

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение  
высшего образования

**«Российский государственный гуманитарный университет»**

**(ФГБОУ ВО «РГГУ»)**

ИСТОРИКО-АРХИВНЫЙ ИНСТИТУТ

ФАКУЛЬТЕТ АРХИВОВЕДЕНИЯ И ДОКУМЕНТОВЕДЕНИЯ

*Кафедра источниковедения*

## **БОЛЬШИЕ ДАННЫЕ И НЕЙРОННЫЕ СЕТИ**

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

**Направление подготовки 46.03.02 – Документоведение и архивоведение**

**Направленность – Электронные архивы и документы**

**Уровень высшего образования: бакалавриат**

**Форма обучения – очно-заочная**

РПД адаптирована для лиц  
с ограниченными возможностями  
здоровья и инвалидов

Москва 2024

БОЛЬШИЕ ДАННЫЕ И НЕЙРОННЫЕ СЕТИ

Рабочая программа дисциплины

Составитель:

*к.и.н., доцент С.В.Ашмарина*

.....

УТВЕРЖДЕНО

Протокол заседания кафедры

источниковедения

№   2   от  27.02.2024

## ОГЛАВЛЕНИЕ

### 1. Пояснительная записка

1.1 Цель и задачи дисциплины

1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

1.3. Место дисциплины в структуре образовательной программы

### 2. Структура дисциплины

### 3. Содержание дисциплины

### 4. Образовательные технологии

### 5. Оценка планируемых результатов обучения

5.1. Система оценивания

5.2. Критерии выставления оценок

5.3. Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

### 6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список источников и литературы

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

6.3. Профессиональные базы данных и информационно-справочные системы

### 7. Материально-техническое обеспечение дисциплины

### 8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

### 9. Методические материалы

9.1. Планы практических (семинарских, лабораторных) занятий

9.2. Методические рекомендации по подготовке письменных работ

9.3. Иные материалы

## Приложения

Приложение 1. Аннотация дисциплины

## 1 \_Пояснительная записка

### 1.1 Цель и задачи дисциплины.

Дисциплина «Большие данные и нейронные сети» входит в состав части, формируемой участниками образовательных отношений, блока дисциплин учебного плана бакалаврской программы «Электронные архивы и документы» по направлению подготовки 46.03.02 – Документоведение и архивоведение.

Цель дисциплины – приобретение знаний в области систем больших данных и нейронных сетей и принятия решений (ПР); изучение программных средств конструирования интеллектуальных систем (ИС) для различных предметных областей.

Задачи дисциплины:

- ознакомление с концепциями и методами, составляющими основу для понимания современных достижений искусственного интеллекта;
- ознакомление с современными областями исследования по искусственному интеллекту;
- ознакомление с основными моделями представления знаний и некоторыми интеллектуальными системами;
- рассмотрение теоретических и некоторых практических вопросов создания и эксплуатации экспертных систем;
- ознакомление с особенностями практического использования интеллектуальных

информационных систем и систем принятия решений.

1.2 Формируемые компетенции, соотнесённые с планируемыми результатами обучения по дисциплине:

Компетенция (код и наименование)	Индикаторы компетенций (код и наименование)	Результаты обучения
ПК-4 - Способность создавать и вести системы документационного обеспечения управления архивов на базе новейших технологий	ПК-4.2 - Способен анализировать ситуацию на рынке информационных продуктов и услуг, давать экспертную оценку современным системам электронного документооборота и ведения электронного архива	<p><b>Знать:</b> определение интеллектуальных систем, структуру статических и динамических экспертных систем; теоретические основы построения и функционирования прикладных интеллектуальных систем и систем поддержки принятия решений, ключевые направления применения новых информационных систем при автоматизации процессов принятия управленческих решений; основные источники научно-технической информации по основным направлениям, методам, моделям и инструментальным средствам конструирования интеллектуальных систем.</p> <p><b>Уметь:</b> разрабатывать постановку задач для решения неформализованных проблем; формулировать цели и задачи автоматизации обработки управленческой информации; применять интеллектуальные системы для решения задач оценки и прогнозирования состояния объектов.</p> <p><b>Владеть:</b> терминологией, навыками поиска и использования научно-технической информации по профессиональной тематике;</p>

		современными методами применения прикладных интеллектуальных систем и систем поддержки принятия решений.
--	--	--

### 1.3 Место дисциплины в структуре образовательной программы

Дисциплина «Большие данные и нейронные сети» относится к части, формируемой участниками образовательных отношений, блока дисциплин учебного плана бакалаврской программы «Электронные архивы и документы» по направлению подготовки 46.03.02 – Документоведение и архивоведение. Дисциплина реализуется кафедрой источниковедения факультета архивного дела историко-архивного института РГГУ. Дисциплина «Большие данные и нейронные сети» имеет своей целью приобретение знаний в области систем искусственного интеллекта (ИИ) и принятия решений (ПР); изучение программных средств конструирования интеллектуальных систем (ИС) для различных предметных областей.

В процессе изучения дисциплины студенты приобретают навыки владения технологиями поиска и использования научно-технической информации по профессиональной тематике; современными методами применения прикладных интеллектуальных систем и систем поддержки принятия решений, построением моделей представления знаний, подходами и техникой решения задач искусственного интеллекта, информационных моделей знаний, методами представления знаний.

Для освоения дисциплины «Большие данные и нейронные сети» необходимо изучение следующих дисциплин: Интернет-коммуникации, Информационная безопасность в архивной сфере, Информационно-аналитические системы государственных и муниципальных органов власти. В результате освоения дисциплины формируются компетенции, необходимые для прохождения практик

## 2 Структура дисциплины

### *Структура дисциплины для очно-заочной формы обучения*

Общая трудоемкость дисциплины для очно-заочной формы составляет 3 з. е., 108 ч., в том числе контактная работа обучающихся с преподавателем 36 ч.

Семестр	Тип учебных занятий	Количество часов
8	Лекции	16
8	Семинары <sup>1</sup>	20
Всего:		36

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 72\_\_ академических часа(ов).

## 3 Содержание дисциплины

№	Наименование раздела дисциплины	Содержание
1.	Тема 1. Введение большие данные и нейронные сети	Актуальность дисциплины. Место дисциплины среди других наук. Основные понятия. Понятие интеллекта. Область ИИ. Подходы к определению ИИ. Информационный, бионический и эволюционный подходы. Интеллектуальные системы. Цели, задачи и возможность создания ИИ. История искусственного интеллекта. Этапы развития и основные направления ИИ. Возражения против ИИ. Классификация ИС. Обобщённая функциональная структура ИС. Основные (базовые) свойства и возможности. Обобщённая типология знаний.
2.	Тема 2. Формализация знаний в интеллектуальных системах (ИС).	Основные понятия и определения. Предметная область. Данные и знания. Свойства, характеристики знаний. Процедурные и декларативные знания. Классификация знаний по глубине, по жесткости. Формализация знаний. Формальные языки. Языки (модели) представления знаний. Классификация моделей знаний и данных.

<sup>1</sup> В соответствии с учебным планом

3.	Тема 3. Моделирование процессов обработки информации для принятия решений.	Организационные структуры и функции отдела ИТ. Системы управления и мониторинга ИТ-инфраструктуры предприятия. Безопасность корпоративных систем. Хранилища данных. Аналитические системы. Архитектура OLAP.
4.	Тема 4. Генетический алгоритм	Теория эволюции Дарвина и ее применение в ИС. Понятие о генетическом алгоритме. Этапы работы генетического алгоритма. Кодирование информации и формирование популяции. Оценивание популяции. Селекция. Скрещивание и формирование нового поколения. Мутация. Настройка параметров генетического алгоритма. Канонический генетический алгоритм. Пример работы генетического алгоритма. Рекомендации к программной реализации генетического алгоритма. Применение генетического алгоритма для решения задач оптимизации и аппроксимации.
5.	Тема 5. Нечеткая логика.	Многозначные логики. Нечеткая логика. Нечеткое множество. Степень вхождения (уровень принадлежности). Основные операции в нечеткой логике. Нечеткие правила вывода в экспертных системах. Фазификация, дефазификация, нечеткий вывод. Сравнение выводов Mamdani и TVFI. Методы дефазификации. Отличие нечеткости и

		вероятности.
6.	Тема 6. Нейронные сети.	Нейронные сети и их применение в ИС. Биологический прототип и искусственный нейрон. Математические модели нейронов. Однослойные искусственные нейронные сети. Многослойные искусственные нейронные сети. Терминология, обозначения и схематическое изображение искусственных нейронных сетей. Персептроны и зарождение искусственных нейронных сетей. Персептронная представляемость. Обучение персептрона. Алгоритм обучения персептрона. Процедура обратного распространения. Обучающий алгоритм обратного распространения. Пример обучения. Область применения алгоритма и ограничения по использованию. Мультиагентные системы.
7.	Тема 7. Экспертные системы.	Понятие экспертной системы. Структура ЭС. Классификации ЭС. Подходы к созданию ЭС. Преимущества и недостатки ЭС. Особенности неформализованных задач. Интегрированность, открытость и переносимость ЭС. Проблемно / предметно – ориентированные ИС. Типология ЭС. Принципиальная технология создания и этапы проектирования ЭС.



#### **4 Образовательные технологии.**

Для проведения учебных занятий по дисциплине используются различные образовательные технологии. Для организации учебного процесса может быть использовано электронное обучение и (или) дистанционные образовательные технологии.

## 5 Оценка планируемых результатов обучения

### 5.1 Система оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль: - опрос - выполнение лабораторных работ - контрольная работа в форме компьютерного тестирования	10 баллов 10 баллов 10 баллов	10 баллов 40 баллов 10 баллов
Промежуточная аттестация (зачет)		40 баллов
<b>Итого за семестр (дисциплину) зачет</b>		100 баллов

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

## 5.2 Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ А,В	«отлично»/ «зачтено (отлично)»/ «зачтено»	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ С	«хорошо»/ «зачтено (хорошо)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,Е	«удовлетвори- тельно»/ «зачтено (удовлетвори- тельно)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».
49-0/ F,FX	«неудовлетворительно»/ не зачтено	Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации. Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами. Демонстрирует фрагментарные знания учебной литературы по дисциплине. Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.

### 5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине.

#### **Примерные вопросы на компьютерное тестирование:**

*ДПК – 2 - Готов к работе с интеллектуальными информационными системами, системами с параллельной обработкой данных в гуманитарной сфере.*

1. Запрос на добавление
  - a. заменяет исходные данные новыми в таблицах базы данных
  - b. выводит указанные данные таблиц базы данных на экран
  - c. копирует указанные данные из одной таблицы в другую существующую таблицу базы данных
  - d. удаляет указанные данные из одной таблицы базы данных и создает на их основе новую таблицу
2. Для решения неформализуемых задач обычно используются
  - a. факты и правила
  - b. правила
  - c. фреймы
  - d. нейронные сети
3. Директива GROUP BY означает
  - a. группировку полей
  - b. группировку баз данных
  - c. группировку записей
  - d. группировку таблиц

4. Семантическая сеть состоит
  - a. из узлов и дуг
  - b. из дуг и отношений
  - c. из объектов и узлов
  - d. из фактов и правил
5. Экспертная система в определенной предметной области может заменить
  - a. конечного пользователя
  - b. экспертов в этой области
  - c. компьютер
  - d. экспертов в области математики
6. Фрейм – это структура, состоящая
  - a. из характеристик и их значений
  - b. из фактов и правил
  - c. из семантических сетей
  - d. из баз данных
7. Директива SELECT означает
  - a. обновить данные
  - b. удалить данные
  - c. выбрать данные
  - d. создать данные
8. Фреймы дают возможность хранить в базе знаний
  - a. модель мышления
  - b. базу данных
  - c. иерархию понятий
  - d. правила вывода
9. Связь 1:1 (один к одному) имеет место между объектами
  - a. группа и студент
  - b. группа и староста группы
  - c. студент и предмет
  - d. товар и магазин
10. Директива FROM определяет
  - a. поля, из которых выбираются данные
  - b. записи, из которых выбираются данные
  - c. базы данных, из которых выбираются данные
  - d. таблицы, из которых выбираются данные

**Вопросы промежуточного контроля (Экзамен):**

*ПК – 8 - Способен программировать приложения и создавать программные прототипы решения прикладных задач.*

1. Уровни понимания. Методы решения задач.
2. Решение задач методом поиска в пространстве состояний.
3. Фреймы. Исчисления предикатов.
4. Системы продукций. Семантические сети.
5. Нечеткая логика.
6. Алгоритмы эвристического поиска.
7. Поиск решений на основе исчисления предикатов.
8. Переход от Базы данных к Базе знаний. Особенности знаний.
9. Продукционные системы. Классификация ядер продукции.
10. Стратегия решений организации поиска.
11. Нечеткое планирование.
12. Сложность решения задач планирования.
13. Назначение экспертных систем.
14. Структура экспертных систем.

15. Этапы разработки экспертных систем.
16. Представление знаний в экспертных системах.
17. Режимы взаимодействия инженера по знаниям с экспертом.
18. Методы работа со знаниями.
19. Основная модель нейросетевой технологии.
20. Методы извлечения знаний

## 6 Учебно-методическое и информационное обеспечение дисциплины

### 6.1 Список источников и литературы

#### Литература

##### Основная

1. Трофимов, В. Б. Интеллектуальные автоматизированные системы управления технологическими объектами: Учебно-практическое пособие / Трофимов В.Б., Кулаков С.М. - Вологда:Инфра-Инженерия, 2016. - 232 с.: ISBN 978-5-9729-0135-7. - Текст : электронный. - URL: <https://new.znanium.com/catalog/product/760121>.
2. Вейнберг, Р. Р. Интеллектуальный анализ данных и систем управления бизнес-правилами в телекоммуникациях: Монография / Р.Р. Вейнберг. - Москва : НИЦ ИНФРА-М, 2016. - 173 с. (Научная мысль) ISBN 978-5-16-011350-0. - Текст : электронный. - URL: <https://new.znanium.com/catalog/product/520998>.
3. Гладков, Л. А. Генетические алгоритмы [Электронный ресурс] / Под ред. В. М. Курейчика. - 2-е изд., исправл. и доп. - Москва : ФИЗМАТЛИТ, 2010. - 368 с. - ISBN 978-5-9221-0510-1. - Текст : электронный. - URL: <https://new.znanium.com/catalog/product/544626>.

##### Дополнительная

1. Программные продукты и системы : международный научно-практический журнал. – Тверь : НИИ Центрпрограммсистем, 2016. - № 2 (114). – 212 с. – ISSN 0236-235X. – Текст : электронный. – URL: <https://new.znanium.com/catalog/product/1016265>.
2. Сети связи и системы коммутации: Учебное пособие / Паринов А.В., Ролдугин С.В., Мельник В.А. - Воронеж:Научная книга, 2016. - 178 с. ISBN 978-5-4446-0906-4. - Текст : электронный. - URL: <https://new.znanium.com/catalog/product/923309>.
3. Доррер, Г. А. Методы и системы принятия решений: Учебное пособие / Доррер Г.А. - Краснояр.:СФУ, 2016. - 210 с.: ISBN 978-5-7638-3489-5. - Текст : электронный. - URL: <https://new.znanium.com/catalog/product/978605>.

### 6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».

1. <http://znanium.com> – Электронно-библиотечная система.
2. <http://window.edu.ru> - Информационная система «Единое окно доступа к образовательным ресурсам».

### 6.3 Профессиональные базы данных и информационно-справочные системы

№п/п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2017 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной

	подписки в 2017 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis Электронные издания издательства Springer
3	Компьютерные справочные правовые системы Консультант Плюс, Гарант

### 7 Материально-техническое обеспечение дисциплины.

Для материально-технического обеспечения дисциплины «Интеллектуальные информационные системы в гуманитарной сфере» необходимо:

1. Компьютерный класс – ауд. № 203.
2. 1 компьютер преподавателя, 12 компьютеров обучающихся, маркерная доска, проектор.
3. Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа

Windows 7	68526624	без даты
Microsoft office 2010 Pro	49420326	08.12.2011
Microsoft Visual Studio 2005	77626-009-0000007-41832	без даты
Mozilla Firefox	свободный доступ 647526	свободный доступ без даты
Matlab	2696062	26.10.2012
Mathcad Education - University edition	1894141205092525	05.12.2014
Kaspersky Endpoint Security		

### 8 Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих:
  - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
  - письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;
  - обеспечивается индивидуальное равномерное освещение не менее 300 люкс;
  - для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;
  - письменные задания оформляются увеличенным шрифтом;
  - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.
- для глухих и слабослышащих:
  - лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования;

- письменные задания выполняются на компьютере в письменной форме;
- экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

- для лиц с нарушениями опорно-двигательного аппарата:
  - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
  - письменные задания выполняются на компьютере со специализированным программным обеспечением;
  - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих:
  - в печатной форме увеличенным шрифтом;
  - в форме электронного документа;
  - в форме аудиофайла.
- для глухих и слабослышащих:
  - в печатной форме;
  - в форме электронного документа.
- для обучающихся с нарушениями опорно-двигательного аппарата:
  - в печатной форме;
  - в форме электронного документа;
  - в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих:
  - устройством для сканирования и чтения с камерой SARA CE;
  - дисплеем Брайля PAC Mate 20;
  - принтером Брайля EmBraille ViewPlus;
- для глухих и слабослышащих:
  - автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих;
  - акустический усилитель и колонки;
- для обучающихся с нарушениями опорно-двигательного аппарата:
  - передвижными, регулируемые эргономическими партами СИ-1;
  - компьютерной техникой со специальным программным обеспечением.



## 9 Методические материалы

Образовательные технологии в преподавании дисциплины направлены на реализацию компетентностного подхода и основываются на принципе профессиональной направленности обучения. Традиционная лекционно-семинарская деятельность дополняется активными и интерактивными формами проведения занятий (презентациями, деловыми играми, разборами конкретных ситуаций, психологическими и иными тренингами и др.). В ходе реализации программы аудиторные занятия проводятся в виде лекций и семинарских занятий с использованием ПК и компьютерного проектора; самостоятельная работа обучающихся проводится под руководством преподавателей (консультации и помощь в написании контрольных работ), а также в библиотеке, в том числе с информационными ресурсами Интернета.

№ п/п	Наименование раздела	Виды учебных занятий	Образовательные технологии
1	2	3	4
1.	Тема 1. Введение в большие данные и нейронные сети	Лекция 1.	Лекция с использованием видеоматериалов.
2.	Тема 2. Формализация знаний в интеллектуальных системах (ИС).	Самостоятельная работа.	Консультирование и проверка домашних заданий посредством электронной почты.
3.	Тема 3. Моделирование процессов обработки информации для принятия решений.	Лекция 3 Лабораторное занятие	Лекция с использованием видеоматериалов. Выполнение лабораторной работы посредством специализированного ПО.
4.	Тема 4. Генетический алгоритм	Лекция 4  Лабораторное занятие	Лекция с использованием видеоматериалов.  Компьютерное тестирование.
5.	Тема 5. Нечеткая логика.	Самостоятельная работа.	Пробное компьютерное тестирование.
6.	Тема 6. Нейронные сети.	Лабораторное занятие	Выполнение лабораторной работы посредством специализированного ПО.
7.	Тема 7. Экспертные системы.	Лекция 7.	Лекция - дискуссия.

### 9.1 Планы лабораторных занятий.

#### Лабораторная работа 1. Тема 3. (8 ч.) Характеристики процесса обработки информации.

Задание.

1. Проработать метод анализа процесса обработки информации, представленной граф - схемой алгоритма (ГСА).

Указания по выполнению заданий:

- 1) Ознакомиться с анализом процесса обработки информации по курсу лекций.
- 2) Получить у преподавателя варианты заданий.
- 3) Определить вероятности перехода  $p_{ij}$  и характеристический вектор  $N$  для каждого оператора.

- 4) Рассчитать среднее число  $n_1, n_2, n_3, \dots, n_{10}$ , пребывания марковского процесса в состояниях  $s_1, s_2, s_3, \dots, s_{10}$ .
- 5) Рассчитать среднее число операций каждого типа в алгоритме управления: короткие, длинные, передачи.
- 6) Определить динамическую длину программы имитационного моделирования алгоритмов управления.
- 7) Определить частотный вектор для анализируемого алгоритма
- 8) Рассчитать внешнюю и внутреннюю связность одного из операторов по указанию преподавателя.
- 9) Оформить отчёт по выполненной работе

**Содержание отчёта по лабораторной работе:**

- 1) Название, цель работы, номер варианта задания.
- 2) Граф - схема алгоритма управления.
- 3) Таблица с вероятностями перехода  $p_{ij}$ .
- 4) Расчеты среднего числа  $n_1, n_2, n_3, \dots, n_{10}$ , пребывания марковского процесса в состояниях  $s_1, s_2, s_3, \dots, s_{10}$ .
- 5) Характеристический вектор для каждого оператора.
- 6) Расчеты среднего число операций каждого типа в алгоритме управления: короткие, длинные, операции обращения.
- 7) Расчеты динамической длины программы.
- 8) Частотный вектор для анализируемого алгоритма.
- 9) Расчеты внешней и внутренней связности анализируемого оператора
- 10) Выводы по результатам исследований.

**Материально-техническое обеспечение занятия:**

Для материально-технического обеспечения дисциплины «Интеллектуальные информационные системы» необходимо:

1. Компьютерный класс – ауд. № 203.

2. 1 компьютер преподавателя, 12 компьютеров обучающихся, маркерная доска, проектор.
3. Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа

Windows 7	68526624	без даты
Microsoft office 2010 Pro	49420326	08.12.2011
Microsoft Visual Studio 2005	77626-009-0000007-41832	без даты
Mozilla Firefox	свободный доступ 647526	свободный доступ без даты
Matlab	2696062	26.10.2012
Mathcad Education - University edition	1894141205092525	05.12.2014
Kaspersky Endpoint Security		

### Лабораторная работа 2. Тема 4. (6 ч.) Генетический алгоритм.

Задания:

1. Исследовать экстремумы функций с помощью генетических алгоритмов.

Указания по выполнению заданий:

- 1) Найти минимум функции одной переменной любым известным способом. Исследовать функцию с помощью генетических алгоритмов. Сравнить полученные результаты. Определить глобальный минимум и значение функции в этой точке. Провести эксперимент при различном размере начальной популяции: 10; 50; 300; 800. Для каждого из этих значений принять следующие операторы отбора родительских особей: • Stochastic uniform; • Uniform; • Roulette. Сделать вывод о том, как влияет размер исходной популяции на результаты, а также какое влияние вносят различные операторы отбора родительских особей. Привести графическое решение, найденное с помощью генетических алгоритмов, которое соответствует максимально точному найденному решению. Объяснить каким образом влияют операторы отбора родительских особей на результаты. Целевые функции одной переменной приведены в таблице у преподавателя.
- 2) Найти минимум функции двух переменных любым известным способом. Провести исследование этой функции с помощью генетических алгоритмов и заполнить табл. 3, изменяя вероятность скрещивания и размер начальной популяции. Выбрать оператор мутации по Гауссу, вероятность мутации принять равным по умолчанию 1,0. По результатам проведенных экспериментов сделать вывод о том, при каком соотношении вероятности скрещивания к мутации результаты максимально точны. Привести графическое решение, найденное с помощью генетических алгоритмов, которое соответствует максимально точному найденному решению. Определить относительную погрешность полученных значений функций при каждом значении размера начальной популяции (для наилучшего соотношения вероятности скрещивания к мутации) и построить график зависимости погрешностей от размера начальной популяции. Сделать вывод о влиянии размера начальной популяции на результаты. Целевые функции двух переменных приведены в табл. 4.
- 3) Оформить отчет, в который должно входить:
  - цель работы;
  - задание;

- краткое описание действий по пунктам;
- графики;
- вывод по каждой части задания.

Материально-техническое обеспечение занятия:

Для материально-технического обеспечения дисциплины «Интеллектуальные информационные системы» необходимо:

1. Компьютерный класс – ауд. № 203.
2. 1 компьютер преподавателя, 12 компьютеров обучающихся, маркерная доска, проектор.
3. Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа

Windows 7	68526624	без даты
Microsoft office 2010 Pro	49420326	08.12.2011
Microsoft Visual Studio 2005	77626-009-0000007-41832	без даты
Mozilla Firefox	свободный доступ 647526	свободный доступ без даты
Matlab	2696062	26.10.2012
Mathcad Education - University edition	1894141205092525	05.12.2014
Kaspersky Endpoint Security		

### Лабораторная работа 3. Тема 5. (8 ч.) Нечеткая логика.

Задания:

1. Изучить основные понятия нечеткой логики и способы задания и этапы функционирования нечеткой экспертной системы.
2. Разработать компьютерную модель нечеткой экспертной системы и исследовать ее работу.

Указания по выполнению заданий:

1. В прикладном пакете Fuzzy logic toolbox программы MATLAB создадим новый проект и зададим нечеткие функции принадлежности для температуры как входные параметры.
2. Зададим нечеткие функции принадлежности для температуры как входные параметры. Задать правила вывода. Правила вывода в созданной компьютерной модели нечеткие правила вывода задаются при помощи вкладки Edit→Rules.

3. Получить отклик системы. Последовательность обработки нечетких для конкретного значения температуры можно просмотреть в окне просмотра правил View→Rules. Передаточную характеристику системы, т.е. зависимость скорости вращения кондиционера от температуры для рассматриваемой модели можно просмотреть при помощи команды View→Surface.
4. В результате выполненной работы заполняется бланк отчета, в котором должны быть отражены:
  - Программа и цель работы.
  - Формальное представление нечеткой экспертной системы.
  - Задание нечетких функций принадлежности для входных и выходных параметров моделируемой системы и правил нечеткого вывода.
  - Выводы по исследованию функционирования нечеткой экспертной системы.

Материально-техническое обеспечение занятия:

Для материально-технического обеспечения дисциплины «Интеллектуальные информационные системы» необходимо:

1. Компьютерный класс – ауд. № 203.
2. 1 компьютер преподавателя, 12 компьютеров обучающихся, маркерная доска, проектор.
3. Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа

Windows 7	68526624	без даты
Microsoft office 2010 Pro	49420326	08.12.2011
Microsoft Visual Studio 2005	77626-009-0000007-41832	без даты
Mozilla Firefox	свободный доступ	свободный доступ
Matlab	647526	без даты
Mathcad Education - University edition	2696062	26.10.2012
Kaspersky Endpoint Security	1894141205092525	05.12.2014

#### Лабораторная работа 4. Тема 6. (8 ч.) Нейронные сети.

Задание:

1. Согласно варианту задания на лабораторную работу по методу «Группового учета аргументов» выберем параметры, характеризующие состояние погоды и воды в качестве выходного (прогнозируемого) для каждой из сетей, которые предполагается построить.

Указания по выполнению заданий:

### Задание 1

Провести анализ значимых входных параметров для каждой нейронной сети. Для каждой нейронной сети представить данные об ошибках распознавания, сделать выводы о достоинствах или недостатках той или иной модели нейронной сети. По представленному графу сети описать наблюдаемую связь между нейронами различных слоев сети. Описать структуру и характеристики нейронных сетей, которые будем использовать для дальнейшей работы в качестве задания на лабораторную работу.

#### Структуры нейронных сетей:

- Обучающее множество = 90%
- Тестовое множество = 10%
- Способ разделения исходного множества данных: случайно
- Слоистость сети: 1 слой, 2 слоя
- Тип функции: сигмоида, гипертангенс
- Алгоритмы обучения: обратное распространение, RPROP
- Число нейронов в слоях: если 1 слой, то полусумма числа входных и выходных нейронов; если 2 слоя, то на первом число нейронов равно числу входных нейронов, а на втором - полусумме числа входных и выходных нейронов для сети.

### Задание 2

С помощью нейронной сети провести прогнозирование выбранных параметров состояния погоды и воды на следующий месяц.

Ход работы.

В качестве примера выберем параметр состояния погоды TEMP\_C. Построим для него одну нейронную сеть. Спрогнозируем его динамику изменения на следующий месяц.

1. Загрузить данные из Excel в Deductor

Рис. 1. Из таблицы Weather с помощью `id_city=2` а извлекаем все данные для `ода с`

В Deductor-е выберем вклад у «Сценарии». Нажмем ПКМ на строку «Сценарии» и выберем «Мастер импорта»

Шаг 1 – выбираем «MS Excel»;

Шаг 2 – выбираем файл с данными и, написав запрос к базе, извлекаем из нее данные о погоде для города с id\_city=2 (рис. 1).

Шаг 3 – жмем «Пуск»

Шаг 4 – для всех параметров погоды выбираем «Вид данных» - «Дискретный».

Шаг 5 – выбираем способ отображения данных на выбор (например «Таблица и «Диаграмма»).

В результате у вас должна добавиться новая запись «MS Excel...» в дереве «Сценарии».

1. Парциальная обработка данных

Проведем парциальную обработку данных для восстановления пропущенных данных, редактирования аномальных значений и удаления шумов пропущенных данных.

ПКМ на «MS Excel...» «Мастер обработки» «Парциальная обработка».

Для восстановления пропущенных данных выберем аппроксимацию, при редактировании аномальных значений используем малую степень подавления.

---

Рис. 2. Способы восстановления пропущенных значений

Согласно варианту лабораторной работы выберем параметр, значение которого

### 1. Выбор значимых параметров

Выбор значащих параметров осуществляется с помощью корреляционного анализа. («Мастер обработки» «Корреляционный анализ») Выберем выходное поле TEMP\_C, не использующиеся поля id\_city и YEARMODA, а все остальные отметим как входные.

В результате выполнения корреляционного анализа получаем значения коэффициентов корреляции для каждой получившейся пары выходного и входных параметров. Установив порог значимости, выберем значащие параметры. Из рис. 5 следует, что в нашем примере все параметры значимые.

---

### Рис. 3. Способы редактирования аномальных значений

После парциальной предобработки данных выводим на диаграмме все параметры погоды. На рис. 4 отметим, что данные по параметру SNDP содержат слишком много пропусков, и даже парциальная предобработка не смогла убрать такое количество пропусков. Значений все остальных параметров, судя по графику, подчиняются гармоническим законам, следовательно, любую их комбинацию можно использовать для прогнозирования значений другого параметра.

Рис. 4. Диаграмма значений параметров после парциальной обработки



## 1. Построение нейронных сетей

«Мастер обработки» «Нейросеть».

Выбрав входные и выходные параметры сети можно сделать нормализацию хранящихся в них значений (рис. 6) (для данного варианта задания параметр TEMP F не берем в качестве входа нейронной сети, так как зависимость между температурой по шкале Цельсия и по шкале Фаренгейта линейная, а мы все-таки хотим проследить нелинейную зависимость между входными и выходными переменными).

Подробнее вопросы нормализации значений параметров при построении нейронных сетей будут рассмотрены в лекционном материале.

Рис. 6. Настройка формализации входных и выходных данных для нейронной сети  
Далее, разобьем исходное множество на обучающее и тестовое подмножества в отношении 9/1 (рис. 7)

Рис. 7. разбиение входного набора данных на обучающее и тестовое подмножества  
Следующим шагом будет выбор структуры нейронной сети и настройка процесса ее обучения (рис. 8-9).

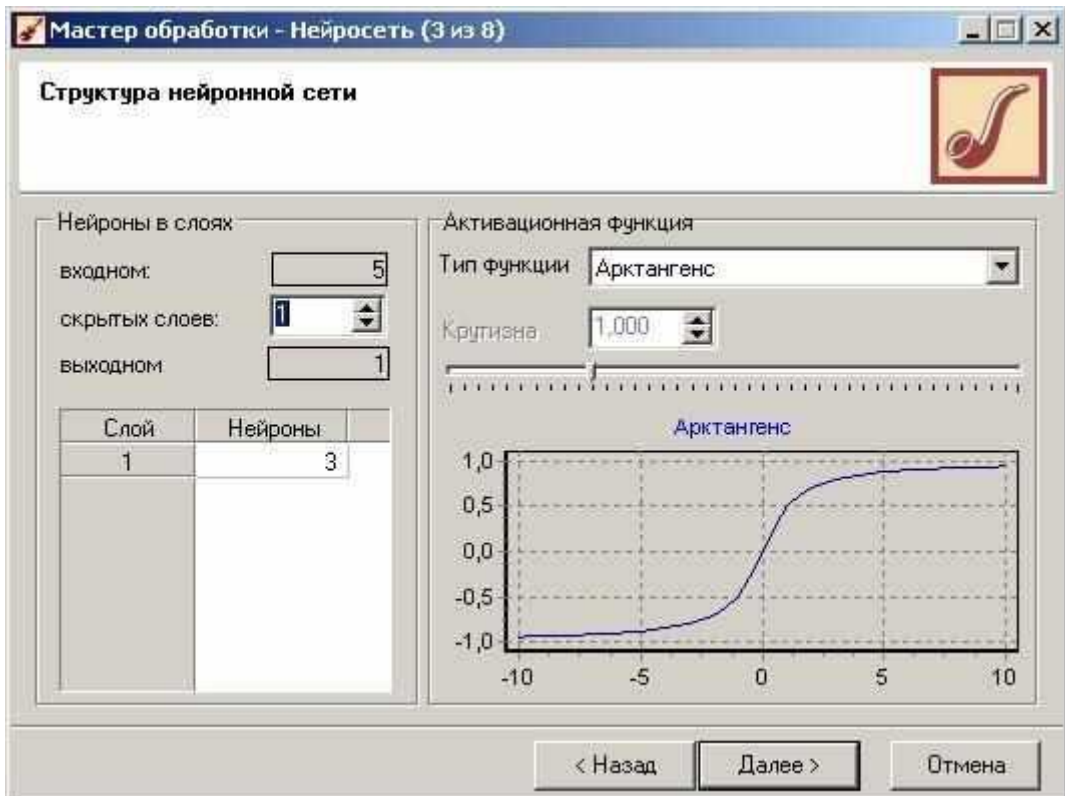


Рис. 8. Настройка структуры нейронной сети

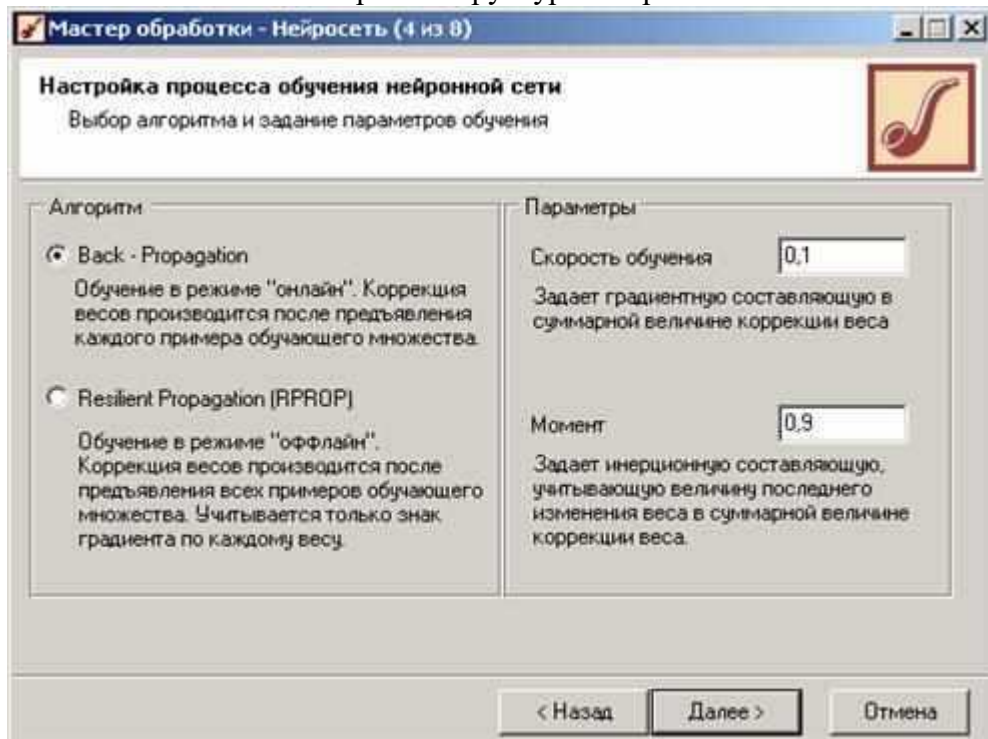


Рис. 9 Настройка процесса обучения нейронной сети

В качестве характеристик работы нейронной сети выберем следующие (рис. 10):

- Максимальная и средняя ошибки на обучающем множестве;
- Максимальная и средняя ошибки на тестовом множестве;
- Процент правильно распознанных случаев на обучающем множестве;
- Процент правильно распознанных случаев на тестовом множестве;
- Время обучения.

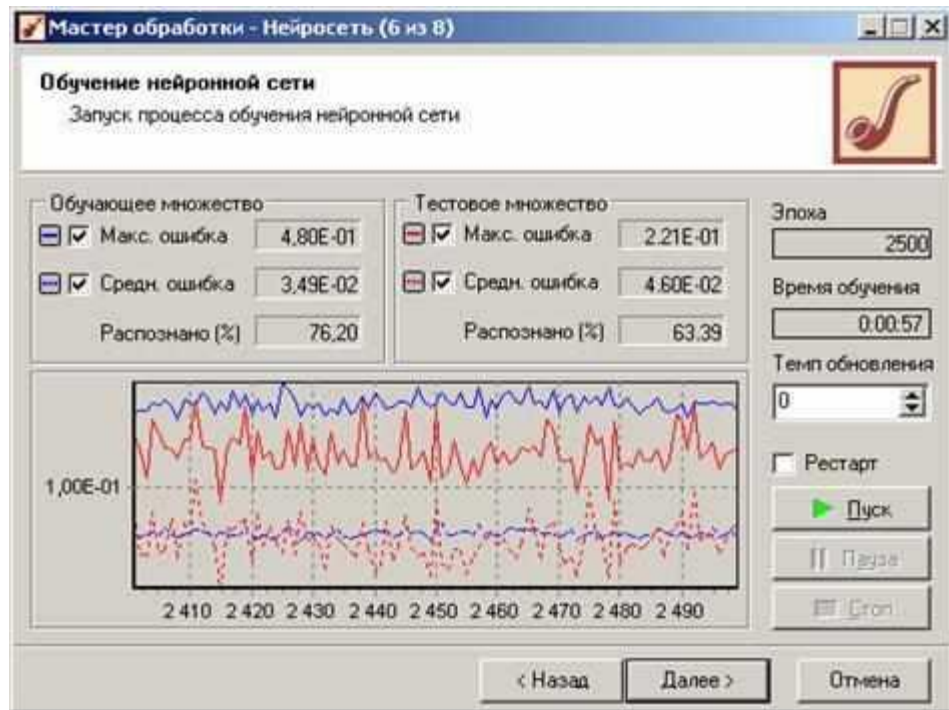


Рис. 10. Результаты процесса обучения нейронной сети

С помощью графа нейронной сети можно проследить распределение весов по связям между нейронами (рис. 11).

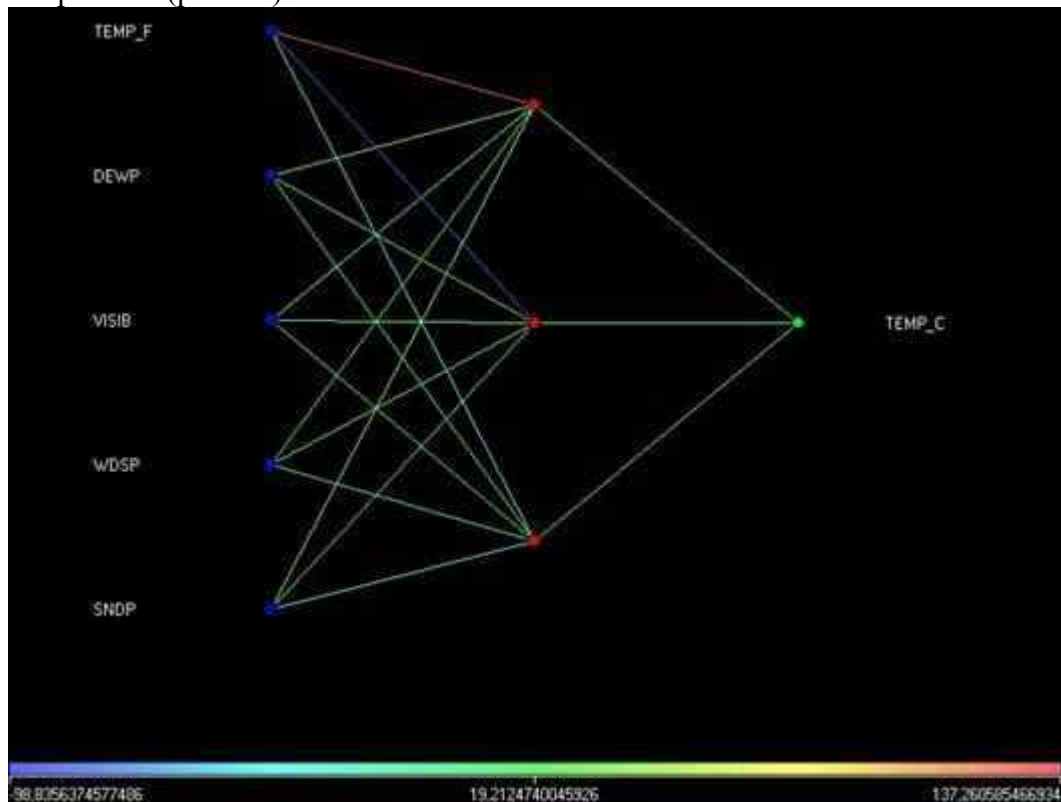


Рис. 11. Граф полученной нейронной сети

С помощью визуализатора «Что-если», меняя значения различных входных параметров можем с помощью нейронной сети спрогнозировать значение выходного параметра (рис. 12).

Поле	Значение
<b>Входные</b>	
9.0 TEMP_C-12	-10
9.0 TEMP_C-11	-8
9.0 TEMP_C-2	-2
9.0 TEMP_C-1	-5
<b>Выходные</b>	
9.0 TEMP_C	-15,9181 324916933

Рис. 12. Использование визуализатора «Что-если»

Теперь вы составили представление о достоинствах той или иной структуры нейронной сети. Далее рассмотрим вопрос, связанный с прогнозированием данных на длительный период.

Выберем проведенную «Парциальную обработку», и щелкнув ПКМ на строку во вкладке «Сценарии» выберем «Мастер обработки» → «Преобразование даты» так, чтобы в таблицу с исходными данными добавился столбец, содержащий данные о годе и месяце (рис. 13).

**Мастер обработки - Дата и время (1 из 3)**

**Преобразование даты и времени**  
Преобразование даты и времени

Имя столбца: YEARMODA  
Тип данных: Дата/Время  
Назначение:  Используемое

Разбиение	Дата	Число	Строка
Год + Квартал	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Год + Месяц	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Год + Неделя	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Год + День	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Год	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Квартал	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Месяц	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Неделя	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Обработать даты по ISO 8601

< Назад    Далее >    Отмена

Рис. 13. Преобразование даты и времени

Далее, сгруппируем данные так, чтобы для каждого месяца получить сведения о среднем значении выбранного параметра. Используем «Мастер обработки» → «Группировка» и в качестве измерения выберем только что добавленный столбец «Год-месяц», а в качестве факта – параметр, выбранный для прогнозирования. Проводимая агрегация для данного параметра – «Среднее» (рис. 14).

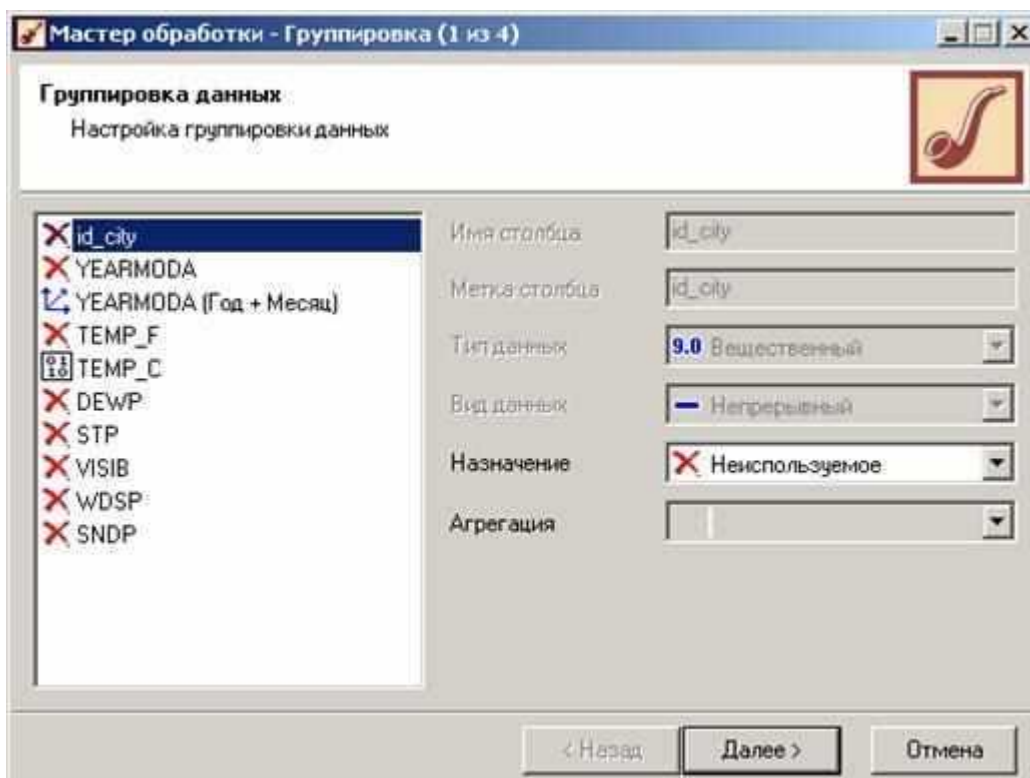


Рис. 14. Настройка группировки данных

Затем отфильтруем полученные значения: «Мастер обработки» → «Фильтрация». В качестве поля для фильтрации используем поле, хранящее данные о дате, а условие фильтрации составим так, чтобы выбрать все данные, за исключением последнего года наблюдения, данные за который будут использоваться в качестве проверочных (например, если наблюдения за года включительно, то условие соответствует рисунку 13).

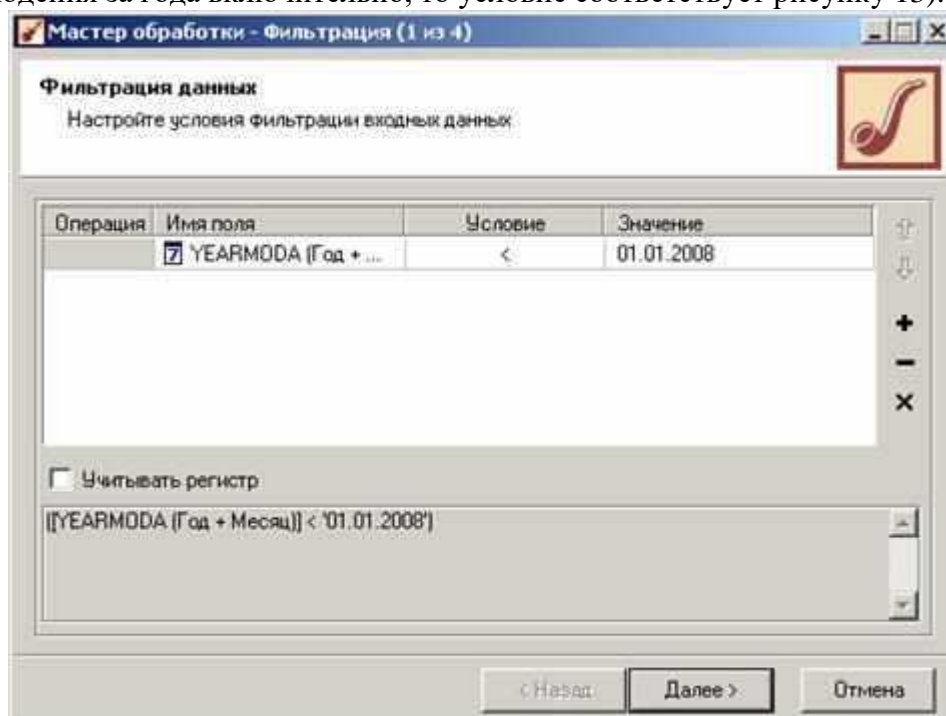


Рис. 15. Настройка условия фильтрации входных данных

Запустим «Мастер обработки», выберем в качестве обработчика «Скользящее окно» и перейдем на следующий шаг. Исходя из того факта, что данные представляют собой временной ряд с периодом сезонности в 12 месяцев, то в скользящем окне укажем глубину погружения в 12 месяцев (рис. 16).

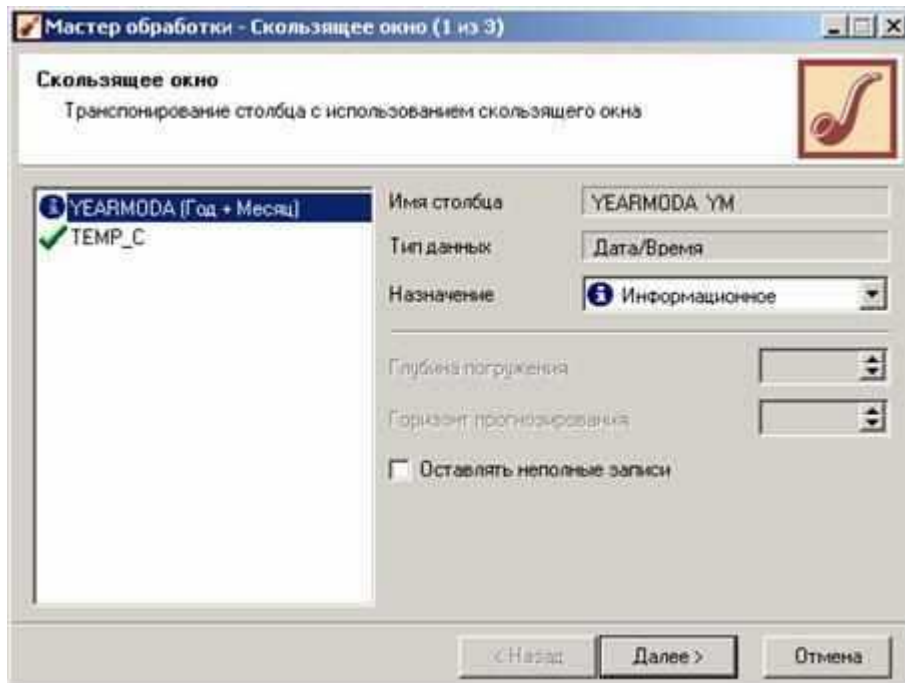


Рис. 16. Транспонирование столбца с использованием скользящего окна

Полученные данные и будем подавать на вход нейронной сети. Но, исходя из того факта, что делаем прогноз на ближайший месяц, то в качестве обучающей выборки будем использовать данные за последние 2 месяца + данные за 12 и 11 месяцев назад (ближайшие к прогнозируемому месяцу). Все остальные столбцы сделаем информационными (рис. 17).

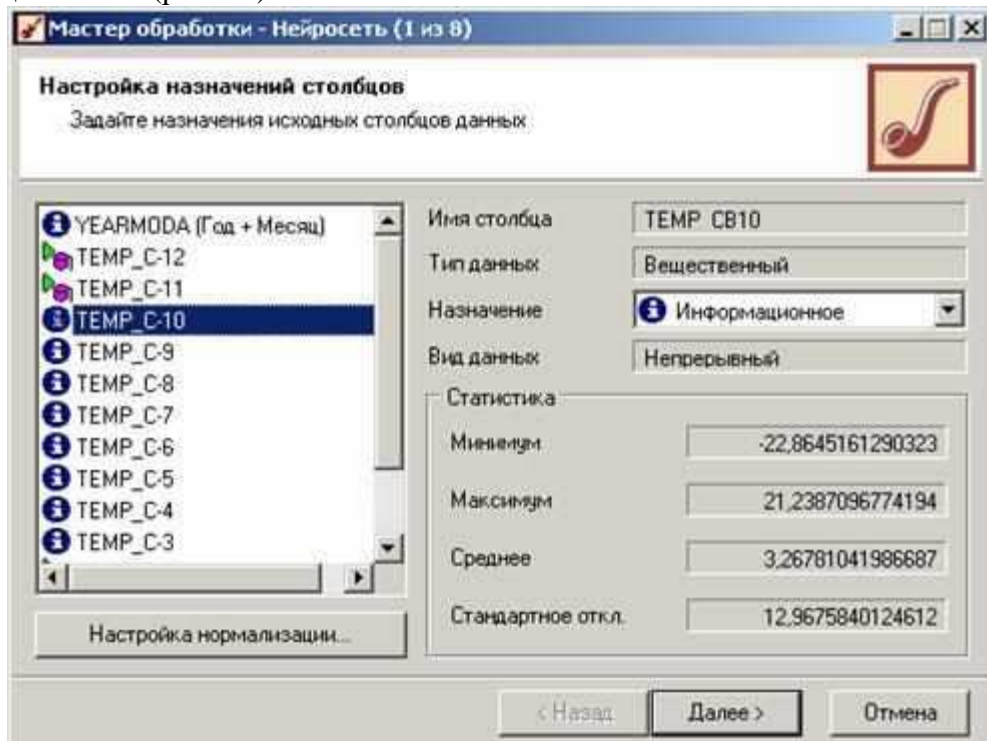


Рис. 17. Настройка назначений столбцов

Используя в созданной нейронной сети визуализатор «Диаграмма рассеяния» можно проанализировать полученные данные, и выяснить уровень значимости, которому они соответствуют.

Полученную нейронную сеть можно использовать для прогнозирования значений температуры, а для этого выберем в качестве мастера обработки «Прогнозирование» и, задав горизонт прогноза в 1 месяц (рис. 18) получаем прогноз для данного параметра.

Рис. 18. Настройка прогнозирования временного ряда

Полученные данные (прогноз) можем посмотреть через визуализатор «**Диаграмма прогноза**» (рис. 19), а точное спрогнозированное значение получить в табличных данных «**Таблица**».

Рис. 19. Диаграмма прогноза выбранного параметра EMP\_C на следующий месяц

1. Составить отчет:

- Титульный лист Цель работы Задание Ход работы
- Сравнительный анализ нейронных сетей различных конфигураций
- Выводы по предобработке входных данных (парциальная обработка и нормирование значений), а также выводы по выбору значимых входных параметров для нейронных сетей.



- Сводная таблица построенных нейронных сетей с описание количества слоев сети, количества нейронов в каждом слое, типа функции активации, выбранного алгоритма обучения.

Нейронная сеть	Кол-во слоев	Кол-во нейронов в слое	Функция активации	Алгоритм обучения
----------------	--------------	------------------------	-------------------	-------------------

- Сводная таблица характеристик функционирования нейронных сетей с описанием максимальной и средней ошибки на обучающем множестве, максимальной и средней ошибки на тестовом множестве, процент правильно распознанных случаев на обучающем множестве, процент правильно распознанных случаев на тестовом множестве, времени обучения.

Нейронная сеть	Max на обуч.	Avg на обуч.	Max на тест.	Avg на тест.	% прав. на обуч.	% прав. на тест.	Время обучения
----------------	--------------	--------------	--------------	--------------	------------------	------------------	----------------

- Выводы о достоинствах и недостатках различных конфигураций нейронных сетей.
- Описание матрицы весов нейронной сети, выводы о силе связей между входными и выходной переменными на основании граф нейронной сети.
- Прогнозирование временных рядов с помощью нейронной сети
- Обоснование выбора структуры нейронной сети, использованной для прогнозирования временного ряда.
- Выводы по результатам прогнозирования на основании диаграмм прогнозов выбранных параметров состояния погоды и воды.
- Выводы по работе

Материально-техническое обеспечение занятия:

Для материально-технического обеспечения дисциплины «Интеллектуальные информационные системы» необходимо:

1. Компьютерный класс – ауд. № 203.
2. 1 компьютер преподавателя, 12 компьютеров обучающихся, маркерная доска, проектор.
3. Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа

Windows 7	68526624	без даты
Microsoft office 2010 Pro	49420326	08.12.2011
Microsoft Visual Studio 2005	77626-009-0000007-41832	без даты
Mozilla Firefox	свободный доступ 647526	свободный доступ без даты
Matlab	2696062	26.10.2012

Mathcad Education - University edition Kaspersky Endpoint Security	1894141205092525	05.12.2014
--	------------------	------------

---

**9.2. Методические рекомендации по подготовке письменных работ****9.3. Иные материалы**

## АННОТАЦИЯ ДИСЦИПЛИНЫ

Дисциплина «Большие данные и нейронные сети» реализуется на факультете архивоведения и документоведения кафедрой источниковедения.

Цель дисциплины – приобретение знаний в области систем искусственного интеллекта (ИИ) и принятия решений (ПР); изучение программных средств конструирования интеллектуальных систем (ИС) для различных предметных областей.

Задачи дисциплины:

- ознакомление с концепциями и методами, составляющими основу для понимания современных достижений искусственного интеллекта;
- ознакомление с современными областями исследования по искусственному интеллекту;
- ознакомление с основными моделями представления знаний и некоторыми интеллектуальными системами;
- рассмотрение теоретических и некоторых практических вопросов создания и эксплуатации экспертных систем;
- ознакомление с особенностями практического использования интеллектуальных информационных систем и систем принятия решений.

Дисциплина направлена на формирование следующих компетенций:

- ПК-4 - Способность создавать и вести системы документационного обеспечения управления архивов на базе новейших технологий

В результате освоения дисциплины обучающийся должен:

**Знать:** определение интеллектуальных систем, структуру статических и динамических экспертных систем; теоретические основы построения и функционирования прикладных интеллектуальных систем и систем поддержки принятия решений, ключевые направления применения новых информационных систем при автоматизации процессов принятия управленческих решений; методы построения эксплуатации и разработки интеллектуальных систем; теорию технологий искусственного интеллекта; архитектуру и методы проектирования экспертных систем; модели представления знаний; современные системы искусственного интеллекта и принятия решений; возможности интеллектуальных систем и имеющихся программных продуктов; основные источники научно-технической информации по основным направлениям, методам, моделям и инструментальным средствам конструирования интеллектуальных систем.

**Уметь:** разрабатывать постановку задач для решения неформализованных проблем; формулировать цели и задачи автоматизации обработки управленческой информации; применять интеллектуальные системы для решения задач оценки и прогнозирования состояния объектов; разрабатывать и программировать диалоги взаимодействия ЭВМ и человека, решать оптимизационные задачи с помощью генетических алгоритмов; применять различные модели представления знаний при реализации экспертных систем на ЭВМ.

**Владеть:** терминологией, навыками поиска и использования научно-технической информации по профессиональной тематике; современными методами применения прикладных интеллектуальных систем и систем поддержки принятия решений; построением моделей представления знаний, подходами и техникой решения задач

искусственного интеллекта, информационных моделей знаний, методами представления знаний

По дисциплине предусмотрена промежуточная аттестация в форме зачета в 7 семестре.

Общая трудоемкость освоения дисциплины составляет 3 зачетные единицы.