

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
«Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»
(ВлГУ)



Проректор
по учебно-методической работе
А.А. Панфилов

« 2 » 04 2015 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

«ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ НА ОСНОВЕ ХРАНИЛИЩ ДАННЫХ»

Направление подготовки 09.03.03 *Прикладная информатика*

Программа подготовки *Прикладная информатика в экономике*

Уровень высшего образования *бакалавриат*

Форма обучения *очная*

Семестр	Трудоем- кость зач, ед, час.	Лек- ций, час.	Практик. занятий, час.	Лаборат. работ, час.	СРС, час.	Форма промежуточного контроля (экз./зачет)
7	5/180	18	36	18	72	экзамен (36 час.)
Итого	5/180	18	36	18	72	экзамен (36 час.)

Владимир, 2015

1. ЦЕЛЬ И ЗАДАЧИ ДИСЦИПЛИНЫ

Курс должен служить руководством по изучению теоретических основ и практическому освоению работы с интеллектуальным анализом данных. Задачей изучения дисциплины является формирование практических навыков по использованию интеллектуальных систем для управления данными.

Задачей изучения дисциплины является формирование теоретических знаний, практических умений и навыков по применению современных методов интеллектуального анализа данных в различных сферах человеческой деятельности. Цели изучения дисциплины:

- изучение существующих технологий подготовки данных к анализу;
- изучение основных методов поиска закономерностей, связей, правил в табулированных массивах данных большого объема; иллюстрированного их применения в различных областях деятельности;
- овладение практическими умениями и навыками реализации технологий интеллектуального анализа данных, формирования и проверки гипотез о их природе и структуре, варьирования применяемыми моделями;
- формирование умений и навыков применения универсальных программных пакетов и аналитических платформ для анализа данных.

Достижение названных целей предполагает **решение следующих задач:**

- ознакомление с основами искусственного интеллекта;
- определение места интеллектуального анализа среди других управляющих систем и обзор современных технологий;
- ознакомление с конкретными примерами прикладных интеллектуальных систем анализа данных;
- изучение методов интеллектуального анализа данных и программного инструментария создания интеллектуальных систем;
- приобретение студентами навыков постановки прикладных задач в терминах изучаемых интеллектуальных методов;
- приобретение студентами навыков решения прикладных задач с использованием интеллектуального программного обеспечения.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

Дисциплина «Интеллектуальный анализ на основе хранилищ данных» относится к дисциплинам вариативной части учебного плана.

К числу дисциплин, связанных с «Интеллектуальный анализ на основе хранилищ данных», относятся «Информационные системы и технологии», «Базы данных», «Интеллектуальные информационные системы», «Компьютерные системы поддержки принятия решений». В результате освоения этих дисциплин студенты приобретают необходимые для изучения данной дисциплины, знания основных понятий информационных систем и технологий; методов и средств интеллектуального моделирования; принципов обработки данных и получения знаний.

В результате изучения дисциплины полученные знания будут применены при выполнении выпускной квалификационной работы и в дальнейшей профессиональной деятельности.

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ

В процессе освоения данной дисциплины студент формирует и демонстрирует следующие компетенции:

- способность решать стандартные задачи профессиональной деятельности на основе информационной и библиографической культуры с применением информационно-коммуникационных технологий и с учетом основных требований информационной безопасности (ОПК-4).

Иметь представление: об основных терминах и понятиях искусственного интеллекта, о месте ИАД среди других технологий информационных систем, о конкретных методах и программных системах интеллектуального анализа данных.

Знать: структуру и общую схему функционирования интеллектуального анализа данных (ИАД), методы представления знаний и методы интеллектуального моделирования в ИАД, области применения ИАД, этапы решения задач интеллектуальными методами, методы и инструментальные средства ИАД.

Уметь: управлять знаниями в условиях формирования и развития информационного общества: анализировать, синтезировать и критически резюмировать и представлять информацию; выбрать форму представления знаний или метод интеллектуального моделирования и инструментальное средство ИАД для конкретной предметной области, проектировать, используя математические методы, интеллектуальную модель в соответствии с выбранным методом, решать задачи и интерпретировать результаты.

Владеть: навыками работы с нейросетевой оболочкой BrainMaker Pro, с инструментами gatool и fuzzytool в системе Matlab, системой Business Intelligence, аналитической платформой «Deductor Studio», инструментом интеллектуального анализа «PolyAnalyst».

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет 5 зачетных единиц, 180 часов.

№ ПП	Раздел (тема) дисциплина	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)						Объем уч работы с применением интерактивных методов (в час/%)	Формы текущего контроля успеваемости. Форма промежуточной аттестации
				Лекции	Практические занятия	Лабораторные работы	Контрольные работы	СРС	КП/КР		
1	Технологии обеспечения качественного анализа данных.	7	1-2	2						1 /500%	
2	Инструменты Data mining	7	3-6	2	8	4		24		7/50 %	1 р-к
3	Классификация и регрессия. Машинное обучение	7	7-14	10	20	8		24		19/50 %	2 р-к
4	Анализ и прогнозирование временных рядов. Ансамбли моделей. Сравнение моделей.	7	15-18	4	8	6		24		9/50%	3 р-к
Всего				18	36	18		72		36/50%	3 р-к, экзамен

СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Раздел 1. Технологии обеспечения качественного анализа данных

Тема 1. Технологии анализа данных

Аналитический и информационный походы к моделированию. Формы представления, типы и виды анализируемых данных. Источники данных для анализа. Технология KDD (Knowledge Discovery in Databases). Введение в технологию Data Mining. Программный инструментарий анализа данных.

Тема 2. Консолидация данных

Введение в консолидацию данных. Общая характеристика OLTP-систем. Предпосылки появления систем поддержки принятия решений (DSS – Decision Support System). Введение в хранилища данных (DW – Data Warehouse). Реляционные хранилища данных (ROLAP – Relational OLAP). Многомерные хранилища данных (MOLAP – Multidimensional OLAP). Гибридные хранилища данных (HOLAP – Hybrid OLAP). Виртуальные хранилища данных. Введение в процесс ETL (Extraction, Transformation, Loading). Извлечение данных в ETL. Преобразование данных в ETL.

Тема 3. Трансформация данных. Визуализация данных

Введение в трансформацию данных. Особенности трансформации временных рядов. Группировка и разгруппировка данных. Слияние данных. Квантование. Нормализация и кодирование данных. Введение в визуализацию данных. Визуализаторы общего назначения. OLAP-анализ. Визуализаторы, применяемые для оценки качества моделей. Визуализаторы, применяемые для интерпретации результатов анализа.

Тема 4. Способы визуализации данных

Тема 5. Оценка качества, очистка и предобработка данных

Введение в оценку качества данных (ADQ – Assessment Data Quality). Технологии и методы оценки качества данных. Очистка и предобработка данных. Фильтрация данных. Обработка дубликатов и противоречий. Выявление аномальных значений. Восстановление пропущенных значений. Введение в сокращение размерности. Сокращение числа признаков. Сокращение числа значений признаков и записей. Сэмплинг.

Раздел 2. Инструменты Data mining

Тема 1. Поиск ассоциативных правил. Кластеризация

Введение в аффинитивный анализ (affinity analysis). Алгоритм a priori. Иерархические ассоциативные правила.

Введение в кластеризацию. Классификация методов кластеризации. Алгоритм кластеризации k-means. Сети Кохонена (KCN – Kohonen network). Карты Кохонена (SOM – self organizing map). Проблемы алгоритмов кластеризации

Тема 2. Классификация и регрессия. Статистические методы

Введение в классификацию и регрессию. Простая линейная регрессия. Оценка соответствия простой линейной регрессии реальным данным. Простая регрессионная модель. Множественная линейная регрессия. Модель множественной линейной регрессии. Регрессия с категориальными входными переменными. Методы отбора переменных в регрессионные модели. Ограничения применимости регрессионных моделей. Основы логистической регрессии. Интерпретация модели логистической регрессии. Множественная логистическая регрессия.

Тема 3. Классификация и регрессия. Машинное обучение

Введение в деревья решений. Алгоритмы построения деревьев решений. Алгоритмы ID3 и C4.5. Алгоритм CART. Упрощение деревьев решений. Введение в нейронные сети. Искусственный нейрон. Принципы построения нейронных сетей. Алгоритмы обучения нейронных сетей. Алгоритм обратного распространения ошибки.

Тема 4. Анализ и прогнозирование временных рядов. Ансамбли моделей. Сравнение моделей.

Временной ряд и его компоненты. Модели прогнозирования.

Введение в ансамбли моделей. Бэггинг. Бустинг. Альтернативные методы построения ансамблей.

Оценка эффективности и сравнение моделей. Оценка ошибки модели. Издержки ошибочной классификации. Lift- и Profit-кривые. ROC-анализ.

Тематика практических занятий

- Демонстрационные примеры решения задач методом генетических алгоритмов.
- Демонстрационные примеры решения задач методом нейросетей.
- Аналитическая платформа «Deductor Studio». Методы подготовки данных.
- Аналитическая платформа «Deductor Studio». Методы трансформации данных.
- Аналитическая платформа «Deductor Studio». Деревья решений.
- Аналитическая платформа «Deductor Studio». Самоорганизующиеся карты Кохонена.
- Аналитическая платформа «Deductor Studio». Нейронные сети.
- Аналитическая платформа «Deductor Studio». Линейная регрессия.
- Аналитическая платформа «Deductor Studio». Логистическая регрессия.
- Аналитическая платформа «Deductor Studio». Ассоциативные правила.
- Аналитическая платформа «Deductor Studio». Пользовательские модели.

Тематика лабораторных занятий

- Аналитическая платформа «Deductor Studio». Решение задач классификации.
- Аналитическая платформа «Deductor Studio». Решение задач поиска ассоциаций.
- Аналитическая платформа «Deductor Studio». Решение задач прогнозирования.
- Нейронные сети: программа Brain Maker Pro.
- Генетические алгоритмы: инструмент «gatoool» в системе Matlab.
- Нечеткая логика: инструмент «fuzzytool» в системе Matlab.
- Программный комплекс PolyAnalist. Поиск зависимостей.

Перечень используемых пакетов программ: Brain Maker Pro, «Deductor Studio», демонстрационные примеры, система Matlab, программный комплекс PolyAnalist.

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

№ п/п	Виды учебной работы	Образовательные технологии
1.	Лекции	- лекция-информация с визуализацией; - использование мультимедиа оборудования
2.	Практические и лабораторные занятия	- выполнение лабораторных и практических работ; - поиск и анализ информации в справочных системах и сети Интернет; - групповые обсуждения, - имитационные (ситуативные) технологии; - анализ конкретных ситуаций и поиск путей решения
3.	Самостоятельная работа	- опережающая самостоятельная работа; - внеаудиторная работа студентов (освоение теоретического материала, подготовка к практическим и лабораторным, работа с электронным учебно-методическим комплексом, подготовка к текущему и итоговому контролю) - доступа к методическим материалам
3.	Контроль	- рейтинг-контроль

6. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Для оценки текущего контроля успеваемости предусмотрено три рейтинг - контроля, проводимых согласно принятому в университете графику.

Промежуточной формой аттестации знаний студентов является экзамен.

Рейтинг-контроль знаний студентов

Вопросы к рейтинг-контролю № 1

- 1 История развития интеллектуальных систем.
- 2 Основные понятия интеллектуального анализа данных.
- 3 Классификация технологий интеллектуального анализа данных.
- 4 Этапы интеллектуального анализа данных.
- 5 Методы очистки данных. Редактирование аномалий.
- 6 Методы очистки данных. Заполнение пропусков.
- 7 Методы очистки данных. Очистка от шумов.
- 8 Методы очистки данных. Сглаживание.
- 9 Методы очистки данных. Поиск дубликатов и противоречий.
- 10 Методы трансформации данных. Преобразование к скользящему окну.
- 11 Методы трансформации данных. Квантование.
- 12 Методы трансформации данных. Группировка и сортировка.
- 13 Методы трансформации данных. Приведение типов.

Вопросы к рейтинг-контролю № 2

- 1 Методы и модели интеллектуального анализа данных.
- 2 Методы первичной проверки гипотез.
- 3 Методы построения числовых моделей и прогноза числовых переменных.
- 4 Полиномиальная нейронная сеть.
- 5 Пошаговая многопараметрическая линейная регрессия.
- 6 Метод «ближайших соседей».
- 7 Классификатор на основе нечеткой логики.
- 8 Статистическая дискриминация.
- 9 Деревья решений.
- 10 Леса решений.
- 11 Самообучающиеся карты Кохонена.
- 12 Эволюционное моделирование.

Вопросы к рейтинг-контролю № 3

- 1 Нечеткие когнитивные схемы.
- 2 Детерминационный анализ.
- 3 Алгоритмы ограниченного перебора
- 4 Поиск ассоциаций.
- 5 Системы поиска на основе аналогичных случаев.
- 6 Логистическая регрессия.
- 7 Предметно-ориентированные аналитические системы.
- 8 Искусственная жизнь и клеточные автоматы.
- 9 Обзор программных систем интеллектуального анализа данных.
- 10 Аналитическая платформа Deductor .
- 11 MicroSoft Buizness Intellegence Suite.
- 12 Программный комплекс PolyAnalist.
- 13 Способы визуализации данных.
- 14 Интерпретация результатов и аналитические отчеты.

Вопросы к самостоятельной работе (темы)

- 1 Microsoft Business Intelligence Suite. Алгоритм взаимосвязей.
- 2 Microsoft Business Intelligence Suite. Алгоритм дерева принятия решений.
- 3 Microsoft Business Intelligence Suite. Алгоритм кластеризации.
- 4 Microsoft Business Intelligence Suite. Алгоритм кластеризации последовательностей.
- 5 Microsoft Business Intelligence Suite. Алгоритм линейной регрессии.
- 6 Microsoft Business Intelligence Suite. Алгоритм логистической регрессии.
- 7 Microsoft Business Intelligence Suite. Упрощенный алгоритм Байеса.
- 8 Microsoft Business Intelligence Suite. Алгоритм нейронной сети.
- 9 STATISTICA Data Miner. Отсеивание признаков и фильтрация переменных.
- 10 STATISTICA Data Miner. Правила связей.
- 11 STATISTICA Data Miner. Интерактивное бурение данных.
- 12 STATISTICA Data Miner. Обобщенный EM и кластерный анализ методом k-средних.
- 13 STATISTICA Data Miner. Обобщенные аддитивные модели (GAM).
- 14 STATISTICA Data Miner. Общие деревья классификации и регрессии (GTrees).
- 15 STATISTICA Data Miner. Общие модели CHAID.
- 16 STATISTICA Data Miner. Интерактивные деревья классификации и регрессии.
- 17 STATISTICA Data Miner. Растущие деревья.
- 18 STATISTICA Data Miner. Случайные леса.
- 19 STATISTICA Data Miner. Машинное обучение – метод опорных векторов.
- 20 STATISTICA Data Miner. Машинное обучение – метод k-ближайших соседей.
- 21 STATISTICA Data Miner. Многомерные адаптивные регрессионные сплайны (MARSplines).
- 22 STATISTICA Data Miner. Быстрое внедрение прогнозирующих моделей.

Экзаменационные вопросы

1. Признаки интеллектуальности информационных систем. Отличие интеллектуальных задач от обычных информационных задач. Примеры интеллектуальных задач в экономике.
2. Основные виды интеллектуальных задач. Структура исследований в области искусственного интеллекта. Основные классы интеллектуальных информационных систем. Общие черты интеллектуальных технологий обработки информации.
3. Знания как особая форма информации. Отличие знаний от данных. Декларативные и процедурные знания. Модели знаний. Базы знаний (БЗ).
4. Технологии оперативной аналитической обработки (OLAP), области их применения.
5. Понятия «фактов» и «измерений» в технологиях оперативной аналитической обработки (OLAP). Многомерное представление данных в технологиях оперативной аналитической обработки.
6. Основные принципы организации инструментов оперативной аналитической обработки (OLAP-анализаторов).
7. Понятие интеллектуального анализа данных (Data Mining). Проблема «сырых данных».
8. Шаблоны, выявляемые методами интеллектуального анализа данных (Data Mining). Примеры из области экономических знаний.
9. Инструментальные средства интеллектуального анализа данных, их виды.
10. Экспертные системы. «Неявные знания», проблема их формализации и использования. Сферы применения экспертных систем.
11. Принципиальная структура экспертной системы, назначение ее элементов.
12. Инструментальные средства построения экспертных систем.
13. Генетические алгоритмы, их сущность, области применения.
14. Основные стадии генетического алгоритма, их сущность.
15. Искусственные нейронные сети. Основные направления применения нейросетевых технологий в экономике. Принципиальные отличия обработки информации методами искусственных

- нейронных сетей и алгоритмической обработки «формальными» методами.
16. Основные классы задач, решаемые методами искусственных нейронных сетей (ИНС).
 17. Принципиальная модель искусственного нейрона (нейрон МакКалока-Питтса).
 18. Активационная функция искусственного нейрона. Виды активационных функций. Активационные функции, применяемые в искусственных нейронных сетях типа MLP (многослойный персептрон) и самоорганизующихся картах Кохонена.
 19. Виды искусственных нейронных сетей (ИНС) по типам обучения. Процесс обучения ИНС. Понятие «эпохи» обучения.
 20. Искусственные нейронные сети типа MLP (многослойный персептрон). Принципиальная схема сети MLP. Связи в сетях MLP. Назначение скрытых слоев.
 21. Проблема переобучения искусственных нейронных сетей типа MLP (многослойный персептрон). Способы преодоления проблемы переобучения.
 22. Основные алгоритмы обучения искусственных нейронных сетей типа MLP (многослойный персептрон), их различия.
 23. Степень обучения и степень обобщения как характеристика качества обучения искусственных нейронных сетей типа MLP (многослойный персептрон). Показатели оценки данных характеристик.
 24. Типовые задачи, решаемые с помощью искусственных нейронных сетей типа MLP (многослойный персептрон). Задачи классификации и задачи аппроксимации.
 25. Визуальные методы и количественные показатели оценки качества обучения искусственных нейронных сетей типа MLP (многослойный персептрон).
 26. Самоорганизующиеся карты Кохонена, их назначение, возможности и направления применения в экономике.
 27. Принципиальная схема нейронной сети, лежащей в основе самоорганизующихся карт Кохонена.
 28. Процесс обучения самоорганизующихся карт Кохонена, его стадии.
 29. Визуальная интерпретация результатов обучения самоорганизующихся карт Кохонена.
 30. Назначение специальных отображений результатов обучения самоорганизующихся карт Кохонена: матрица расстояний, матрица плотности попадания, проекция Саммона.
 31. Карты Кохонена.
 32. Деревья решений.
 33. Ассоциативные правила. Алгоритм ограниченного перебора.

7.УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература

1. Популярное введение в современный анализ данных в системе STATISTICA [Электронный ресурс] : Учебное пособие для вузов / Боровиков В.П. - М. : Горячая линия - Телеком, 2013. - <http://www.studentlibrary.ru/book/ISBN9785991203265.html>
2. Основы построения интеллектуальных систем [Электронный ресурс] : учеб. пособ./ Г.В. Рыбина. - М. : Финансы и статистика, 2014. Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785279034123.html>
3. Модели и методы искусственного интеллекта. Применение в экономике [Электронный ресурс] /: учеб. пособие / М.Г. Матвеев, А.С. Свиридов, Н.А. Алейникова. - М. : Финансы и статистика, 2014. – Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785279032792.html>

Дополнительная литература

1. Основы нечеткой технологии и примеры решения аналитических задач в государстве и бизнесе [Электронный ресурс] / Свешников С.В., Бочарников В.П. - М. : ДМК Пресс, 2014. - <http://www.studentlibrary.ru/book/ISBN9785940749561.html>

2. MATLAB 6.5 SP1/7/7 SP1/7 SP2 + Simulink 5/6. Инструменты искусственного интеллекта и биоинформатики. [Электронный ресурс] / В. П. Дьяконов, В. В. Круглов - М. : СОЛОН-ПРЕСС, 2009. Режим доступа: <http://www.studentlibrary.ru/book/ISBN5980032551061.html>
3. Самоорганизующиеся карты [Электронный ресурс] / Т. Кохонен. - М. : БИНОМ, 2014. – Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785996313488.html>
4. Информационные технологии моделирования финансовых рынков [Электронный ресурс] / В.П. Романов, М.В. Бадрина. - М. : Финансы и статистика, 2010. - (Прикладные информационные технологии) Режим доступа: <http://www.studentlibrary.ru/book/ISBN9785279034444.html>

Интернет-ресурсы

<http://sdb.su/system-intellekt/>
<http://www.gotai.net/documents.aspx>
<http://www.aiportal.ru/>
<http://www.artint.ru/>
<http://www.raai.org/>
<http://techvesti.ru/robot>
<http://ru.wikipedia.org/wiki/>

Периодические издания

1. «Информационные технологии». Ежемесячный теоретический и прикладной научно-технический журнал(с приложением) ISSN 1684-6400
2. «Прикладная информатика» Научно-практический журнал
3. «[Информатика и системы управления](#)». Научный журнал. ISSN 1814-2400

8.МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Практические занятия проводятся в аудитории, обеспеченной мультимедийной аппаратурой, позволяющей использовать различные варианты демонстрации изучаемого материала.

Лабораторные занятия проводятся в компьютерном классе на 12 рабочих мест, что позволяет работать студентам в индивидуальном режиме.

Студенты имеют возможность доступа к локальной сети кафедры и сети университета.

Рабочая программа дисциплины составлена в соответствии с требованиями ФГОС ВО по направлению **«Прикладная информатика»**

Рабочую программу составил



Д.Н.Васильев

к.т.н., доцент

Рецензент
Зам.исполнительного директора
Владимирского городского ипотечного фонда
к.э.н.



А.П.Чернявский

Программа рассмотрена и одобрена на заседании кафедры УИТЭС

Протокол № 3/1 от 02.04.15 года

Заведующий кафедрой



А.Б.Градусов

Рабочая программа рассмотрена и одобрена на заседании учебно-методической комиссии направления **«Прикладная информатика»**

Протокол № 5 от 02.04.15 года

Председатель комиссии



А.Б.Градусов