	30
3	Некоторые избранные методы обработки и моделирования временных рядов	38	6	6	20
4	Модификации стандартных методов скорейшего обнаружения для случая неизвестных параметров сигнала после разладки	38	6	6	30
5	Обнаружение аномалий	38	6	6	30
	Итого	190	30	30	130



I. Источники информации

Список литературы

Основная литература

1. Вальд А. Последовательный анализ: Пер. с англ. / Севастьянов Б.А., ред.- М.: Физматлит, 1960. - 328 с.
2. Леман Э. Проверка статистических гипотез. Изд-во "Наука", М., 1964.
3. Ширяев А.Н. Статистический последовательный анализ: Оптимальные правила остановки. - 2-е изд., перераб. - М.: Наука. Физматлит, 1976. - 272 с. - (Теория вероятностей и математическая статистика).
4. Ширяев А.Н. Вероятность-1,2. 4-ое издание. М.:МЦМНО, 2007.
5. Н. Клигине, Л. Телькснис. Методы обнаружения моментов изменения свойств случайных процессов (обзор) // Автоматика и телемеханика. 1983. № 10. С. 5-56.
6. И.В. Никифоров Последовательное обнаружение изменения свойств временных рядов. М.: Наука, 1983.
7. М. Бассевиль, А. Банвенист. Обнаружение изменения свойств сигналов и динамических систем. М.: Наука, 1989.
8. А.Г. Тартаковский. Последовательные методы в теории информационных систем. М.: Радио и связь, 1991.
9. Peskir Goran, Shiryaev Albert. Optimal stopping and free-boundary problems. Basel etc.: Birkhauser-Verl., 2006. - 500 с. - Bibliogr.: p. 477-492.
10. M. Basseville, I. V. Nikiforov. Detection of abrupt changes: theory and application. Prentice-Hall, 1993.
11. D. Siegmund Sequential analysis : Tests and confidence intervals.- New York etc.: Springer-Verl., 1985. - 272 pp. - (Springer series in statistics). - Bibliogr.: pp.263-270. - ISBN 0-387-96134-8.
12. A.S. Willsky. A survey of design methods for failure detection in dynamic systems // Automatica. 1976. V. 12. P. 601-611. 2
13. B.K. Ghosh, P.K. Sen. Handbook of sequential analysis. Marcel Dekker, Inc., 1991.
14. E. Carlstein, H.-G. Muller, D. Siegmund (eds.). Change-point problems. Papers presented at the AMS-IMS-SIAM Summer Research Conference on Change-Point Problems, Mt. Holyoke College, July 11-17, 1992.

II. Формы контроля и структура итоговой оценки

Текущий контроль - домашняя работа в первом модуле, контрольная работа в первом модуле.



Итоговый контроль – письменный экзамен (120 мин.)

Итоговая оценка вычисляется следующим образом:

$0,1 \cdot \text{оценка за домашнюю} + 0,2 \cdot \text{оценка за контрольную} + 0,7 \cdot \text{оценка за экзамен.}$

Таблица соответствия оценок по десятибалльной и системе зачет/незачет

Оценка по 10-балльной шкале	Оценка по 5-балльной шкале
1	Незачет
2	
3	
4	Зачет
5	
6	
7	
8	
9	
10	

Таблица соответствия оценок по десятибалльной и пятибалльной системе

По десятибалльной шкале	По пятибалльной системе
1 – неудовлетворительно 2 – очень плохо 3 – плохо	неудовлетворительно – 2
4 – удовлетворительно 5 – весьма удовлетворительно	удовлетворительно – 3
6 – хорошо 7 – очень хорошо	хорошо – 4
8 – почти отлично 9 – отлично 10 – блестяще	отлично – 5

III. Программа дисциплины «Вероятностно-статистические методы в теории принятия решений»

Тема 1. Основы теории скорейшего обнаружения разладки

- Обзор основных задач в теории принятия решений
- Ключевые статистики и тесты в теории принятия решений в задачах различения двух гипотез по фиксированному числу наблюдений. Дискретное время

- Ключевые статистики и последовательные тесты в задачах различения двух гипотез. Дискретное время
- Некоторые широко используемые статистики в проблеме скорейшего обнаружения
- Об определении броуновского движения, производной Радона-Никодима и условного математического ожидания
- О различении двух гипотез для броуновского движения. Сравнение методов Неймана-Пирсона и Вальда
- Об основных формулировках задач скорейшего обнаружения для броуновского движения (Варианты А, В, С, D и E)
- "Вальдовская" система наблюдения в варианте E
- "Нейман-Пирсоновская" система наблюдения в варианте В
- Постановка задачи о "разладки" в варианте А и ее редукция к задаче об оптимальной остановке для марковского процесса
- Решение задачи об оптимальной остановке для марковского процесса в варианте А

Тема 2. Практическая реализация стандартных статистик для скорейшего обнаружения разладки

- Контрольные карты Шухарта
- CUSUM, "двухсторонний" CUSUM
- Ширяев-Робертс
- Модификация стандартных статистик для выявления разладки в дисперсии и ковариационной структуре (случай многомерного сигнала)

Тема 3. Некоторые избранные методы обработки и моделирования временных рядов

- Моделирование временных рядов на основе моделей типа ARMA
- Обработка сигналов: очистка от шума, сегментация сигналов на основе скрытых марковских моделей, сглаживания с адаптивными весами
- Экспоненциальное сглаживание: оценки среднего значения и параметра масштаба, экспоненциальное сглаживание в случае неравномерного расположения наблюдений, метод Хольта-Винтерса, робастный вариант экспоненциального сглаживания

Тема 4. Модификации стандартных методов скорейшего обнаружения для случая неизвестных параметров сигнала после разладки



- Практическая реализация стандартных методов скорейшего обнаружения. Статистическое моделирование распределения ложных тревог
- Байесовские предположения о параметрах распределения после разладки
- Использование ОМП для задания значений параметров распределения после разладки
- Мгновенная оценка среднего значения после разладки на основе текущего наблюдения
- Оценка среднего значения и параметра масштаба на основе экспоненциального сглаживания
- Численное оценивание среднего времени до ложной тревоги и среднего времени до обнаружения разладки для стандартных алгоритмов скорейшего обнаружения
- Настройка параметров алгоритмов скорейшего обнаружения

Тема 5. Обнаружение аномалий

- Основные методы обнаружения аномалий и их теоретические свойства
- Примеры использования методов наискорейшего обнаружения и детектирования аномалий

IV. Методические указания студентам

Самостоятельная работа студента предусматривает написание программ для работы с кластером и практическое использование изучаемых архитектур.

Автор программы: _____ / <Пузыревский И.В.> /