



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал
Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Нализко Наталья Александровна
Должность: Директор
Дата подписания: 13.05.2024 16:15:21
Уникальный программный ключ:
25467908655d9e0abdc452e51caba97c16aa1bca

Версия документа - 1

стр. 1 из 74

Первый экземпляр _____

КОПИЯ № _____



УТВЕРЖДАЮ

Директор Костанайского филиала
ФГБОУ ВО «ЧелГУ»

[Signature] / Тюлегенова Р.А.

13 мая 2023 г.

**Фонд оценочных средств
текущего контроля
по дисциплине**

Интеллектуальный анализ данных

**Направление подготовки
38.03.02 Менеджмент**

**Направленность (профиль)
Бизнес логистика**

**Присваиваемая квалификация (степень)
Бакалавр**

**Форма обучения
очная**

**Год набора
2023, 2024**

Костанай, 2023 г.



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 2 из 74

Первый экземпляр _____

КОПИЯ № _____

Фонд оценочных средств принят

Учёным советом Костанайского филиала ФГБОУ ВО «ЧелГУ»

Протокол заседания № 10 от 25 мая 2023 г.

Председатель учёного совета
филиала

Р.А. Тюлегенова

Секретарь учёного совета
филиала

Н.А. Кравченко

Фонд оценочных средств рекомендован

Учебно-методическим советом Костанайского филиала ФГБОУ ВО «ЧелГУ»

Протокол заседания № 10 от 18 мая 2023 г.

Председатель
Учебно-методического совета

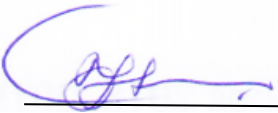
Н.А. Нализко


Фонд оценочных средств разработан и рекомендован кафедрой социально-гуманитарных и естественнонаучных дисциплин

Протокол заседания № 10 от 16 мая 2023 г.

Заведующий кафедрой

Т.К. Нуртаzenов

Автор (составитель)  Карасева Э.М., профессор кафедры социально-гуманитарных и естественнонаучных дисциплин, кандидат педагогических наук

 Рак О.В., старший преподаватель кафедры социально-гуманитарных и естественнонаучных дисциплин, магистр



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 3 из 74

Первый экземпляр _____

КОПИЯ № _____

1. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ

Направление подготовки: 38.03.02 Менеджмент

Направленность (профиль): Бизнес логистика

Дисциплина: Интеллектуальный анализ данных

Семестр (семестры) изучения: 4

Форма (формы) текущего контроля: устный опрос, отчетность по практической работе

Оценивание результатов учебной деятельности обучающихся при изучении дисциплины осуществляется по балльно-рейтинговой системе

2. КОМПЕТЕНЦИИ, ЗАКРЕПЛЁННЫЕ ЗА ДИСЦИПЛИНОЙ

2.1. Компетенции, закреплённые за дисциплиной

Изучение дисциплины «Интеллектуальный анализ данных» направлено на формирование следующих компетенций:

Коды компетенции (по ФГОС)	Содержание компетенций согласно ФГОС	Перечень планируемых результатов обучения по дисциплине	
1	2	3	
ОПК-2.2	Выбирает соответствующие содержанию управленческих задач методы и приемы сбора, обработки и анализа данных с применением современного инструментария и интеллектуальных информационно-аналитических систем	пороговый	Знать: методологию сбора, обработки и систематизации данных, необходимых для решения поставленных управленческих задач
			Уметь: выбирать методы и приемы сбора, обработки и анализа данных, соответствующие содержанию управленческих задач
			Владеть: навыками сбора, обработки и систематизации данных, необходимых для решения поставленных управленческих задач
		продвинутый	Знать: современный инструментарий и интеллектуальные информационно-аналитические системы для сбора, обработки и анализа данных
			Уметь: выбирать современный инструментарий и интеллектуальные информационно-аналитические системы для сбора, обработки и анализа данных
			Владеть: современным инструментарием и интеллектуальными информационно-аналитическими системами для сбора, обработки и анализа данных
		высокий	Знать: методологию анализа данных, необходимых для решения поставленных управленческих задач
			Уметь: анализировать данные, необходимые для решения поставленных управленческих задач
			Владеть: навыками анализа данных с применением современного инструментария и интеллектуальных информационно-



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по
основной профессиональной образовательной программе высшего образования – программе
бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 4 из 74

Первый экземпляр _____

КОПИЯ № _____

ОПК-2.3	Интерпретирует полученные результаты анализа с помощью современного инструментария и интеллектуальных информационно-аналитических систем для решения управленческих задач	пороговый	Знать:	аналитических систем объективные тенденции экономического развития, закономерности функционирования экономических систем, взаимодействие экономических процессов и их социальное содержание
			Уметь:	на основе изучения экономических процессов и явлений в их взаимосвязи и взаимозависимости, выявлять происходящие в них положительные и отрицательные тенденции для решения управленческих задач
			Владеть:	навыками экономических расчетов и формирования аналитических выводов по результатам проведенного анализа для решения управленческих задач
		продвинутый	Знать:	приемы систематизации, детализации, моделирования различных факторов деятельности экономических субъектов для интерпретации результатов анализа при решении управленческих задач
			Уметь:	комплексно и системно интерпретировать полученные результаты анализа деятельности экономических субъектов во всех существующих аспектах
			Владеть:	навыками интерпретации результатов анализа на основе систематизации, детализации и моделирования различных факторов деятельности экономических субъектов для решения управленческих задач
		высокий	Знать:	приемы интерпретации результатов анализа на основе применения современных методов социально-экономической диагностики и информационно-аналитических систем для решения управленческих задач
			Уметь:	интерпретировать полученные результаты анализа на основе применения современных методов социально-экономической диагностики, получения обобщенных характеристик, агрегирования информации, ее обработки с помощью современного инструментария и интеллектуальных информационно-аналитических систем
			Владеть:	приемами интерпретации результатов проведенных исследований и расчетов с помощью современного инструментария и интеллектуальных информационно-аналитических систем для объяснения пользователям в системе управления организацией
ОПК-5.3	Применяет принципы	пороговый	Знать:	сущность и содержание понятий «информационные технологии» и



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 5 из 74

Первый экземпляр _____

КОПИЯ № _____

управления крупными массивами данных и их интеллектуальный анализ при решении профессиональных задач		«программные средства».
		Уметь: применять информационные технологии для решения профессиональных задач
		Владеть: навыками применения информационных технологий в управлении
	продвинутый	Знать: понятия «базы данных» и «массивы данных»
		Уметь: структурировать данные
		Владеть: навыками работы с информационными системами и базами данных
	высокий	Знать: принципы интеллектуального анализа
		Уметь: анализировать данные при решении профессиональных задач
		Владеть: принципами управления крупными массивами данных и их интеллектуальным анализом при решении профессиональных задач

3. СОДЕРЖАНИЕ ОЦЕНОЧНЫХ СРЕДСТВ ПО ДИСЦИПЛИНЕ

3.1 Виды оценочных средств

№ п/п	Контролируемые разделы	Код компетенции/планируемые результаты обучения	Наименование оценочного средства
1.	Системные задачи	ОПК-2.2, ОПК-2.3, ОПК-5.3	Устный опрос, отчетность по практической работе
2.	Статистические задачи	ОПК-2.2, ОПК-2.3, ОПК-5.3	Устный опрос, отчетность по практической работе
3.	Когнитивные технологии	ОПК-2.2, ОПК-2.3, ОПК-5.3	Устный опрос, отчетность по практической работе

3.2 Содержание оценочных средств

Оценочные средства представлены в форме перечня теоретических вопросов для устного опроса, МУ для практических работ и заданий для СРС.

Для очно-заочной и заочной формы обучения задания прикрепляются в СДО «Moodle». Количество СРС может быть изменено в соответствии с рабочей программой дисциплины.

3.2.1 Темы для конспектирования

Раздел 1. Системные задачи

Системные задачи ИАД Введение в ИАД. Data Mining и аналитические информационные технологии. Цели, задачи, математический и программный инструментария ИАД. Прикладные задачи. ИАД в задачах математического моделирования сложных систем. Применение ИАД в задачах моделирования сложных динамических систем, включая технологические и экономические процессы Проблема неопределенности. Вероятностный дескрипт. Проблема неопределенности, как главная задача анализа данных в интересах задач управления. Вероятностные методы описания неопределенности. /Лек/



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по
основной профессиональной образовательной программе высшего образования – программе
бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 6 из 74

Первый экземпляр _____

КОПИЯ № _____

Раздел 2. Статистические методы

Статистические методы ИАД

Принципы статистического анализа данных средствами ИАД. Статистические методы и алгоритмы ИАД. Особенности применения. Ограничения. Восстановление зависимостей. Линейная регрессия в задачах анализа данных. Восстановление зависимостей и связей одна из основных задач ИАД. Методы регрессионного анализа данных.

Статистические решения: Методы проверки гипотез. Статистические решения как терминальная задача ИАД. Решения на основе метода проверки гипотез. Задачи классификации. Дискриминантный анализ данных. Классификационные задачи ИАД. Классификация «с учителем». Алгоритмы дискриминантного анализа данных.

Раздел 3. Когнитивные технологии

Когнитивные технологии ИАД. Современные технологии прогнозирования. Прогнозирование развития сложных многомерных процессов методами ИАД. Введение в искусственные нейронные сети (ИНС). Модели нейронов. Перцептроны. Математическое описание ИНС. Элементы и простейшие модели ИНС. Эволюционные алгоритмы прогноза и распознавания. Прогнозирование и Распознавание состояния сложных систем на основе метода эволюционного моделирования.

3.2.2.Перечень примерных теоретических вопросов для проведения устного опроса

1. Современные подходы к моделированию: аналитический и информационный.
2. Этапы моделирования в интеллектуальном анализе данных (ИАД). Принципы анализа данных.
3. Виды данных и их структурирование.
4. Подготовка данных к анализу: извлечение, формализация.
5. Методика сбора данных.
6. Этапы анализа данных: выборка, очистка, трансформация, интерпретация.
7. Машинное обучение в ИАД: обучающая выборка, обучение с учителем и без, ошибка обучения, ошибка обобщения, эффект переобучения.
8. Программное обеспечение в ИАД, аналитические платформы ИАД
9. Методики оценки качества данных.
10. Организация мониторинга качества данных.
11. Очистка и предобработка данных.
12. Фильтрация данных
13. Выявление аномальных значений.
14. Сокращение размерности данных.
15. Постановки задачи ассоциации: основные категории, ассоциативное правило, достоверность правила.
16. Методы поиска ассоциативных правил: алгоритм AIS, алгоритм SETM, алгоритм

3.2.3.Методические рекомендации к выполнению практических заданий

Раздел 1. Системные задачи



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 7 из 74

Первый экземпляр _____

КОПИЯ № _____

Практическая работа 1

Анализ признаков и оценка их информативности

Продолжительность работы – 8 час.

Цель работы: ознакомиться с возможностями аналитического пакета *Deductor Academic*.

Программа работы

1. Выполнить импорт данных в программный комплекс *Deductor*.
2. Выполнить задание по предварительной парциальной обработке данных.
3. Выполнить задание по предварительной обработке путем удаления аномалий в данных.
4. Выполнить задание по предварительной обработке путем сглаживания данных методом спектральной обработки.
5. Выполнить задание по удалению шумов на этапе предварительной обработки данных.
6. Ознакомиться с возможностями автоматического анализа качества импортируемых данных.

Методические указания по выполнению работы

Импорт данных в программный комплекс *Deductor Academic*

Импорт данных является отправной точкой анализа данных. Импорт в *Deductor* может осуществляться из популярных форматов хранения данных, таких как *Excel*, *Access*, *MS SQL*, *Oracle*, Текстовый файл и прочих. Кроме того, имеется универсальный доступ к любому источнику данных посредством ADO или ODBC (Только в коммерческой версии, в бесплатной версии возможен импорт из *.txt, *.csv и *.ded).

Импорт данных из текстового файла с разделителями осуществляется путем вызова мастера импорта на панели «Сценарии» (рис. 1.1).



Рис. 1.1. - Панель сценарии

После запуска мастера импорта укажем тип импорта «Текстовый файл» и перейдем к настройке импорта (рис. 1.2-3). Укажем имя файла, из которого необходимо получить данные. В окне просмотра, выбранного файла можно увидеть содержание данного файла.



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 8 из 74

Первый экземпляр _____

КОПИЯ № _____

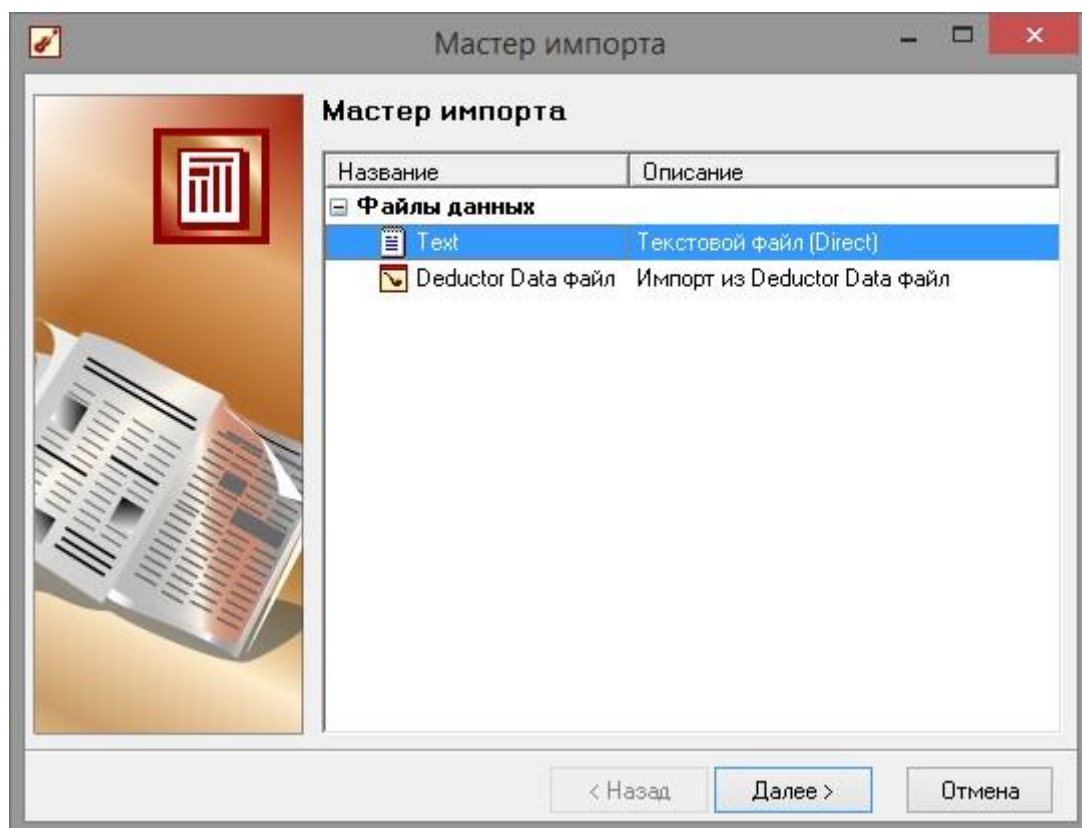


Рис. 1.2 - Мастер импорта

Далее перейдем к настройке параметров импорта (рис. 1.4). На этой странице мастера предоставляется возможность указать, с какой строки следует начать импорт, указать то, что первая строка является заголовком, возможность добавить первичный ключ. Указать, что является символом-разделителем столбцов, а также указать ограничитель строк, разделитель целой и дробной части вещественного числа, разделитель компонентов даты и ее формат.



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 9 из 74

Первый экземпляр _____

КОПИЯ № _____

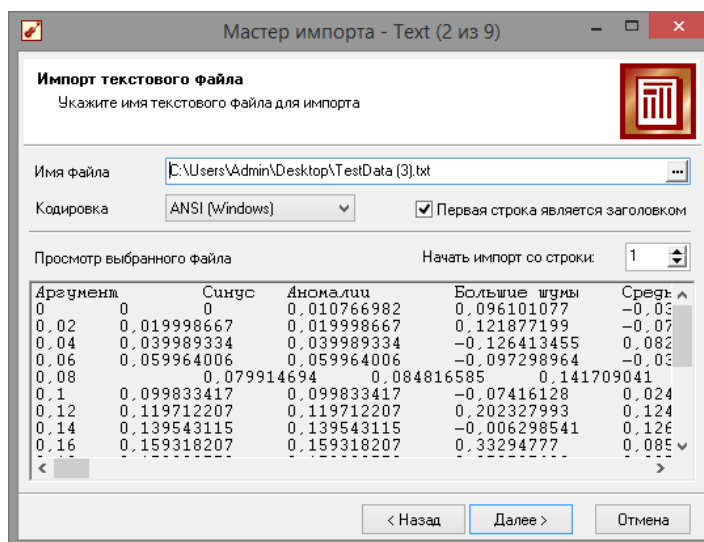


Рис. 1.3 - Выбор файла

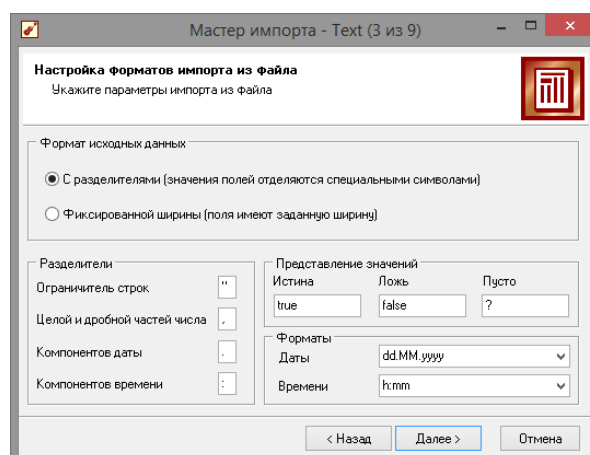


Рис. 1.4 - Параметры импорта

В данном случае параметры по умолчанию на этой страницемастера установлены правильно, а именно: начать импорт с первой строки, первая строка является заголовком, разделителем между столбцами является знак табуляции, разделителем целой и дробной частей является запятая. Далее перейдем к настройке свойств полей (рис. 1.5).



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по
основной профессиональной образовательной программе высшего образования – программе
бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 10 из 74

Первый экземпляр _____

КОПИЯ № _____

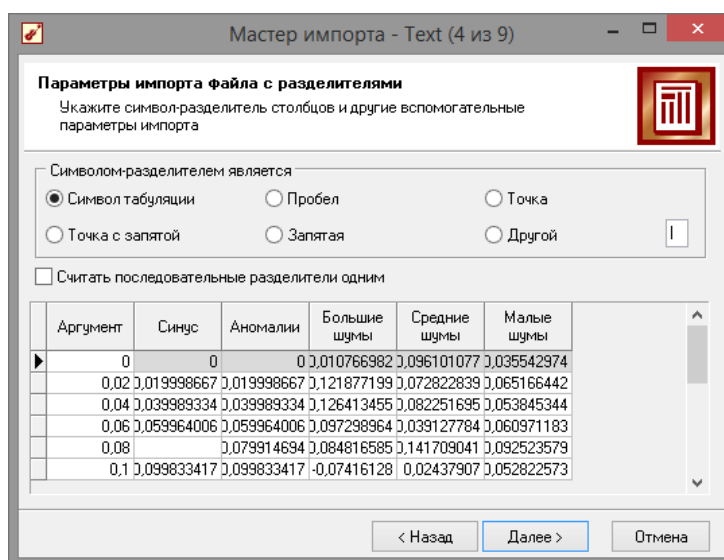


Рис. 1.5 - Параметры разделителей

На этом шаге мастера предоставляется возможность настроить имя, название (метку), размер, тип данных, вид данных и назначение (рис 1.6). Некоторые свойства (например, тип данных) можно задавать для выделенного набора столбцов. Вид данных определяет – конечный ли это набор (дискретные) или бесконечный (непрерывные). Назначение столбцов определяет характер их использования в алгоритмах обработки (при импорте можно оставить значение по умолчанию). Необходимо убедиться, что в данном случае тип данных у все столбцов выставлен как вещественный, так как в более старых версиях тип определялся по первой строке, а в демо-примере столбцы «Аргумент», «Синус» и «Аномалии» имеют в первой строке значение «0», что могло приводить к неправильному определению типа данных.

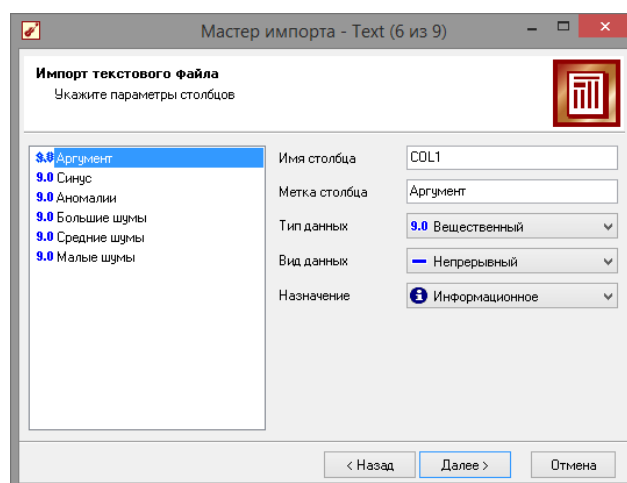


Рис. 1.6 - Параметры столбцов



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 11 из 74

Первый экземпляр _____

КОПИЯ № _____

Далее осталось только выполнить импорт данных, нажав на кнопку «Пуск» на следующем шаге мастера импорта (рис. 1.7). После импорта данных на следующем шаге мастера необходимо выбрать способ отображения данных (рис. 1.8). В данном случае самым информативным является диаграмма, выберем ее.

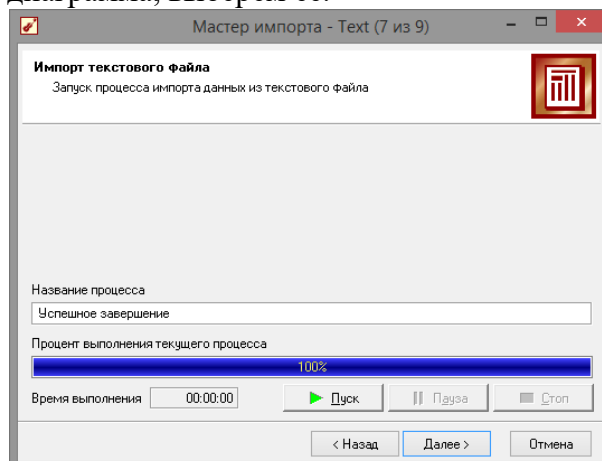


Рис. 1.7 - Импорт файла

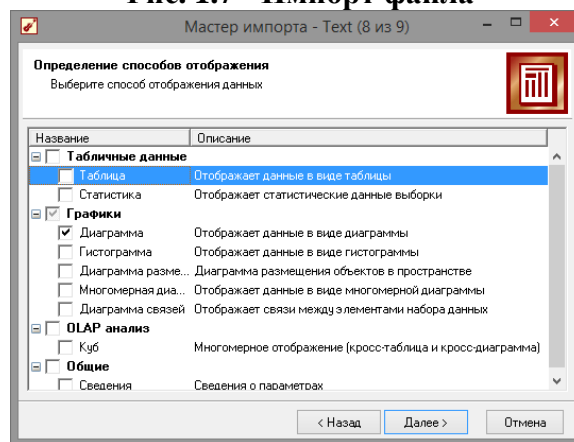


Рис. 1.8 - Способ отображения

От того, какие способы отображения будут выбраны на этом этапе, зависят последующие шаги мастера. В данном случае необходимо настроить, какие столбцы диаграммы следует отображать и как именно. Выберем для отображения поле «СИНУС» и тип диаграммы «Линии» (рис. 1.9).



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по
основной профессиональной образовательной программе высшего образования – программе
бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 12 из 74

Первый экземпляр _____

КОПИЯ № _____

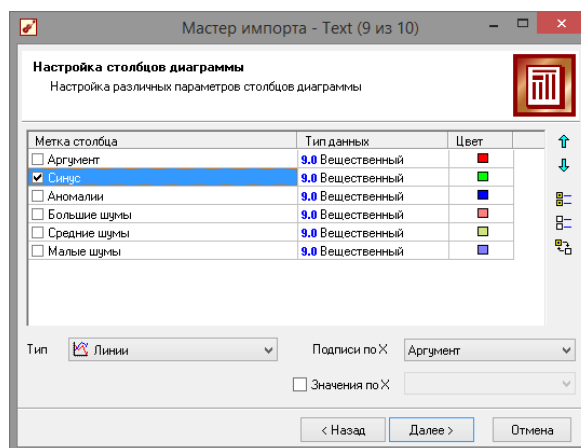


Рис. 1.9 - Настройка столбцов

На последнем шаге мастера необходимо указать название ветки в дереве сценариев. Напишем в поле заголовка окна «Импорт примера для демонстрации предобработки данных» и нажмем «Готово» (рис. 1.10). На этом работа мастера импорта заканчивается. Теперь в дереве сценариев появится новый узел с необходимыми данными. В главном окне программы представлены все выбранные отображения данных этого узла. В данном случае только диаграмма. Примечание: для отображение диаграммы в 3D-виде, необходимо нажать кнопку «3-х мерный вид» в левом верхнем углу панели «Диаграмма». А для просмотра другой диаграммы, нажать на значок лупы «Отображать поля» (рис. 1.11).

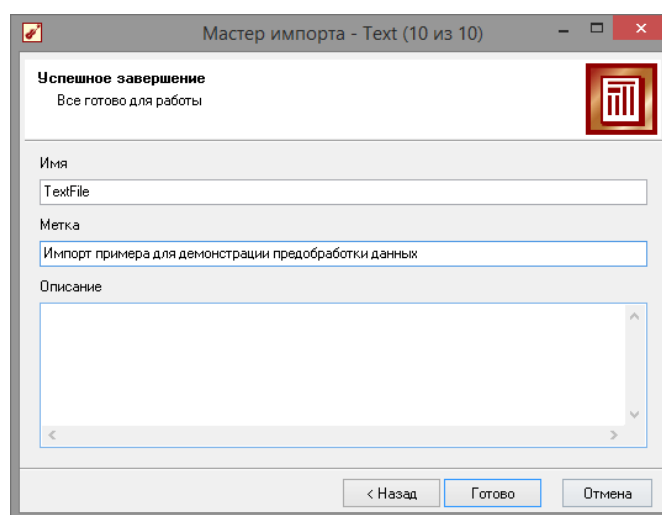
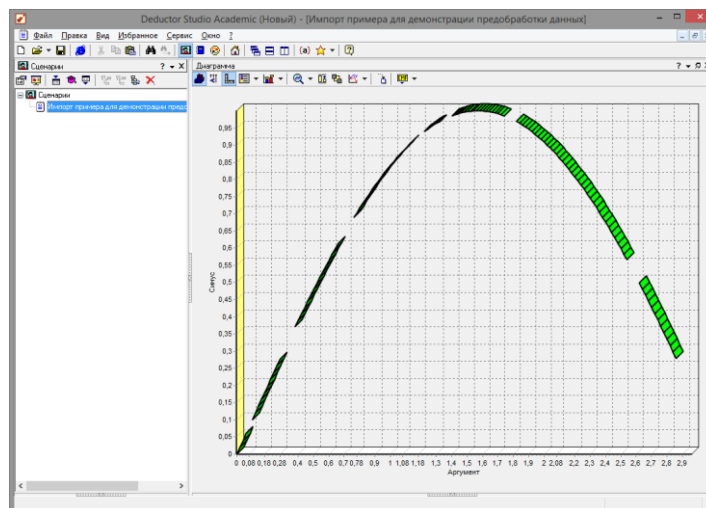


Рис. 1.10 – Завершение импорта

**Рис. 1.11 - Диаграмма синуса**

1.1 Предварительная парциальная обработка

Часто исходные данные для анализа не годятся, а качество данных влияет на качество результатов, поэтому вопрос подготовки данных для последующего анализа является очень важным. Обычно

«сырые» данные содержат в себе различные шумы, за которыми трудно увидеть общую картину, а также аномалии – влияние случайно, либо редко происходивших событий. Очевидно, что влияние этих факторов на общую модель необходимо минимизировать, т.к. модель, учитывающая их, получится неадекватной.

Парциальная предобработка служит для восстановления пропущенных данных, редактирования аномальных значений и спектральной обработке данных (например, сглаживания данных). Именно этот шаг часто проводится в первую очередь.

Рассмотрим применение обработки на примере данных из файла

«TestData.txt». Он содержит таблицу со следующими полями:

«АРГУМЕНТ» – аргумент, «СИНУС» – значения синуса аргумента (некоторые значения пустые), «АНОМАЛИИ» – синус с выбросами,

«БОЛЬШИЕ ШУМЫ» – значения синуса с большими шумами,

«СРЕДНИЕ ШУМЫ» – значения синуса со средними шумами,

«МАЛЫЕ ШУМЫ» – значения синуса с малыми шумами. Все данные можно увидеть на диаграмме после импорта из текстового файла.

Часто бывает так, что в столбце некоторые данные отсутствуют в силу каких-либо причин (данные не известны, либо их забыли внести и т.п.). Обычно из-за этого пришлось бы убрать из обработки все строки, которые содержат пропущенные данные. Но механизмы



Deductor Studio позволяют решить эту проблему. Один из шагов парциальной обработки как раз отвечает за восстановление пропущенных значений. Если данные упорядочены (например, по времени), то рекомендуется в качестве восстановления пропущенных значений использовать аппроксимацию. Алгоритм сам подберет значение, которое должно стоять на месте пропущенного значения, основываясь на близлежащих данных. Если же данные не упорядочены, то следует использовать режим максимального правдоподобия, когда алгоритм подставляет вместо пропущенных данных наиболее вероятные значения, основываясь на всей выборке.

Для демонстрации воспользуемся мастером заполнения пропусков. Импортировав файл можно увидеть, что в столбце «СИНУС» содержатся пустые значения. На диаграмме выше видно, что некоторые значения синуса пропущены. Для дальнейшей обработки необходимо их восстановить. Для этого следует запустить мастер заполнения пропусков. На рис.1.12 показаны различные варианты столбцов: с пропущенными данными; с аномалиями (выбросами); с большими шумами; со средними шумами; с малыми шумами.

Для запуска мастера необходимо выделить нужный сценарий и нажать F7, либо правый клик по необходимому сценарию откроет контекстное меню, где так же можно выбрать мастер обработки. Поскольку данные в исходном наборе упорядочены, на следующем шаге мастера обработки поставим галочку – «обрабатывать как упорядоченный набор» (рис. 1.13).

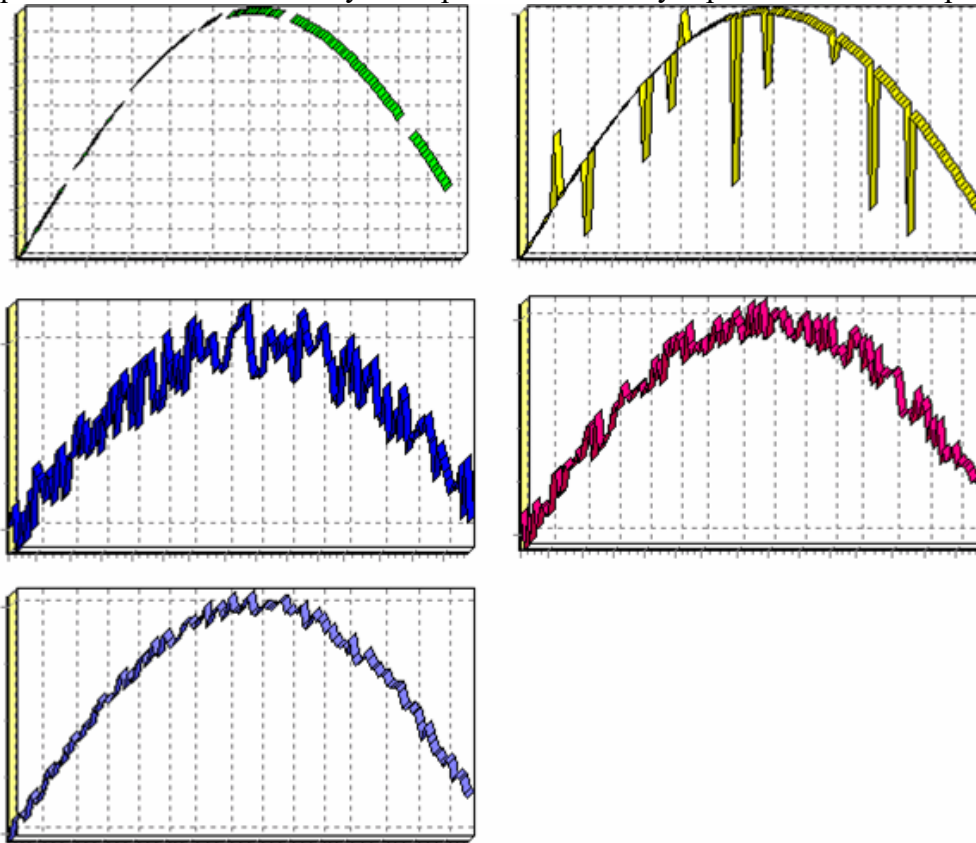


Рисунок 1.12 - Варианты столбцов



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 15 из 74

Первый экземпляр _____

КОПИЯ № _____

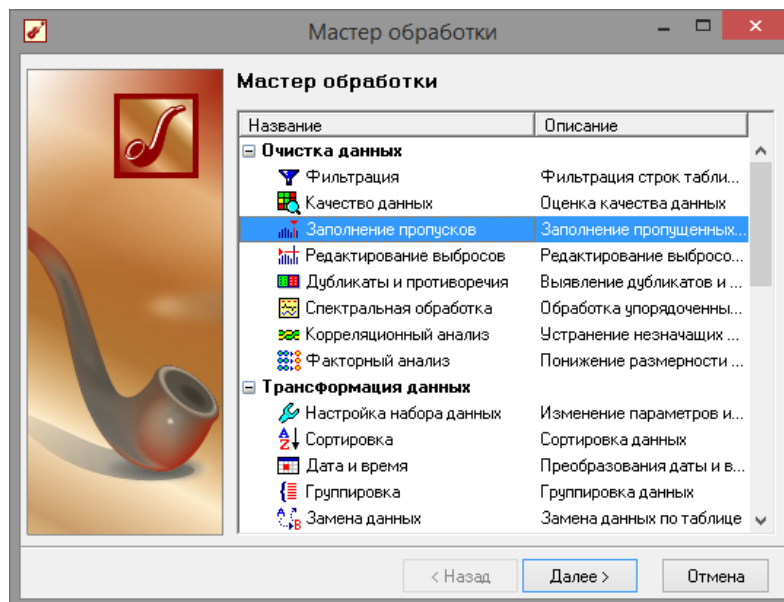


Рисунок 1.13 - Мастер обработки

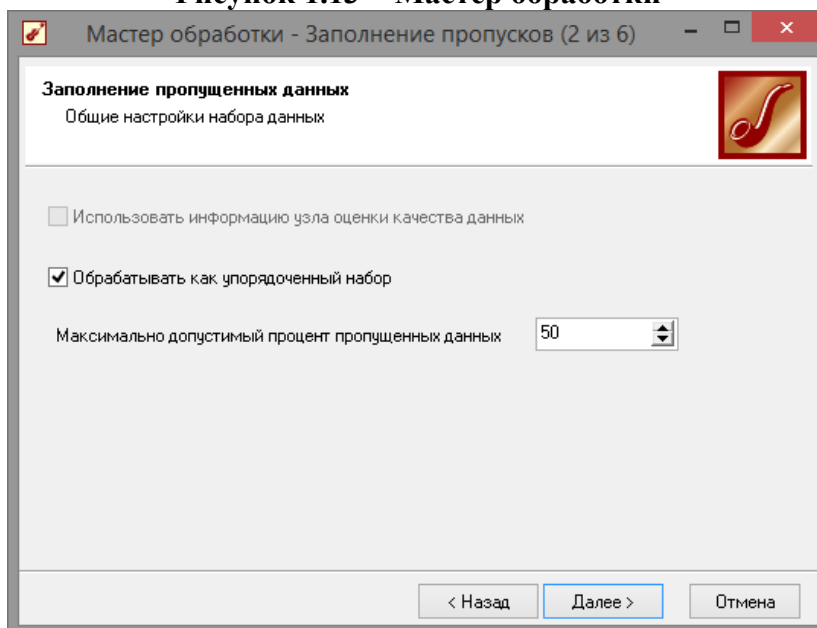


Рис. 1.14 - Мастер заполнения пропусков

Далее следует выбрать необходимый столбец и метод заполнения, в данном случае интерполяция (рис. 1.14). Перейдя на страницу запуска процесса обработки, выполняем ее, нажав на пуск, и далее выбираем тип визуализации обработанных данных (как в примере импорта) (рис. 1.15).. После выполнения процесса обработки на диаграмме видно, что пропуски в данных исчезли, что и было необходимо сделать (рис. 1.16).



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 16 из 74

Первый экземпляр _____

КОПИЯ № _____

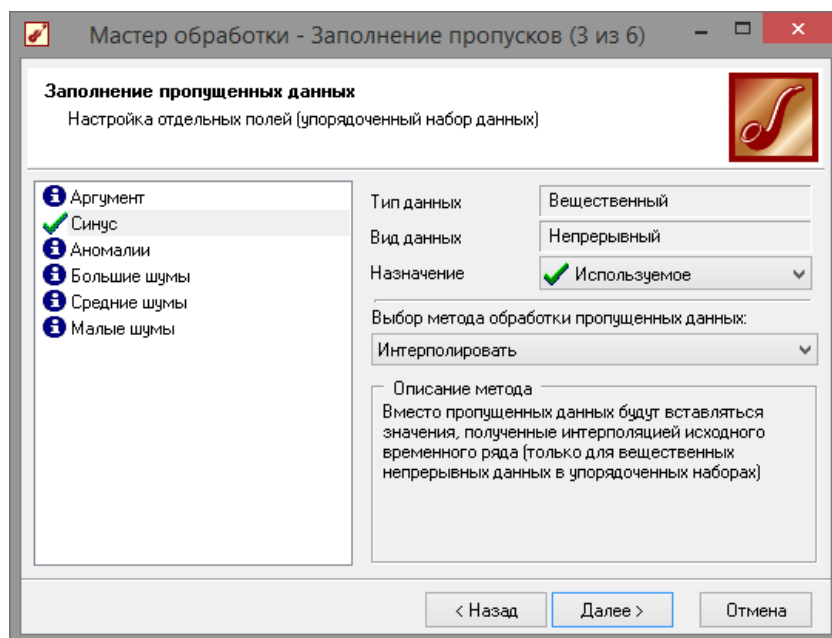


Рис. 1.15 - Настройки мастера заполнения пропусков

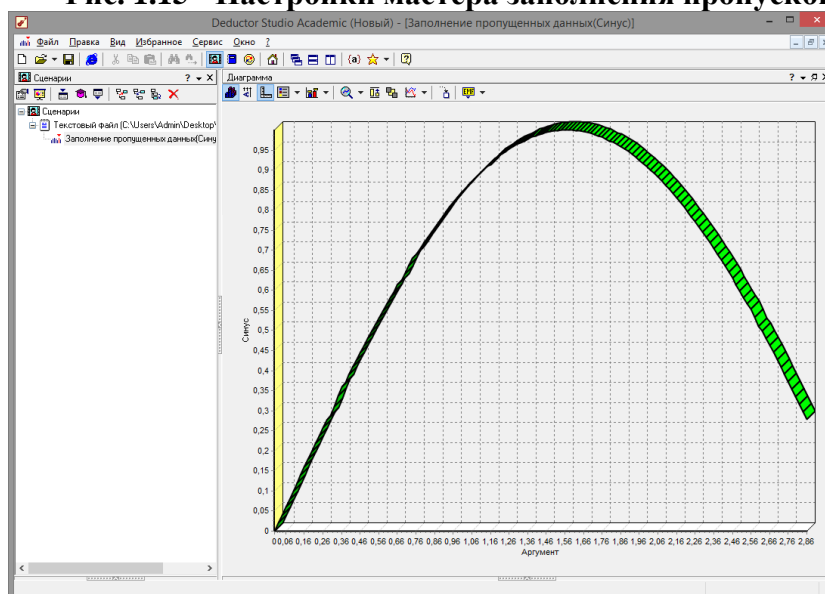


Рис.1.16 - Заполнение пропусков методом интерполяции

1.2 Удаление аномалий на этапе предобработки данных

Аномалии встречаются в «сырых» данных не реже шумов. По существу, они вообще не должны оказывать никакого влияния на результат. Если же они присутствуют при построении модели, то оказывают на нее весьма большое влияние и их предварительно необходимо устранить. Также они портят статистическую картину распределения данных. К примеру, вот как выглядят данные с аномалиями, а также гистограмма их распределения



(рис. 1.17).

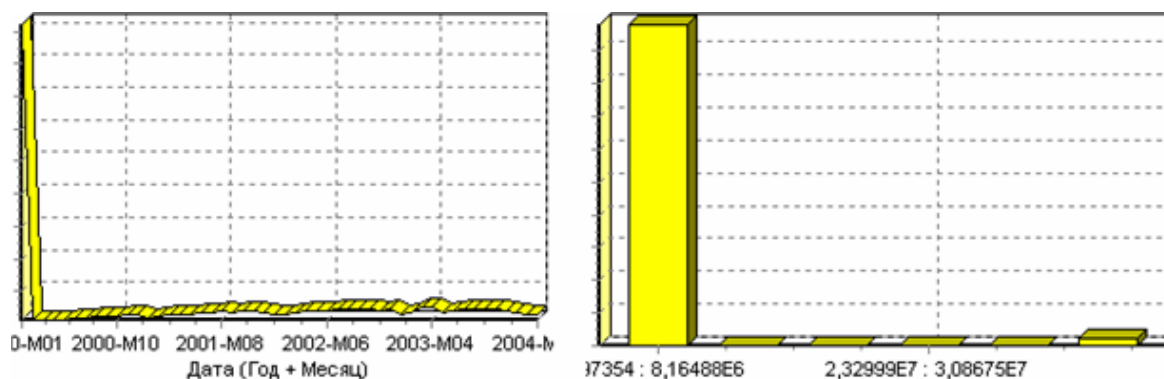


Рис. 1.17 - Гистограмма с аномалиями

Очевидно, что аномалии не позволяют определить, как характер самих данных, так и статистическую картину. После устранения аномалий те же данные представляются как показано на рис. 1.18

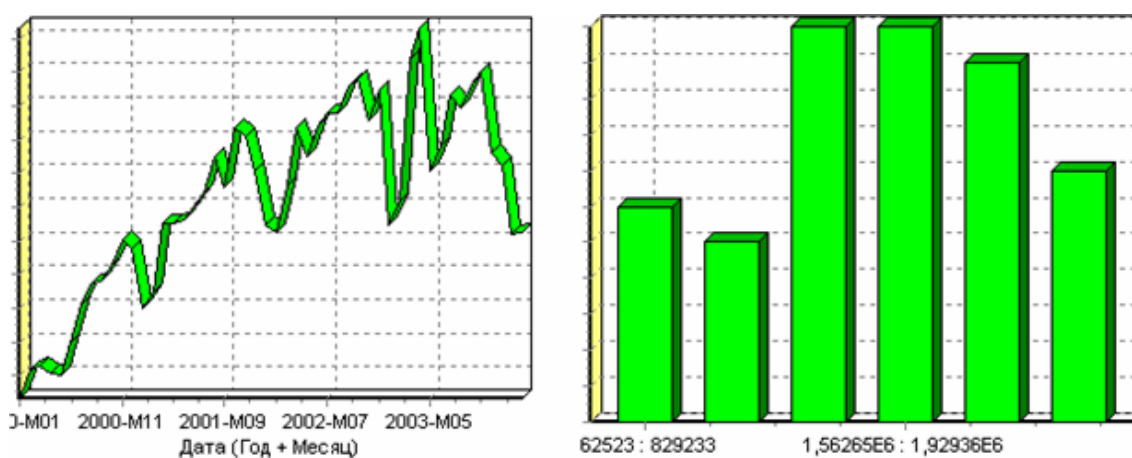


Рис. 1.18 - Гистограмма без аномалий

Следует открыть мастер обработки и выбрать редактирование выбросов (рис. 1.19).
Поставить галочку - «Обрабатывать как упорядоченный набор данных» (рис. 1.20).



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 18 из 74

Первый экземпляр _____

КОПИЯ № _____

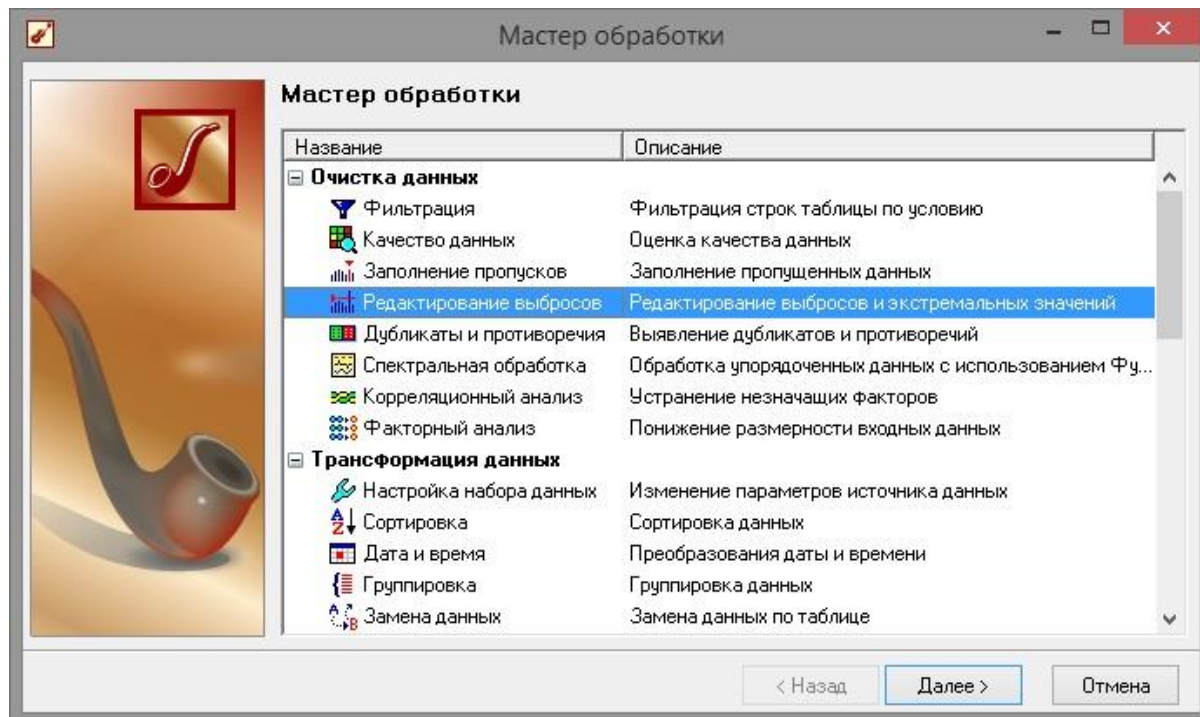


Рис. 1.19 - Мастер редактирования выбросов

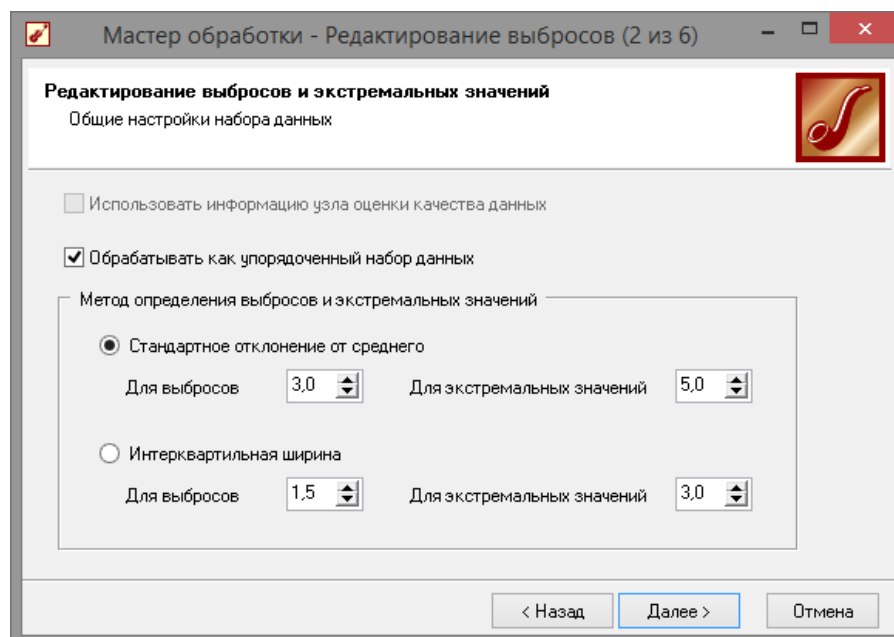



Рис. 1.20 - Настройки мастера редактирования выбросов

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ») Костанайский филиал Кафедра социально-гуманитарных и естественнонаучных дисциплин			
Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент			
Версия документа - 1	стр. 19 из 74	Первый экземпляр _____	КОПИЯ № _____

На следующем шаге (рис. 1.21) необходимо выбрать назначение «Используемое» только для необходимого столбца данных. И выставить большую степень подавления, так как выбросы существенны. Для множественного выделения столбцов можно использовать клавишу *Shift* и *Ctrl*.

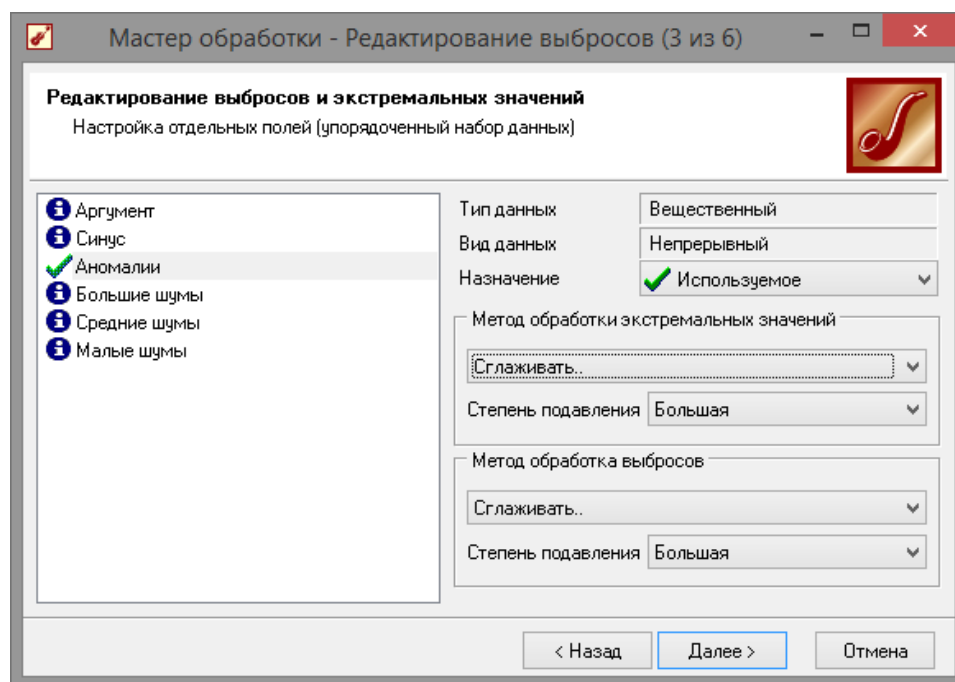


Рис. 1.21 - Настройки отдельных полей мастера редактирования выбросов

Далее нажать кнопку «Пуску» и выбрать данные для отображения как в предыдущих пунктах. После выполнения процесса обработки на диаграмме видно, что выбросы исчезли, остались лишь небольшие возмущения, которые легко сгладить при помощи спектральной обработки (рис. 1.22).

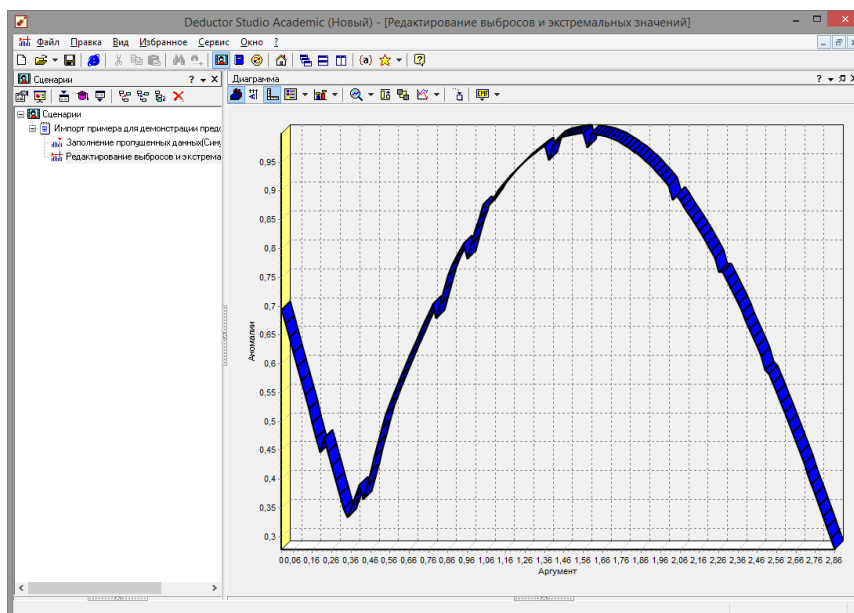


Рис. 1.22 - Диаграмма после удаления аномалий

1.3 Сглаживания данных методом спектральной обработки

Сглаживание данных применяется для удаления шумов из исходного набора, а также для выделения тенденции, которая трудно видна в исходном наборе. Платформа *Deductor Studio* предлагает несколько видов спектральной обработки: сглаживание данных путем указания полосы пропускания, вычитание шума путем указания степени вычитания шума и вейвлет преобразование путем указания глубины разложения и порядка вейвлета.

Для выбора способа вейвлет преобразования открыть мастер обработки для сценария «Редактирование выбросов и экстремальных значений». В мастере следует выбрать пункт «спектральная обработка» (рис. 1.23)



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 21 из 74

Первый экземпляр _____

КОПИЯ № _____

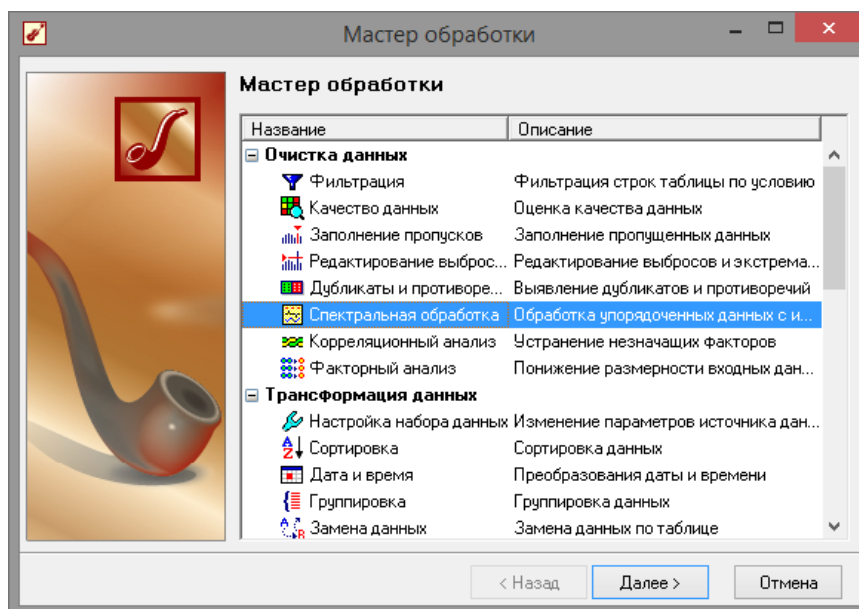


Рис. 1.23 - Мастер обработки

В мастере спектральной обработки необходимо выбрать назначение используемое для столбца «Аномалии», и в качестве метода сглаживания данных выбрать ветвь преобразование (рис. 1.24).

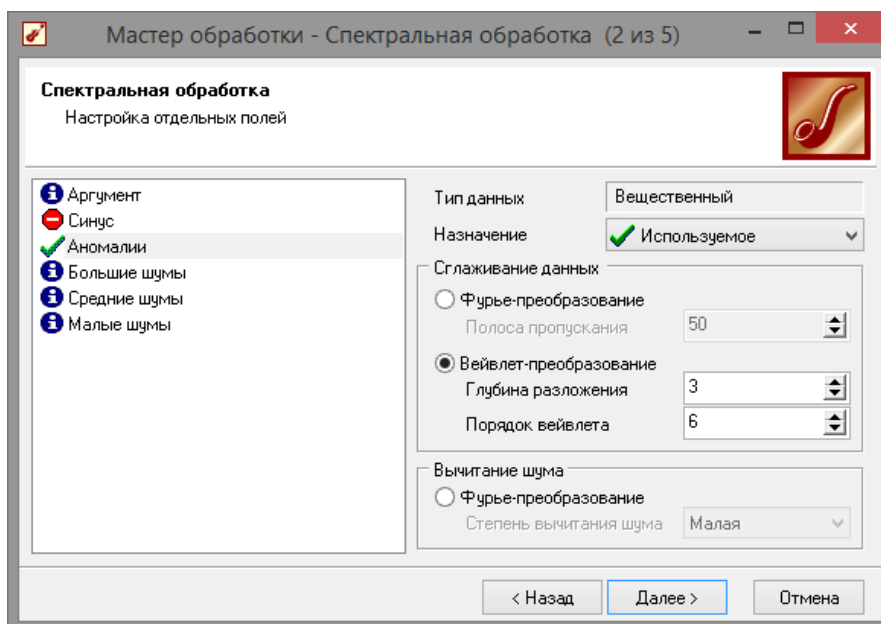
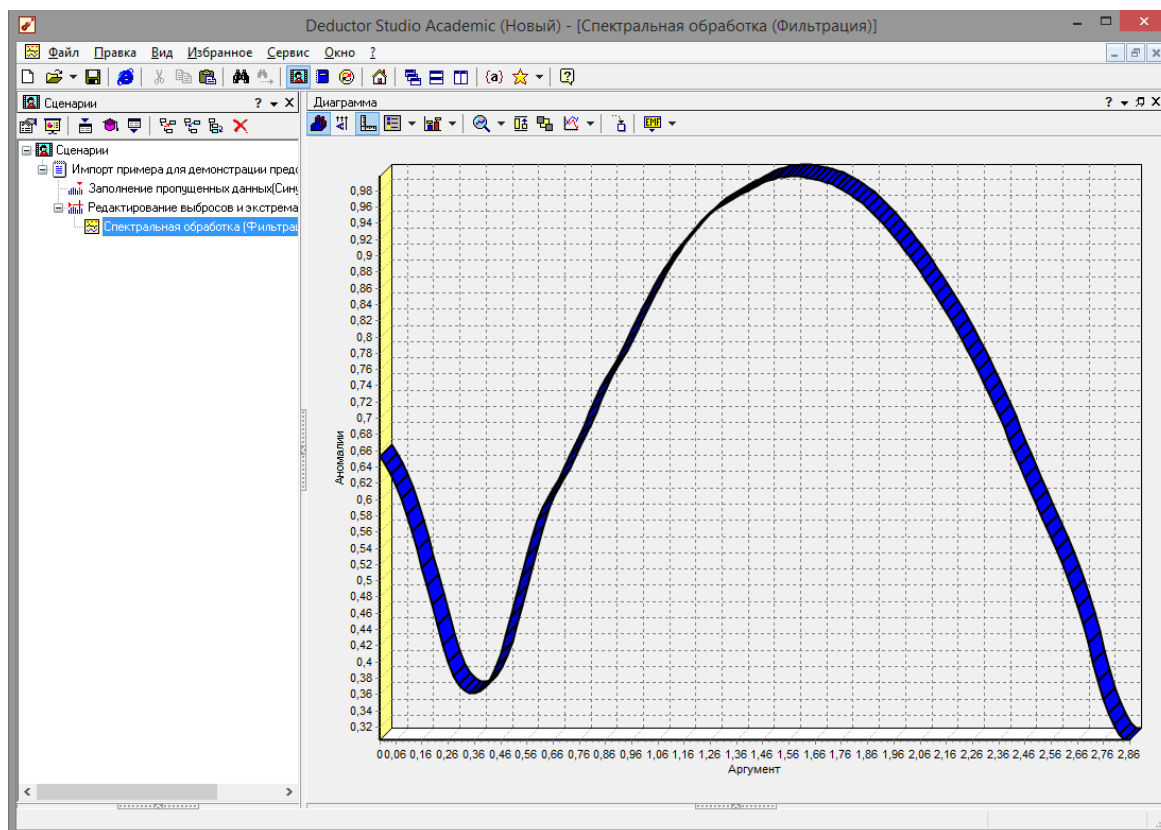


Рис. 1.24 - Мастер спектральной обработки

Далее выбрать визуализацию «Диаграмма» и столбец «Аномалии». После обработки можно убедиться на диаграмме в отсутствии выбросов (рис.



1.25).

**Рис. 1.25 - Диаграмма после применения спектральной обработки**

1.4 Удаление шумов на этапе предварительной обработки данных

Шумы в данных не только скрывают общую тенденцию, но и проявляют себя при построении модели прогноза. Из-за них модель может получиться с плохими обобщающими качествами.

В примере по парциальной обработке есть 3 столбца с шумами: «БОЛЬШИЕ ШУМЫ», «СРЕДНИЕ ШУМЫ», и «МАЛЫЕ ШУМЫ» - соответственно синус с большими, средними и малыми шумами. Ясно, что для дальнейшей работы с данными эти шумы необходимо устранить. Спектральная обработка позволяет сделать это с помощью указания для этих полей в качестве типа обработки «Вычитание шума». Настройки обладают определенной гибкостью. Так, существует большая, средняя и малая степень вычитания шума. Аналитик может подобрать степень, устраивающую его.

В мастере спектральной обработки (рис. 1.26) по очереди выбрать поля «БОЛЬШИЕ ШУМЫ», «СРЕДНИЕ ШУМЫ» и «МАЛЫЕ ШУМЫ», задать тип обработки «Вычитание шума» и указать степень



МИНОБРАЗОВАНИЯ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 23 из 74

Первый экземпляр _____

КОПИЯ № _____

подавления – «большая», «средняя» и «малая» соответственно. В некоторых случаях неплохие результаты удаления шумов дает вейвлет преобразование. Повысить качество сглаживания шумов таким способом можно, путем подбора удовлетворительных параметров обработки (рис. 1.27).

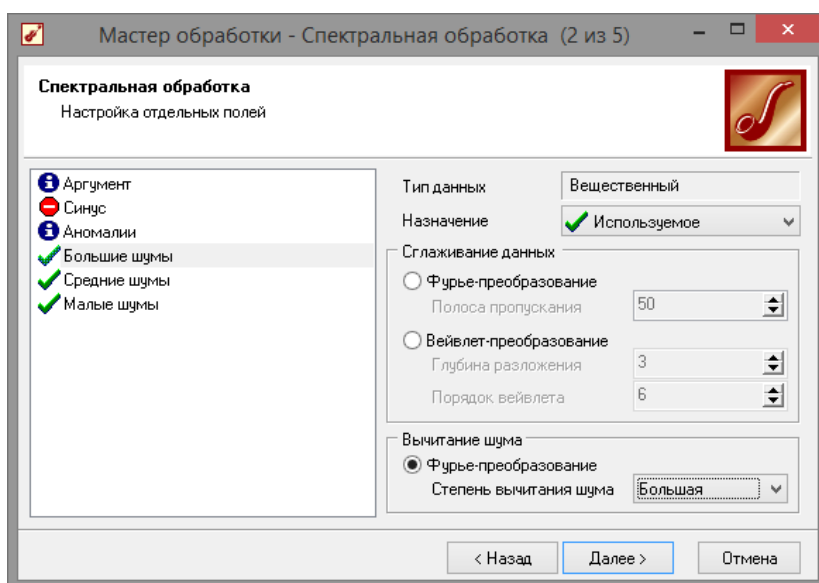


Рис. 1.26 - Настройки мастера спектральной обработки

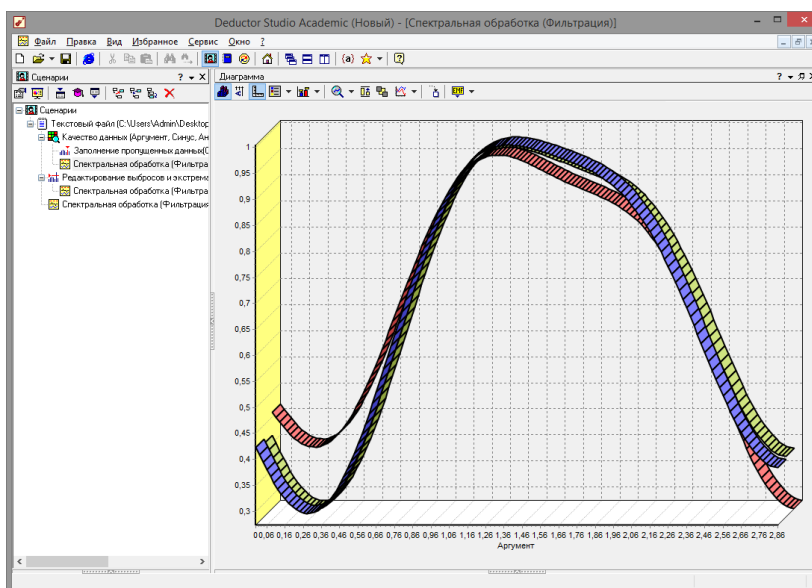


Рис 1.27 - Диаграмма после применения спектральной обработки



1.5 Возможности автоматического анализа качества импортируемых данных в *Deductor Academic*

В мастере обработки выбрать пункт «качество данных»(рис. 1.28).

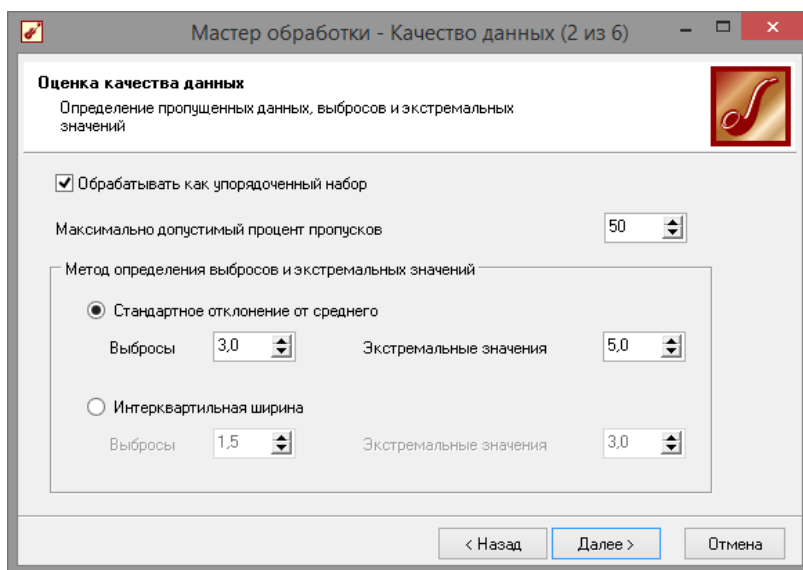


Рис. 1.28 - Мастер качества данных

После анализа мастер дает рекомендации к обработке данных и возможность автоматического исправления (рис. 1.29). Следует отметить, что автоматическое исправление далеко не всегда дает желаемые результаты (рис. 1.30).

№	Столбец	Тип данных	Вид данных	Пропуски		Выбросы		Экстремальные		Количество уникальн...	Качество данных	Рекоменд
				Колво	Действие	Колво	Действие	Колво	Действие			
1	Аргумент	Вещест...	Непер...			21	Интерполиров...	6	Сглаживае...		1.0000	Пропуски
2	Сигур	Вещест...	Непер...					5	Сглаживае...		0.8956	Приведба.
3	Анокали	Вещест...	Непер...					5	Сглаживае...		0.9120	Приведба.
4	Большие ...	Вещест...	Непер...					5	Сглаживае...		0.9408	Приведба.
5	Средние ш...	Вещест...	Непер...					7	Сглаживае...		0.9076	Приведба.
6	Малые шу...	Вещест...	Непер...					6	Сглаживае...		0.9206	Приведба.

Рис. 1.29 - Результаты мастера качества до обработки данных



№	Столбец	Тип данных	Вид данных	Пропуски		Выбросы		Экстремальные		Количество уникальных	Качество данных	Резюме
				Кол-во	Действие	Кол-во	Действие	Кол-во	Действие			
1	Аргумент	9.0 Вещест...	Непер...							1,0000		Пригоден
2	Синус	9.0 Вещест...	Непер...							0,8963		Пригоден
3	Аномалии	9.0 Вещест...	Непер...							0,9120		Пригоден
4	Большие ...	9.0 Вещест...	Непер...							0,9408		Пригоден
5	Средние ш ...	9.0 Вещест...	Непер...							0,9076		Пригоден
6	Малые ш ...	9.0 Вещест...	Непер...							0,9206		Пригоден

Рис. 1.30 - Вывод мастера после всех внесенных нами изменений

«Грязные данные» представляют собой очень большую проблему. Фактически они могут свести на нет все усилия по анализу данных. Причем, речь идет не о разовой операции, а о постоянной работе в этом направлении. Чисто не там, где не сорят, а там, где убирают.

Описанные выше варианты решения проблем не единственные. Есть еще достаточно много методов обработки, начиная от экспертных систем и заканчивая нейросетями. Главное, суметь грамотно ими воспользоваться. Обязательно нужно учитывать то, что методы очистки сильно привязаны к предметной области. От сферы деятельности организации и назначения хранилища данных зависит практически все. То, что для одних является шумом для других очень ценная информация. Если у нас будет априорная информация о задаче, то качество очистки данных можно увеличить на порядки.

1.6 Задание на самостоятельную работу

Сгенерировать собственный набор данных провести его анализ и выполнить предобработку данных. Данные сгенерировать в электронной таблице с помощью *MS Excel* с использованием формул и автозаполнений. Готовый файл необходимо сохранить в формате «*.csv» (*MS-DOS*) и при импорте в *Deductor*, выбрать в качестве разделителя «Точка с запятой».

Содержание отчёта

1. Описание предметной области
2. Зашумлённые данные (90-100 векторов размерностью не менее 4)



3. Скорректированные данные
4. Краткий порядок обработки
5. Выводы

Контрольные вопросы

1. Для чего служит программа *Deductor Academic*?
2. Зачем нужна предобработка данных?
3. Что такое парциальная предобработка?
4. Что такое вейвлет?
5. Какие данные можно импортировать в программу?

Библиографический список

1. Официальный сайт *Deductor Academic Studio* [Электронный ресурс] Режим доступа: <http://www.basegroup.ru/>
2. *Deductor 5.3* Руководство аналитика.
3. Встроенная справка приложения.

Раздел 2. Статистические методы

Практическая работа 2

Базовые методы интеллектуального анализа данных

Продолжительность работы – 8 час.

Цель работы: ознакомиться с возможностями классификации данных с помощью аналитического пакета *Deductor Academic*.

Программа работы

1. Выполнить классификацию данных с использованием алгоритма *g-mean*.
2. Выполнить классификацию данных с использованием алгоритма *k-mean*.
3. Выполнить классификацию данных с использованием нейронной сети.

Методические указания по выполнению работы

Задача разбиения множества объектов или наблюдений на априорно заданные группы, называемые классами, внутри каждой из которых они предполагаются похожими друг на друга, имеющими примерно одинаковые свойства и признаки. При этом решение получается на основе анализа значений атрибутов (признаков).

Классификация является одной из важнейших задач *Data Mining*. Она применяется в



маркетинге при оценке кредитоспособности заемщиков, определении лояльности клиентов, распознавании образов, медицинской диагностике и многих других приложениях. Если аналитику известны свойства объектов каждого класса, то когда новое наблюдение относится к определенному классу, данные свойства автоматически распространяются и на него.

Если число классов ограничено двумя, то имеет место бинарная классификация, к которой могут быть сведены многие более сложные задачи. Например, вместо определения таких степеней кредитного риска, как «Высокий», «Средний» или «Низкий», можно использовать всего две - «Выдать» или «Отказать».

Для классификации в *Data Mining* используется множество различных моделей: нейронные сети, деревья решений, машины опорных векторов, метод *k*-ближайших соседей, алгоритмы покрытия др., при построении которых применяется обучение с учителем, когда выходная переменная (метка класса) задана для каждого наблюдения. Формально классификация производится на основе разбиения пространства признаков на области, в пределах каждой из которых многомерные векторы рассматриваются как идентичные. Иными словами, если объект попал в область пространства, ассоциированную с определенным классом, он к нему и относится.

Рассмотрим классификацию данных на примере Ирисов Фишера. Ирисы Фишера - это набор данных для задачи классификации, на примере которого Рональд Фишер в 1936 году продемонстрировал работу разработанного им метода дискриминантного анализа. Этот набор данных стал уже классическим, и часто используется в литературе для иллюстрации работы различных статистических алгоритмов. Ирисы Фишера состоят из данных о 150 экземплярах ириса, по 50 экземпляров из трёх видов - Ирис щетинистый (*Iris setosa*), Ирис виргинский (*Iris virginica*) и Ирис разноцветный (*Iris versicolor*). Для каждого экземпляра измерялись четырехарактеристики (в сантиметрах):

- длина чашелистика (англ. *sepal length*);
- ширина чашелистика (англ. *sepal width*);
- длина лепестка (англ. *petal length*);
- ширина лепестка (англ. *petal width*).

На основании этого набора данных требуется построить правило классификации, определяющее вид растения по данным измерений. Это задача многоклассовой классификации, так как имеется три класса - три вида ириса.

2.1 Классификация данных с использованием алгоритма «*g-mean*»

Импортируем в программу данные из файла «Ирисы.txt». Для начала попробуем провести классификацию ирисов, встроенным методом кластеризации *g-mean* (рис. 2.1). Интересно то, что обучение будет проходить без учителя, т.е. выходные данные не будут указаны.



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 28 из 74

Первый экземпляр _____

КОПИЯ № _____

Длина чашелистика	Ширина чашелистика	Длина лепестка	Ширина лепестка	Класс
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
5	3.6	1.4	0.2	Iris-setosa
5.4	3.9	1.7	0.4	Iris-setosa
4.6	3.4	1.4	0.3	Iris-setosa
5	3.4	1.5	0.2	Iris-setosa
4.4	2.9	1.4	0.2	Iris-setosa

Рис. 2.1 - Импортированные данные

Более того, не будет даже указано количество кластеров, на которое необходимо разделить данную выборку, проверим эффективность встроенной системы кластеризации, для этого необходимо выбрать в мастере обработок пункт кластеризация(рис. 2.2).

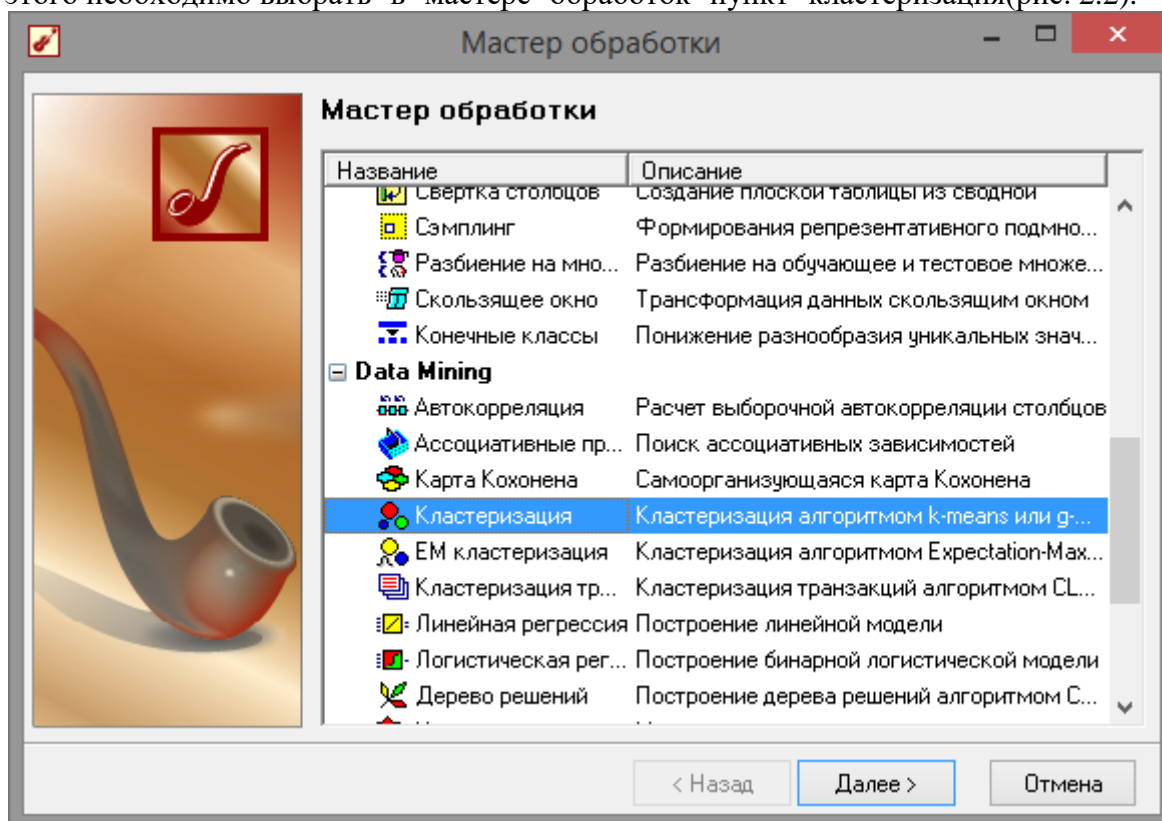


Рис. 2.2 - Мастер кластеризации

На следующем шаге (рис. 2.3) со столбца «Класс» уберем назначение выходное и поставим информационное, теперь обучение будет происходить без учителя.



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 29 из 74

Первый экземпляр _____

КОПИЯ № _____

Рис. 2.3 - Настройка мастера кластеризации

Далее следует настроить параметры обучения, в данном конкретном случае, параметры по умолчанию отлично подходят (рис. 2.4).

Множество	Размер		Порядок сортировки
	В процентах	В строках	
<input checked="" type="checkbox"/> Обучающее	95,00	142	По возрастанию
<input checked="" type="checkbox"/> Тестовое	5,00	8	По возрастанию
ИТОГО:	100,00	150	

Рис. 2.4 - Настройка параметров обучения



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 30 из 74

Первый экземпляр _____

КОПИЯ № _____

На следующем шаге мастера необходимо выбрать алгоритм кластеризации. На и так заранее известно, что количество кластеров должно равняться «3», но протестируем возможности программы и предоставим ей самой выбрать количество кластеров. Это так же будет полезно, если не известно на какое количество групп следует разбивать выборку (рис. 2.5).

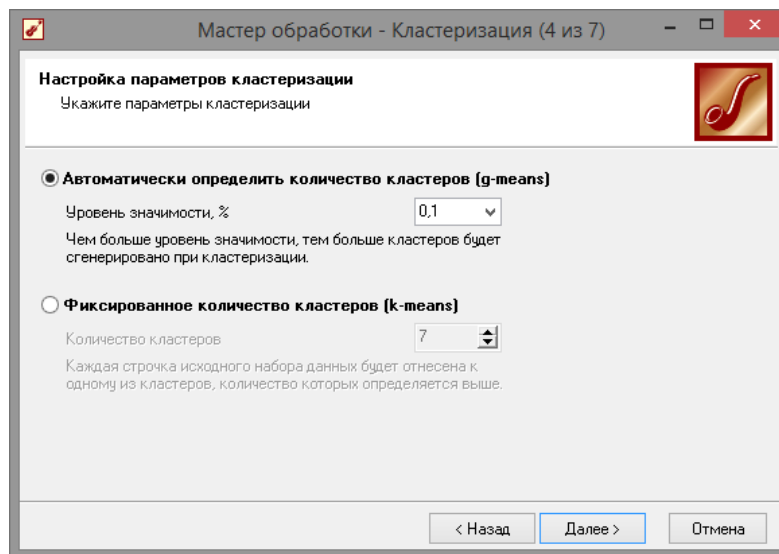


Рис. 2.5 - Выбор алгоритма кластеризации

Следующий шаг мастера предлагает запустить процесс обучения и наблюдать в процессе обучения величину ошибки, а также процент распознанных примеров. Параметр «Частота обновления» отвечает за то, через какое количество эпох обучения выводится данная информация (рис. 2.6).

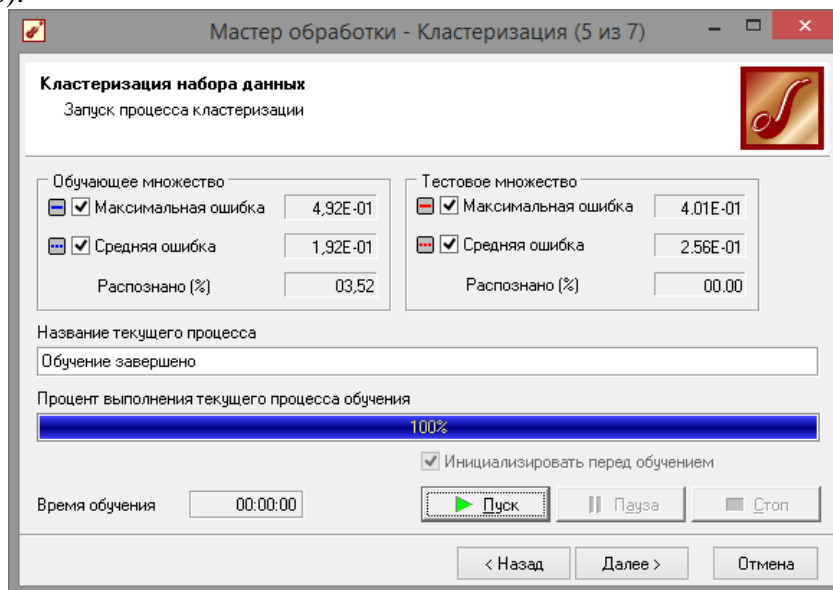



Рис. 2.6 - Обучение сети

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ») Костанайский филиал Кафедра социально-гуманитарных и естественнонаучных дисциплин			
Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент			
Версия документа - 1	стр. 31 из 74	Первый экземпляр _____	КОПИЯ № _____

После обучения сети, в качестве визуализаторов выберем: «Связи кластеров»; «Профили кластеров»; «Матрицу сравнения»; «Что-если»(рис. 2.7).

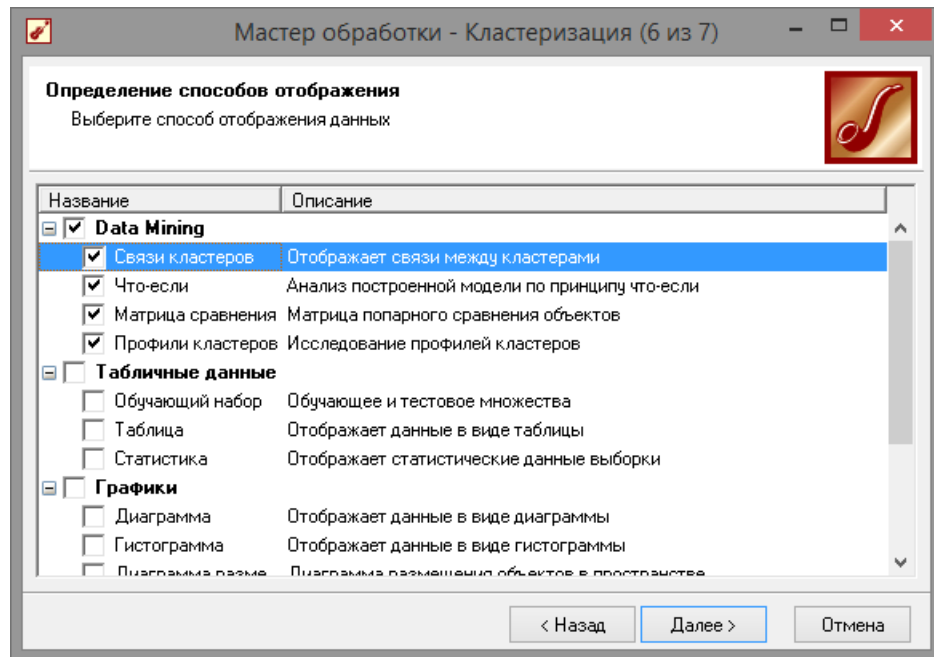


Рис. 2.7 - Визуализация данных

Как видно из рис. 2.8 приложение верно переделало количество групп, также здесь можно увидеть на сколько значим тот или иной параметр, для присвоения цветку того или иного вида.

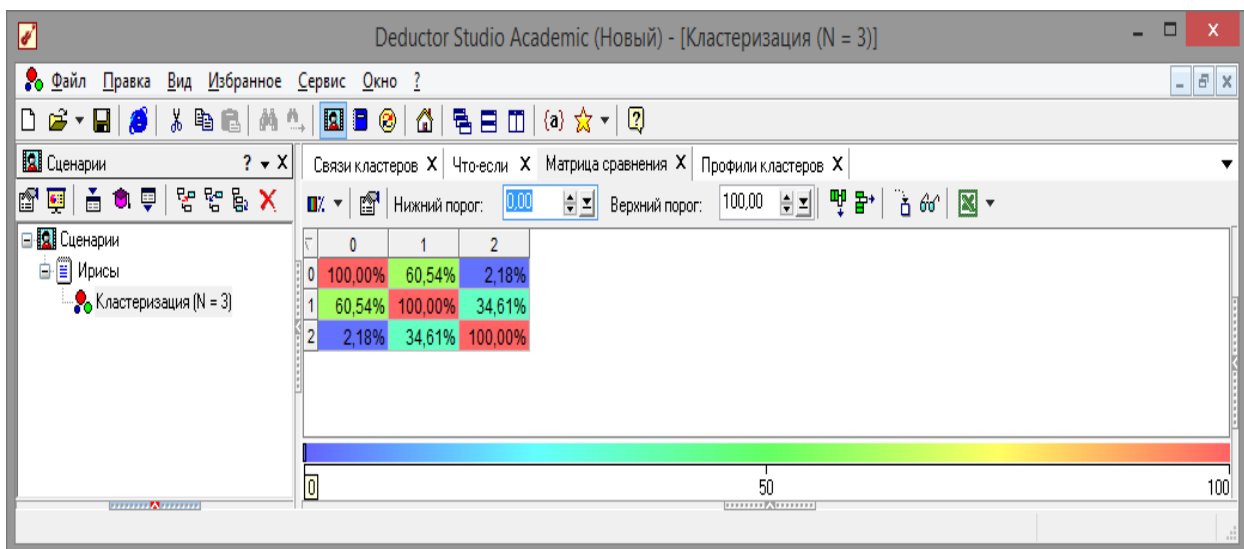


Рис. 2.8 - Профили кластеров



На матрице сравнения (рис. 2.9) видно что больше всех от остальных отличается «кластер 2» (ему соответствует «iris-setosa»).

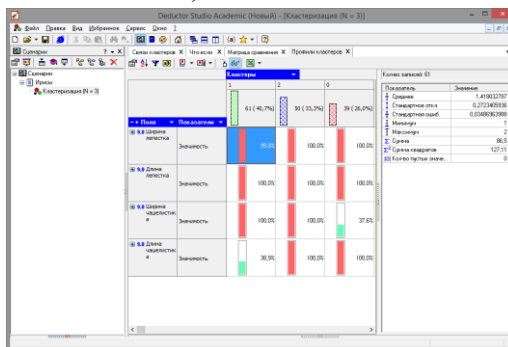


Рис. 2.9 - Матрица сравнения

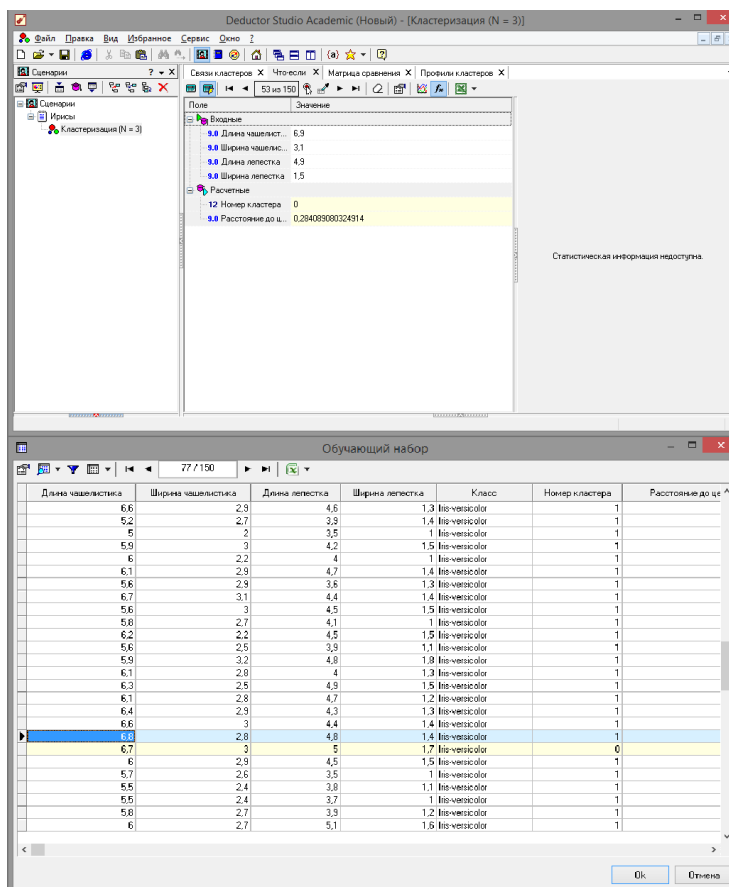


Рис. 2.10 - Инструмент «Что-Если»

Визуализатор «Что-если» (рис. 2.10) позволит провести эксперимент, введя любые



значения параметров. Если же нажать кнопку «Загрузить данные» из исходной выборки, то можно заметить неточности определения кластеров из-за схожих параметров цветков.

2.2 Классификация данных с помощью нейронной сети

Теперь попробуем провести классификацию ирисов при помощи обычной нейронной сети. Входные данные брать из файла «Ирисы.txt», только в данном случае обучение будет происходить с учителем. Импортируем данные и в мастере обработок выберем пункт нейросеть (рис. 2.11).

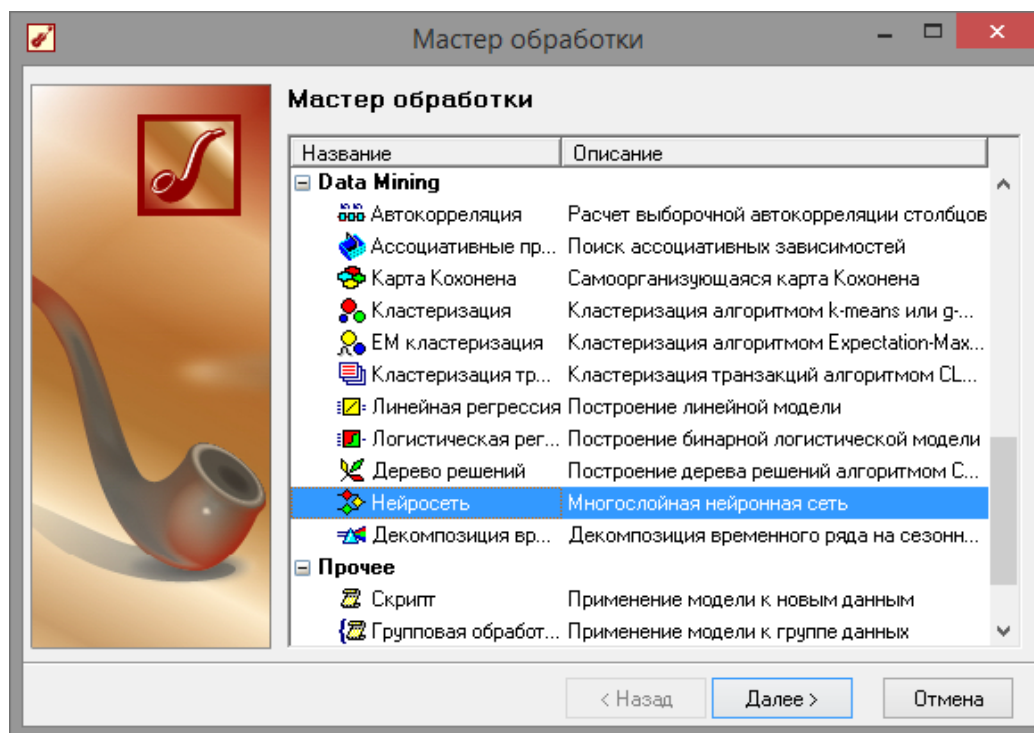


Рис. 2.11 - Мастер обработок

Установим в качестве выходного значения параметр класс. В качестве нормализатора для выхода «Класс» необходимо установить пункт «Уникальные значения».

Примечание. Уникальные значения используются для дискретных значений. Такими являются строки, числа или даты, заданные дискретно. Чтобы привести непрерывные числа в дискретные, можно, например, воспользоваться обработкой «Квантование». Так следует поступать для величин, для которых можно задать отношение порядка, то есть, если для двух любых дискретных значений можно указать, какое больше, а какое меньше. Тогда все значения необходимо расположить в порядке возрастания (рис. 2.12-2.13). Далее они нумеруются по порядку, и значения заменяются их порядковым номером



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 34 из 74

Первый экземпляр _____

КОПИЯ № _____

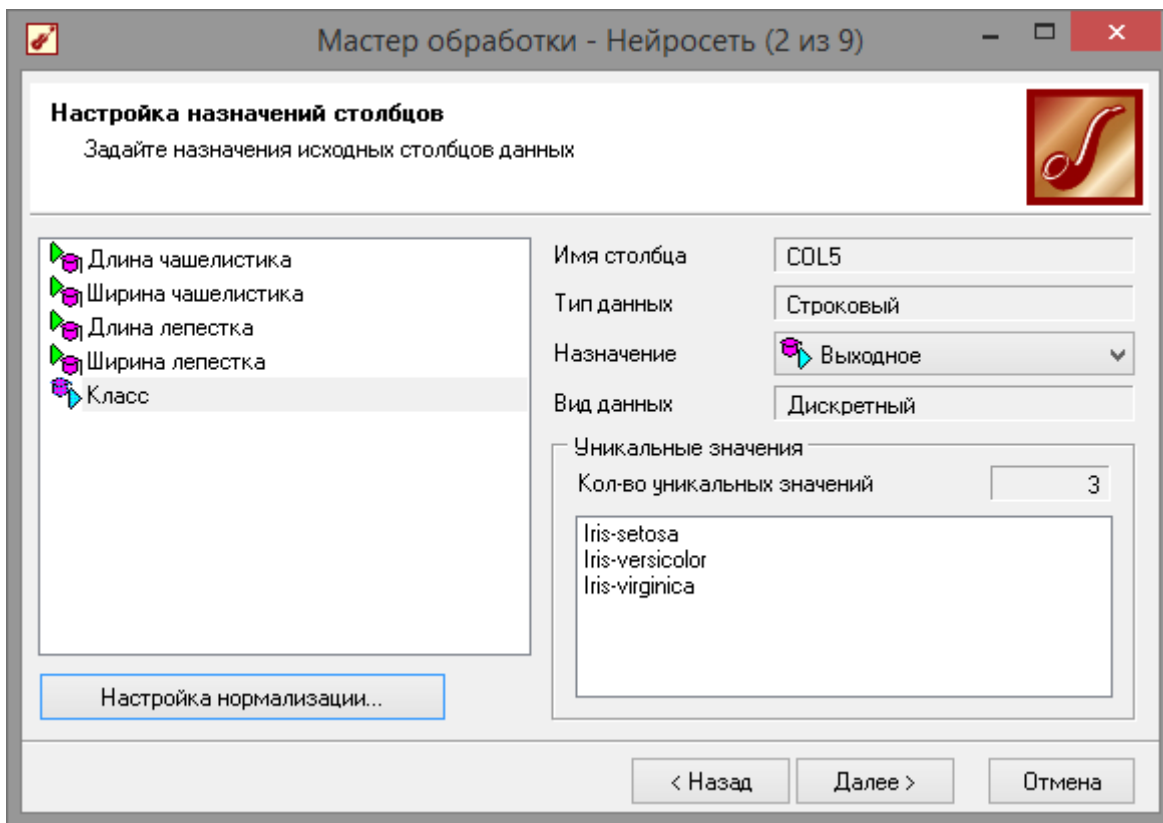


Рис. 2.12 - Назначение входов и выходов

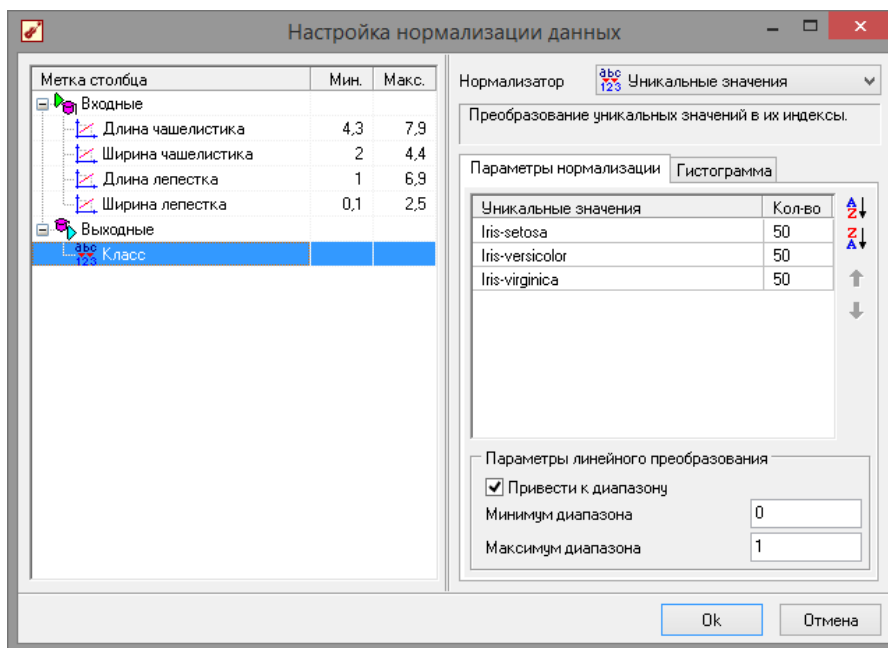


Рис. 2.13 - Настройка нормализации



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 35 из 74

Первый экземпляр _____

КОПИЯ № _____

Следующим шагом указываем параметры обучения (рис. 2.14).

Разбиение исходного набора данных на подмножества

Настройте разбиение исходного множества данных на обучающее и тестовое множества

Способ разделения исходного множества данных: Случайно

Столбец для разделения исходного множества: _____

Множество	Размер		Порядок сортировки
	В процентах	В строках	
<input checked="" type="checkbox"/> Обучающее	95,00	142	По возрастанию
<input checked="" type="checkbox"/> Тестовое	5,00	8	По возрастанию
ИТОГО:	100,00	150	

Количество строк (всего): 150

< Назад Далее > Отмена

Рис. 2.14 - Параметры обучения

Количество нейронов первого слоя экспериментально было выставлено в значение 5 (рис. 2.15).

Структура нейронной сети

Нейроны в слоях:

входном: 4

скрытых слоев: 1

выходном: 1

Слой	Нейроны
1	5

Активационная функция:


Тип функции: Сигмоида

Круглизна: 1,000

Сигмоида

< Назад Далее > Отмена

Рис. 2.15 - Параметры обучения

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ») Костанайский филиал Кафедра социально-гуманитарных и естественнонаучных дисциплин			
Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент			
Версия документа - 1	стр. 36 из 74	Первый экземпляр _____	КОПИЯ № _____

Далее выбираем алгоритм обучения (рис. 2.16).

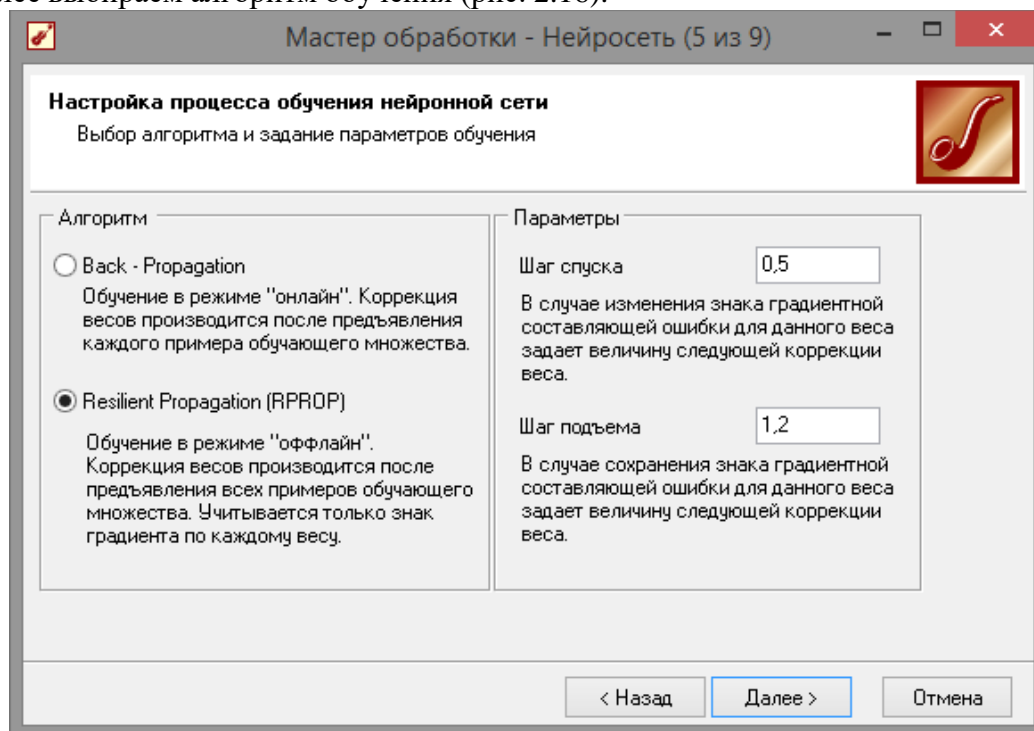


Рис. 2.16 - Алгоритм обучения

На следующем шаге (рис. 2.17) устанавливаем значение погрешности и количество эпох обучения. И запускаем обучение сети(рис. 2.18).

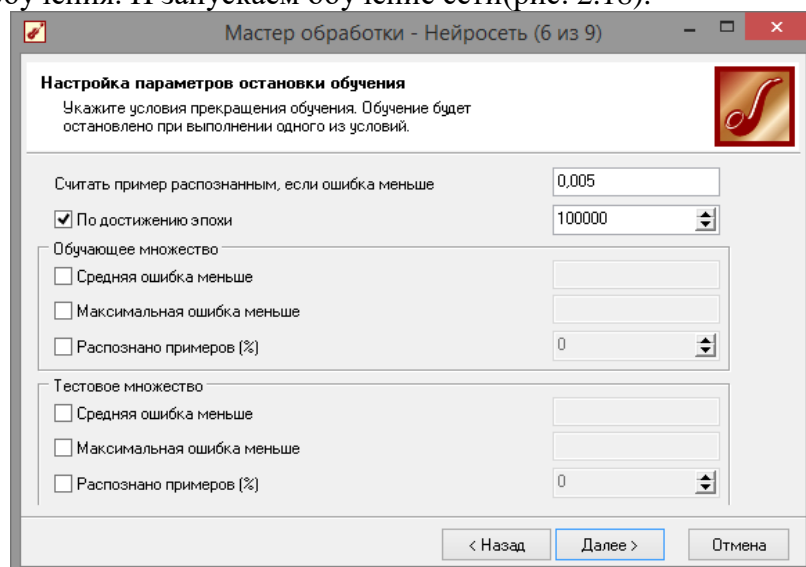


Рис. 2.17 - Установка погрешности и числа эпох обучения

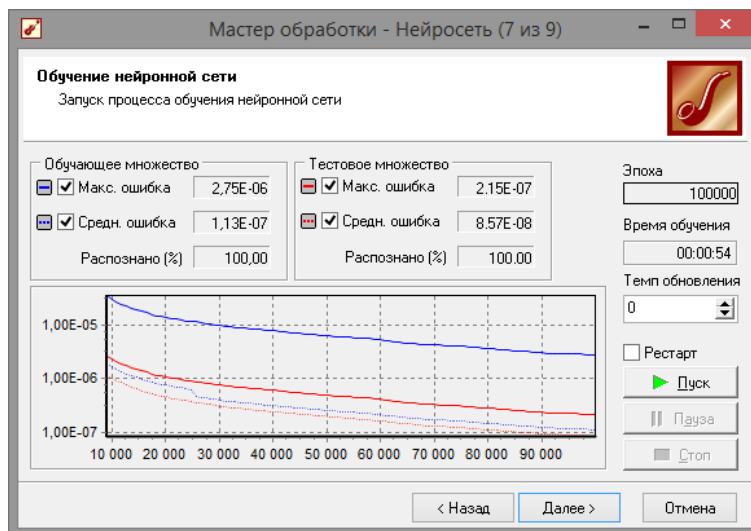


Рис. 2.18 - Обучение сети

Как видно из рис.2.18, сеть обучалась дольше, но дала более точные результаты. В качестве визуализаторов выбираем «Граф нейросети», далее пункт «Таблица сопряжения» и «Что-Если» (рис. 2.19).

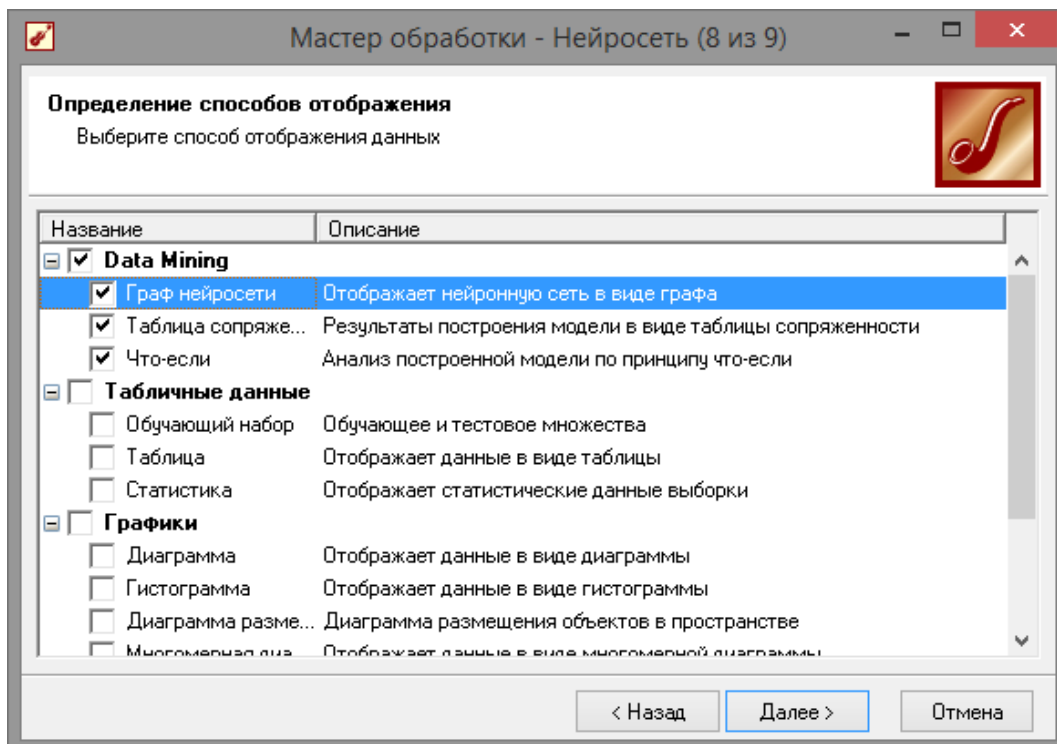


Рис. 2.19 - Визуализаторы



На графе нейросети видно, как выглядит обученная сеть (рис. 2.20).

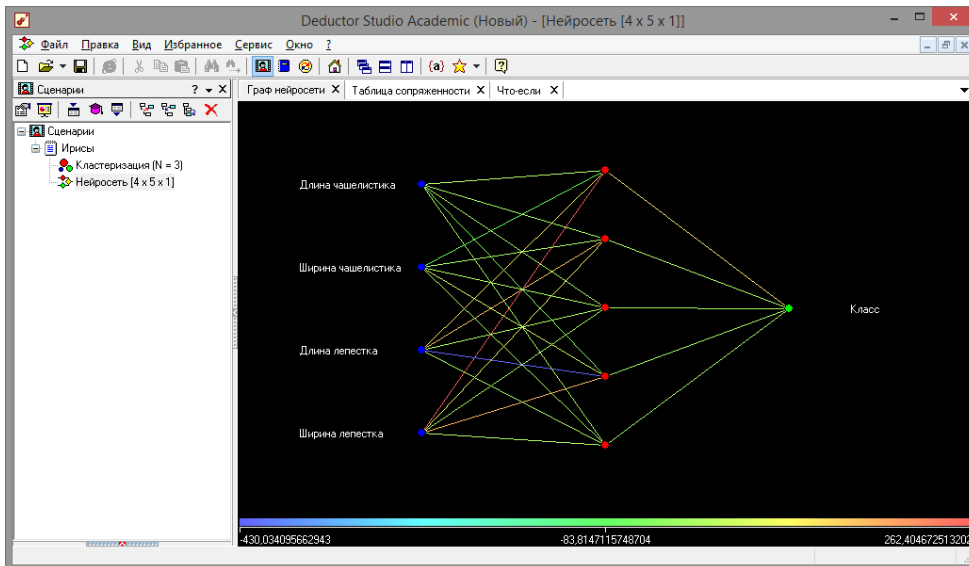


Рис. 2.20 - Граф нейросети

Диаграмма сопряженности (рис. 2.21) показывает как распределились входные значения. На ней видно что сеть обучилась абсолютно точно.

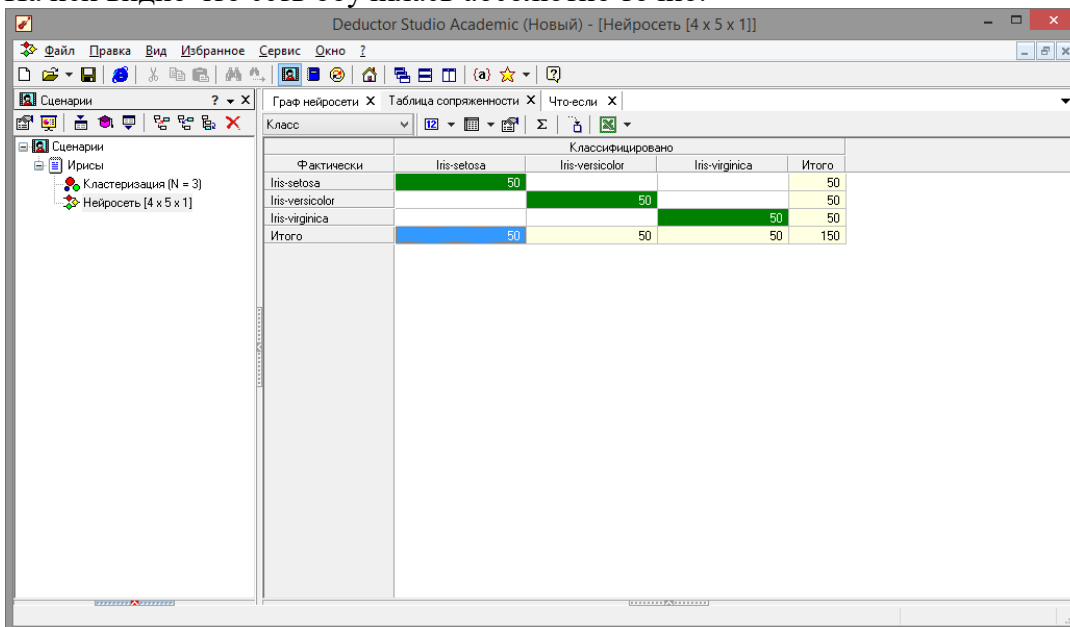


Рис. 2.21 - Диаграмма сопряженности

Убедимся в высказанном выше заключении взглянув на таблицу «Что-Если» (рис. 2.22). Даже при вводе значений отсутствующих в выборке, сеть верно



реагирует на них. Далее можно опять нажать на кнопку «Загрузить данные из исходной выборки» чтобы убедиться в правильности распознавания результатов.

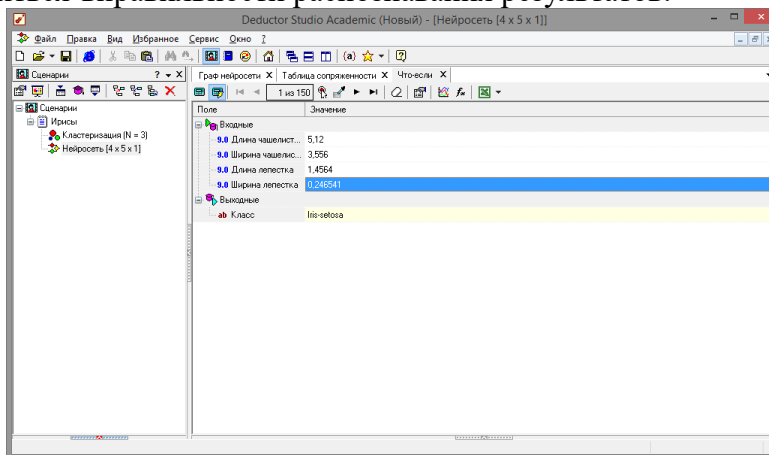


Рис. 2.22 - Инструмент «Что-Если»

Данный пример показал, как можно обучить сеть с учителем. Был изучен новый

Длина чашелистика	Ширина чашелистика	Длина лепестка	Ширина лепестка	Класс	Класс_OUT
5,5	3,5	1,3	0,2	Iris-setosa	Iris-setosa
4,9	3,1	1,5	0,1	Iris-setosa	Iris-setosa
4,4	3	1,3	0,2	Iris-setosa	Iris-setosa
5,1	3,4	1,5	0,2	Iris-setosa	Iris-setosa
5	3,5	1,3	0,3	Iris-setosa	Iris-setosa
4,5	2,3	1,3	0,3	Iris-setosa	Iris-setosa
4,4	3,2	1,3	0,2	Iris-setosa	Iris-setosa
5	3,5	1,6	0,6	Iris-setosa	Iris-setosa
5,1	3,8	1,9	0,4	Iris-setosa	Iris-setosa
4,8	3	1,4	0,3	Iris-setosa	Iris-setosa
5,1	3,8	1,6	0,2	Iris-setosa	Iris-setosa
4,6	3,2	1,4	0,2	Iris-setosa	Iris-setosa
5,3	3,7	1,5	0,2	Iris-setosa	Iris-setosa
5	3,3	1,4	0,2	Iris-setosa	Iris-setosa
7	3,2	4,7	1,4	Iris-versicolor	Iris-versicolor
6,4	3,2	4,5	1,5	Iris-versicolor	Iris-versicolor
6,9	3,1	4,9	1,5	Iris-versicolor	Iris-versicolor
5,5	2,3	4	1,3	Iris-versicolor	Iris-versicolor
6,5	2,8	4,6	1,5	Iris-versicolor	Iris-versicolor
5,7	2,8	4,5	1,3	Iris-versicolor	Iris-versicolor
6,3	3,3	4,7	1,6	Iris-versicolor	Iris-versicolor
4,9	2,4	3,3	1	Iris-versicolor	Iris-versicolor
6,6	2,9	4,6	1,3	Iris-versicolor	Iris-versicolor
5,2	2,7	3,9	1,4	Iris-versicolor	Iris-versicolor
5	2	3,5	1	Iris-versicolor	Iris-versicolor
5,9	3	4,2	1,5	Iris-versicolor	Iris-versicolor
6	2,2	4	1	Iris-versicolor	Iris-versicolor
6,1	2,9	4,7	1,4	Iris-versicolor	Iris-versicolor

инструмент «Нейросеть» и сделано предположение, что обучение с учителем более эффективно, хотя и занимает большой промежуток времени.

Пример показал простоту и удобство применения «Нейросеть» для классификации Ирисов Фишера. Мастер предлагает широкие возможности по настройке процесса обучения. После обучения сети стали видны ее достоинства для анализа. Также были продемонстрированы широкие возможности визуализации построенной сети. Все это говорит о эффективности инструмента «Нейросеть».



2.3 Задания

1. Сгенерировать данные для классификации, провести классификацию двумя способами (инструмент «Кластеризация» и «Нейронная сеть»), сделать выводы по эффективности этих двух способов.

2. Провести эксперимент по изменению параметров обучения нейросети, и сделать выводы по эффективности процесса обучения при разном количестве нейронов.

3.

Контрольные вопросы

1. Для чего служат алгоритмы g-mean и k-mean?

2. Какие алгоритмы обучения нейронной сети предлагает программный комплекс Deductor Academic?

3. В чем их отличие?

4. Что такое обучение с учителем?

5. Что такое обучение без учителя?

Библиографический список

1. Официальный сайт *Deductor Academic Studio* [Электронный ресурс] Режим доступа: <http://www.basegroup.ru/>

2. *Deductor 5.3* Руководство аналитика.

3. Встроенная справка приложения.

Раздел 3. Когнитивные технологии

Практическая работа 3

Применение интеллектуального анализа данных в задачах поддержки принятия решений

Продолжительность работы – 8 час.

Цель работы: освоить методы и средства прогнозирования в пакете *Deductor Academic* при интеллектуальном анализе данных в задачах поддержки принятия решений.

Программа работы

1. Выполнить пример прогнозирования с помощью нейронных сетей в пакете *Deductor Academic*.

2. Выполнить пример прогнозирования с помощью временных рядов в пакете *Deductor Academic*.

3. Выполнить прогнозирование с помощью нейронных сетей и временных рядов



на данных согласно индивидуальному заданию.

Методические указания по выполнению работы

Основное направление программы *Deductor Studio* – анализ, прогнозирование, классификация и кластеризация данных. Программа предоставляет следующие механизмы анализа: нейронные сети, линейный регрессионный анализ, построение деревьев решений, самоорганизующиеся карты Кохонена, прогнозирование временного ряда, обнаружение дубликатов и противоречий. Нейросети – механизм, который используют для прогнозирования и решения задач классификации. Они применяются в основном там, где существует нелинейные зависимости результата от входных факторов.

3.1 Прогнозирование умножения с помощью нейронных сетей

Рассмотрим прогнозирование с помощью нейронных сетей на примере прогнозирования результата умножения двух чисел – файл «Произведение.txt». В нем содержится таблица со следующими полями: «АРГУМЕНТ1», «АРГУМЕНТ2» – множители,

«ПРОИЗВЕДЕНИЕ» – их произведение. Причем произведение некоторых чисел пропущено в обучающей выборке, например, $7 \times 7 = 49$. Импортировав данные из файла, можно посмотреть результат умножения, используя таблицу.

Пусть необходимо построить модель прогноза умножения, подавая на вход которой два множителя получать на выходе их произведение. Для этого необходимо, находясь на узле импорта, открыть мастер обработки, показанный на рис. 3.1. В нем выбрать в качестве обработки нейронную сеть и перейти к следующему шагу мастера. На втором шаге мастера необходимо установить назначение полей «АРГУМЕНТ1» и «АРГУМЕНТ2» как входные, а поле «ПРОИЗВЕДЕНИЕ» – как выходное и задать тип данных как «Целый» (рис. 3.2).

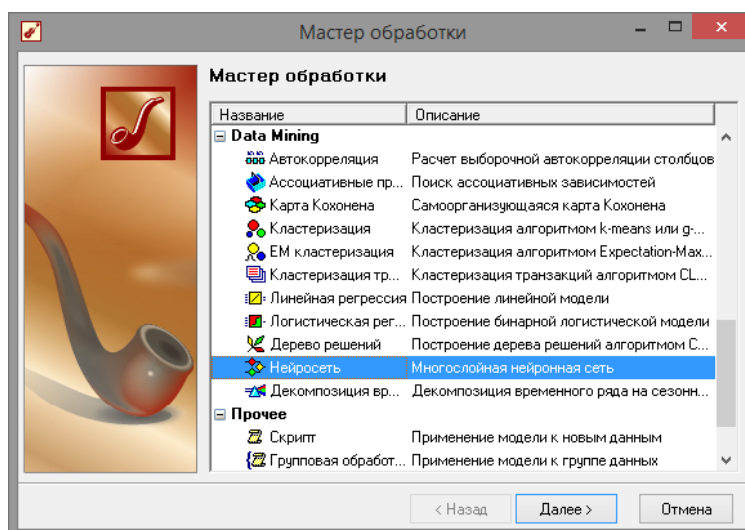


Рис. 3.1 - Мастер обработки



На следующем шаге предлагается настроить разбиение исходного множества данных на обучающее и тестовое. Здесь необходимо указать способ разбиения исходного множества данных «Случайно» и поставить обучающее множество размером 100 %, так как выборка слишком мала (рис. 3.3).

Далее необходимо указать функцию активации, количество скрытых слоев, и количество нейронов в слое. Данные необходимо подбирать экспериментально, так как большое количество нейронов и слоев, может сильно замедлить процесс обучения. Эмпирически были выставлены настройки, показанные на рис. 3.4.

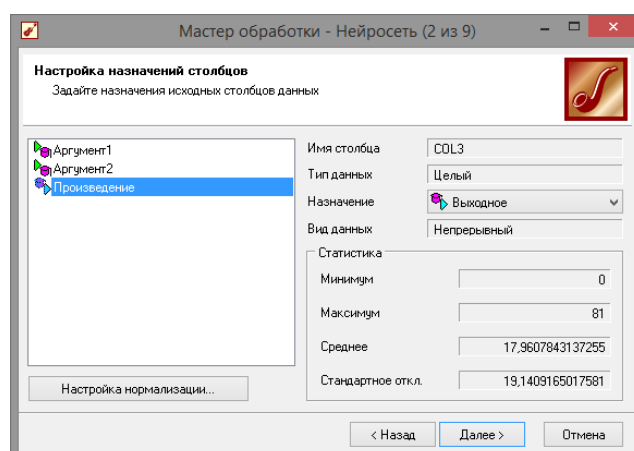


Рис. 3.2 - Мастер нейросети

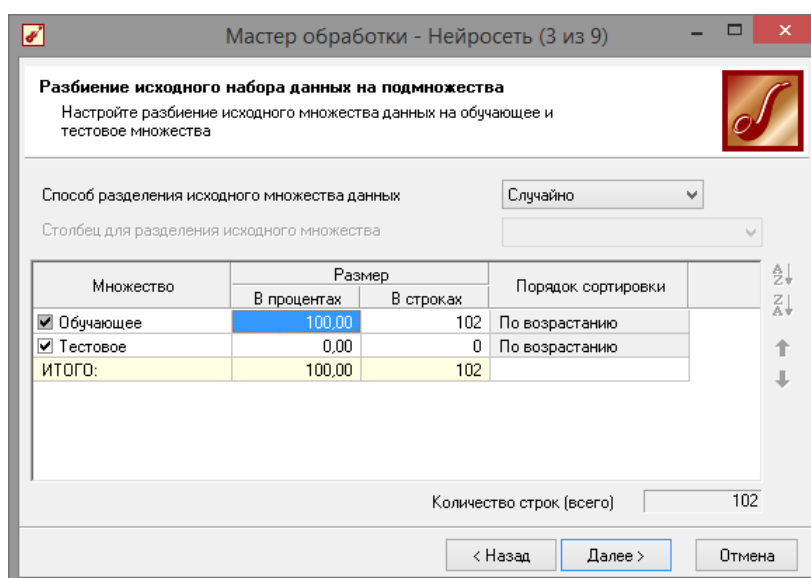


Рис. 3.3 - Настройки параметров обучения



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 43 из 74

Первый экземпляр _____

КОПИЯ № _____

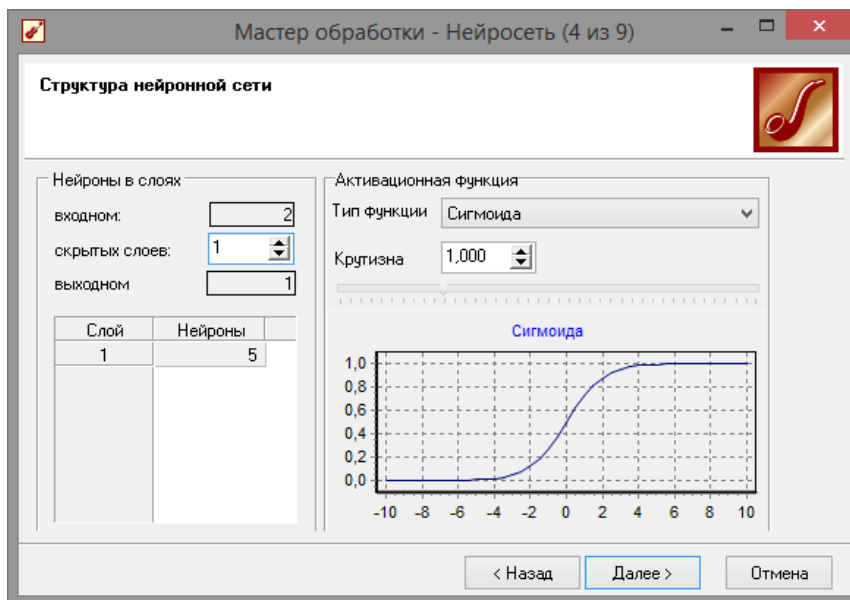


Рис. 3.4 - Структура нейронной сети

Следующий шаг предлагает выбрать алгоритм обучения и его параметры (рис. 3.5).

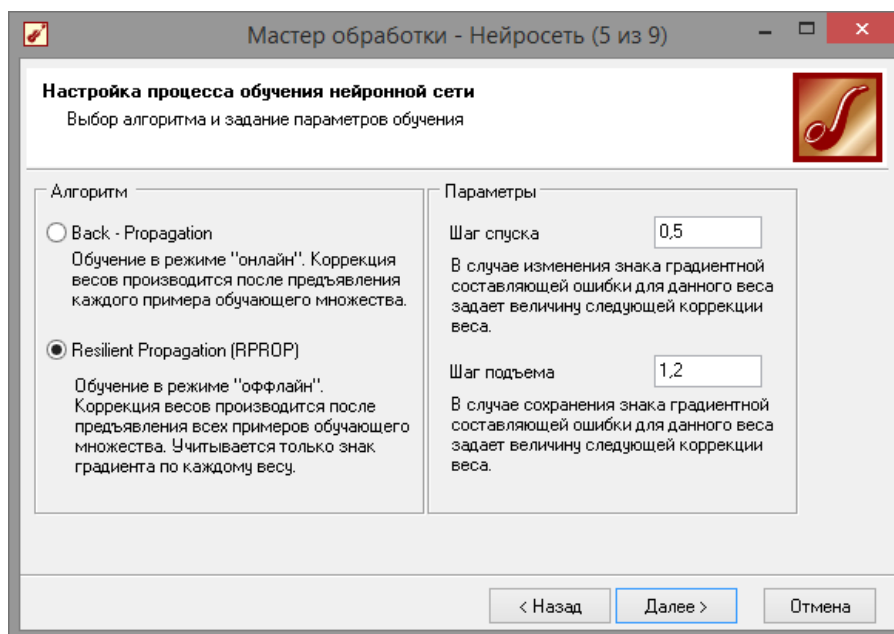


Рис. 3.5 - Алгоритм обучения

Далее настроим условия остановки обучения (рис. 3.6). Пусть пример считаем распознанным, если ошибка меньше 0,005. Также укажем условие остановки обучения при достижении эпохи 100000.



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 44 из 74

Первый экземпляр _____

КОПИЯ № _____

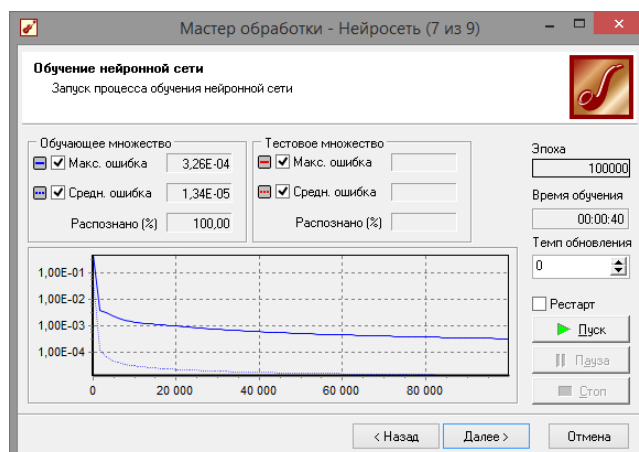


Рис. 3.6 - Настройка параметров остановки обучения

Следующий шаг мастера (рис. 3.7) предлагает запустить процесс обучения и наблюдать в процессе обучения величину ошибки, а также процент распознанных примеров. Параметр «Частота обновления» отвечает за то, через какое количество эпох обучения выводится данная информация.

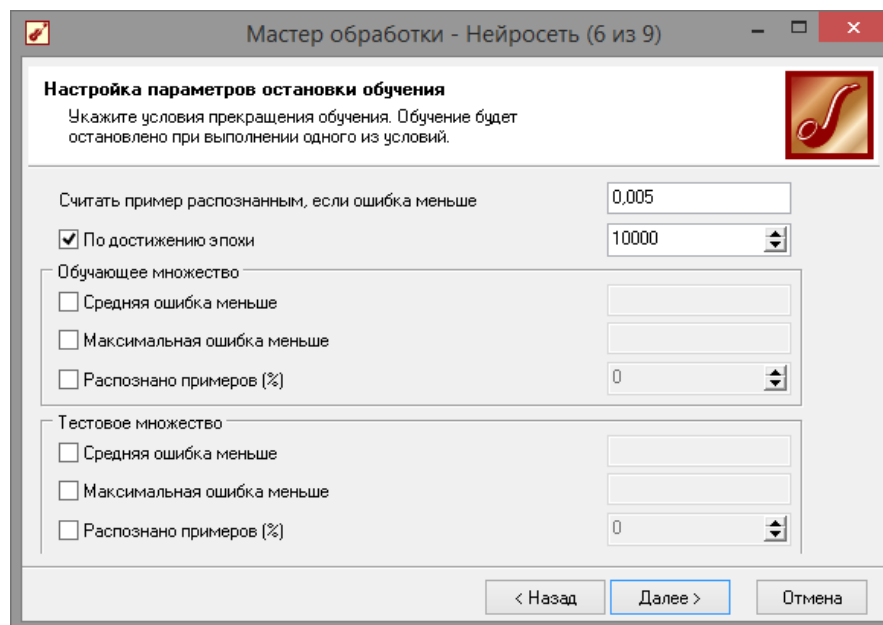



Рис. 3.7 - Обучение сети

 МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ») Костанайский филиал Кафедра социально-гуманитарных и естественнонаучных дисциплин			
Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент			
Версия документа - 1	стр. 45 из 74	Первый экземпляр _____	КОПИЯ № _____

После обучения сети, в качестве визуализаторов выберем варианты, показанные на рис. 3.8: «Диаграмма», «Диаграммарассеяния», «Граф нейросети», «Что-если».

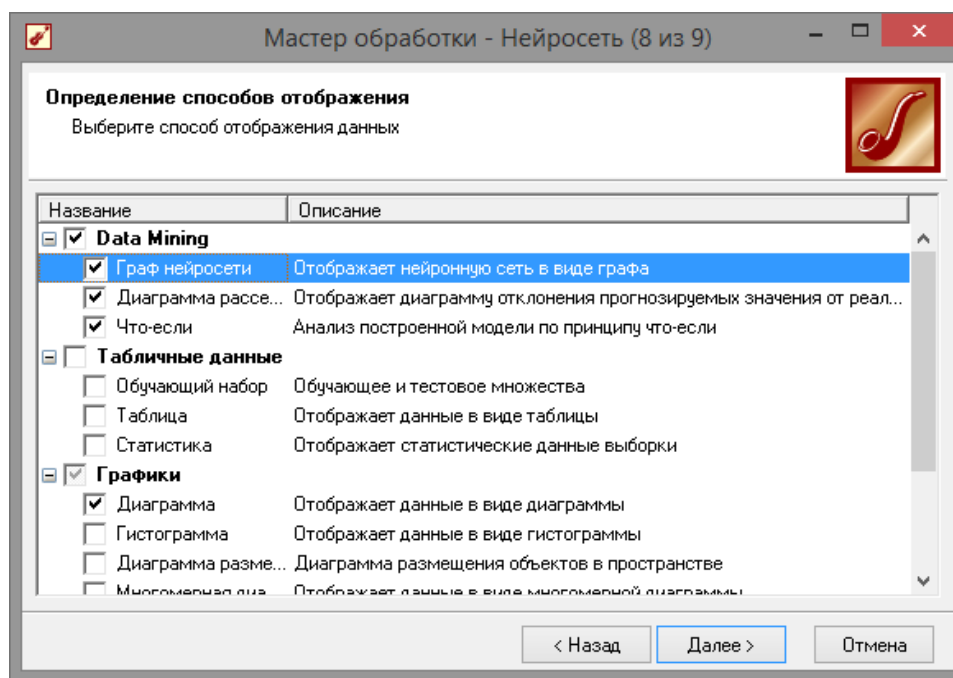


Рис. 3.8 - Визуализация данных

Результаты наглядно видны на диаграмме рассеяния (рис. 3.9), которая показывает рассеяние прогнозируемых данных относительно эталонных. По диаграмме можно судить, что сеть обучена недостаточно хорошо, из-за малого объема обучающей выборки. Также можно сравнить эталонные данные с прогнозируемыми, выбрав на обычной диаграмме два поля – «ПРОИЗВЕДЕНИЕ» и «ПРОИЗВЕДЕНИЕ_OUT». Если масштабировать диаграмму и включить отображение меток, то можно увидеть достаточно большую ошибку (рис. 3.10), но на цели этого задания она сказываться не будет.

Визуализатор «Что-если» позволит провести эксперимент, введя любые значения множителей АРГУМЕНТ1 и АРГУМЕНТ2 и рассчитав результат их произведения. Попробуем ввести аргументы которые отсутствуют в обучающей выборке, например, 7x7. Как видно на рис. 3.11 сеть обучена достаточно точно, так как получен верный результат, несмотря на то, что сеть никогда не видела такую комбинацию аргументов.



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 46 из 74

Первый экземпляр _____

КОПИЯ № _____

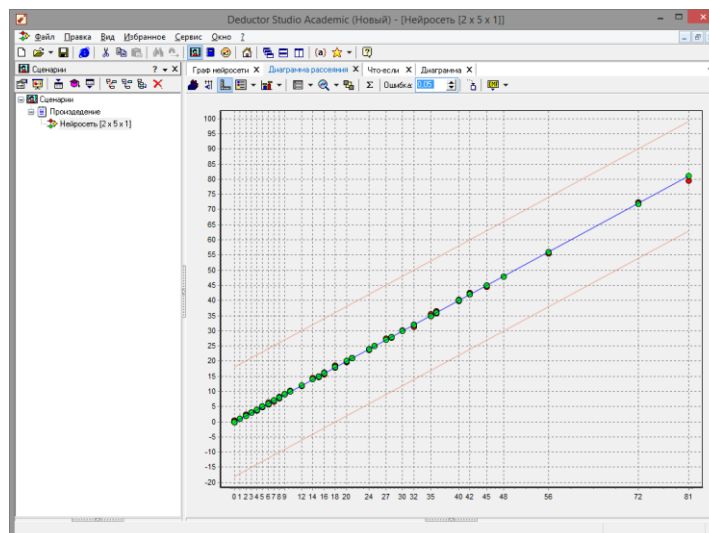


Рис. 3.9 - Диаграмма рассеяния

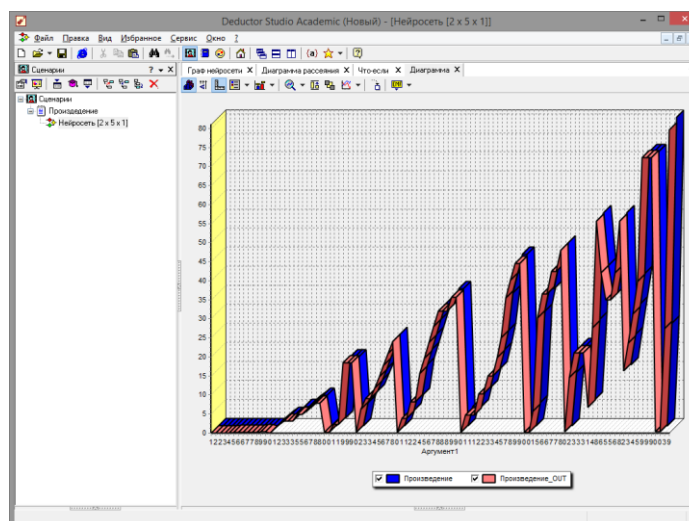
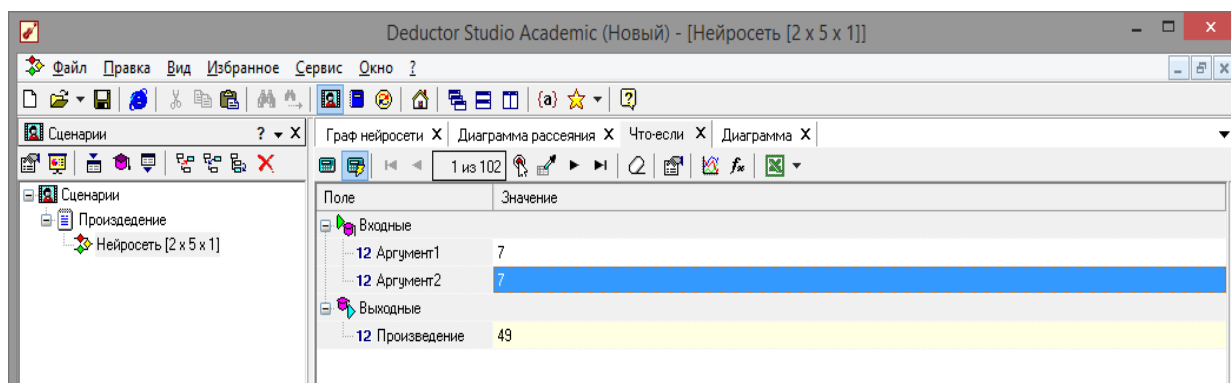
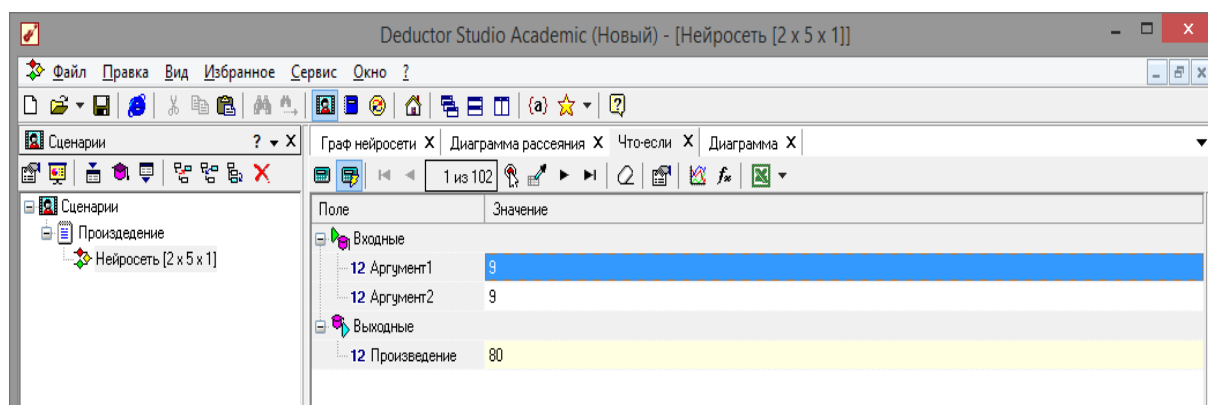


Рис. 3.10 - Сравнение эталонных данных с прогнозируемыми

**Рис. 3.11 - Инструмент «Что-Если»**

Далее введем в окно «Что-Если» аргументы, которые как следует из диаграммы и диаграммы рассеивания большую ошибку. Как видим по произведению, данные диаграммы верны относительно ошибки.

**Рис. 3.12 - Ошибки прогнозирования**

Вид построенной сети можно посмотреть, выбрав визуализатор «Граф нейронной сети» (рис. 3.13).

Данный пример показал, как можно построить модель прогноза, используя нейронную сеть. Пример показал, что для построения нет необходимости в строгой математической спецификации модели, что особенно ценно при анализе плохо формализуемых процессов. А большинство бизнес задач плохо формализуется. Это означает, что наличие достаточно развитых и удобных инструментальных программных средств позволяет аналитику при построении модели прогнозируемого процесса руководствоваться такими понятиями, как опыт и интуиция.

Настройки мастера позволяют увидеть широкие возможности *Deductor Studio* касательно структуры сети, способов обучения и т.д. Аналитику предоставляется широкие возможности по настройке



нормализации столбцов, разбиения данных на обучающее и тестовое множество, определения структуры сети, количества слоев и нейронов в каждом слое, выборе функции активации и ее параметров, выборе различных алгоритмов обучения и настройки их параметров.

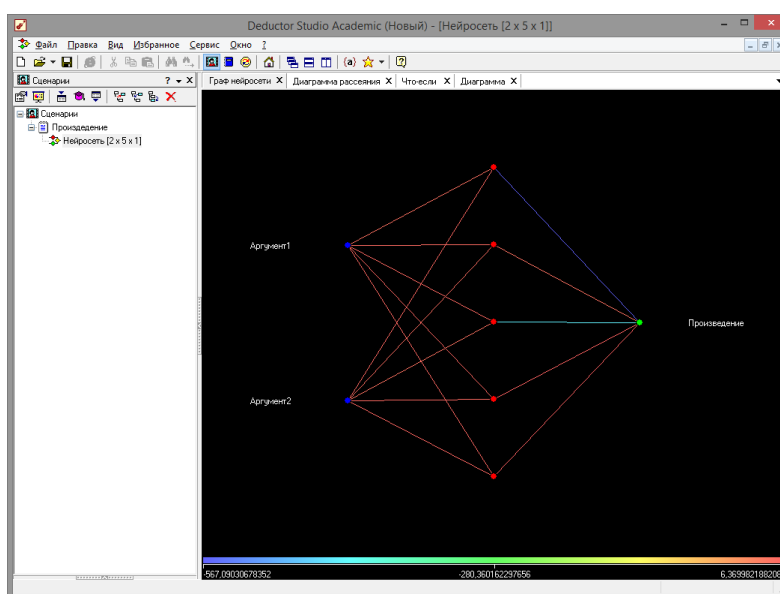


Рис. 3.13 - Граф нейронной сети

Все это позволяет построить модель, описывающую практически любые закономерности. Также было показано, как можно спрогнозировать результат, введя любые значения входных факторов, используя визуализатор «Что-если». Качество подготовки данных для модели, а также качество самой модели аналитик может оценить разными способами: посмотреть диаграмму рассеяния, провести ряд экспериментов при помощи «Что-если», построить гистограмму распределения ошибки и т.п.

3.2 Прогнозирование данных на основе временного ряда

Прогнозирование результата на определенное время вперед, основываясь на данных за прошедшее время – задача, встречающаяся довольно часто (к примеру, перед большинством торговых фирм стоит задача оптимизации складских запасов, для решения которой требуется знать, чего и сколько должно быть продано через неделю, и т.п.; задача предсказания стоимости акций какого-нибудь предприятия через день и т.д. и другие подобные вопросы). *Deductor Studio* предлагает для этого инструмент «Прогнозирование».

Прогнозирование появляется в списке мастера обработки только после построения какой-либо модели прогноза: нейросети, линейной регрессии и т.д. Прогнозировать на несколько шагов вперед имеет смысл только временной ряд (к примеру, если есть данные по недельным суммам продаж за определенный период, можно спрогнозировать сумму продаж на две недели вперед). Поскольку при построении модели прогноза необходимо учитывать



много факторов (зависимость результата от данных день, два, три, четыре назад), то методика имеет свои особенности. Покажем ее на примере.

У аналитика имеются данные о месячном количестве проданного товара за несколько лет. Ему необходимо, основываясь на этих данных, сказать, какое количество товара будет продано через неделю и через две. Исходные данные по продажам находятся в файле «Продажи.txt» Выполним импорт данных из файла (рис. 3.14).

После импорта данных воспользуемся диаграммой для их просмотра. На ней видно, что данные содержат аномалии (выбросы) и шумы, за которыми трудно разглядеть тенденцию. Поэтому перед прогнозированием необходимо удалить аномалии и сгладить данные. Сделать это можно при помощи спектральной обработки. Запустим мастер обработки (рис. 3.15), выберем в качестве обработки данных спектральную обработку и перейдем на следующий шаг мастера. Следующий шаг отвечает за удаление аномалий из исходного набора. Выберем поле для обработки «КОЛИЧЕСТВО» и укажем для него вычитание шума (степень вычитания – малая).

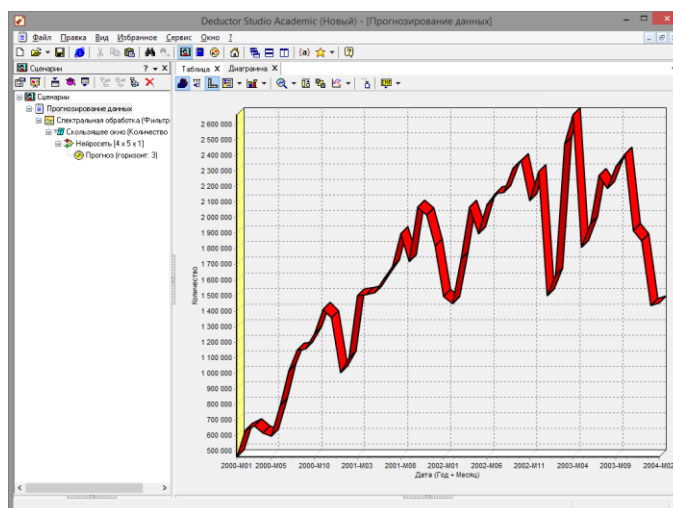


Рис. 3.14 - Диаграмма продаж

На следующем шаге запустим обработку, нажав на «пуск» и посмотрим полученный результат (рис. 3.16). Видно, что данные сгладились, аномалии и шумы исчезли. Также видна тенденция. Теперь перед аналитиком встает вопрос, а как, собственно, прогнозировать временной ряд. Во всех предыдущих примерах мы сталкивались с ситуацией, когда есть входные столбцы - факторы и есть выходные столбцы – результат. В данном случае столбец один.

Строить прогноз на будущее необходимо, основываясь на данных прошлых периодов. Предполагается, что количество продаж на следующий месяц зависит от количества продаж за предыдущие месяцы. Входными факторами для модели могут быть продажи за текущий месяц, продажи за месяц ранее и т.д., а результатом должны быть продажи за следующий месяц. Здесь явно необходимо трансформировать данные к скользящему окну.



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 50 из 74

Первый экземпляр _____

КОПИЯ № _____

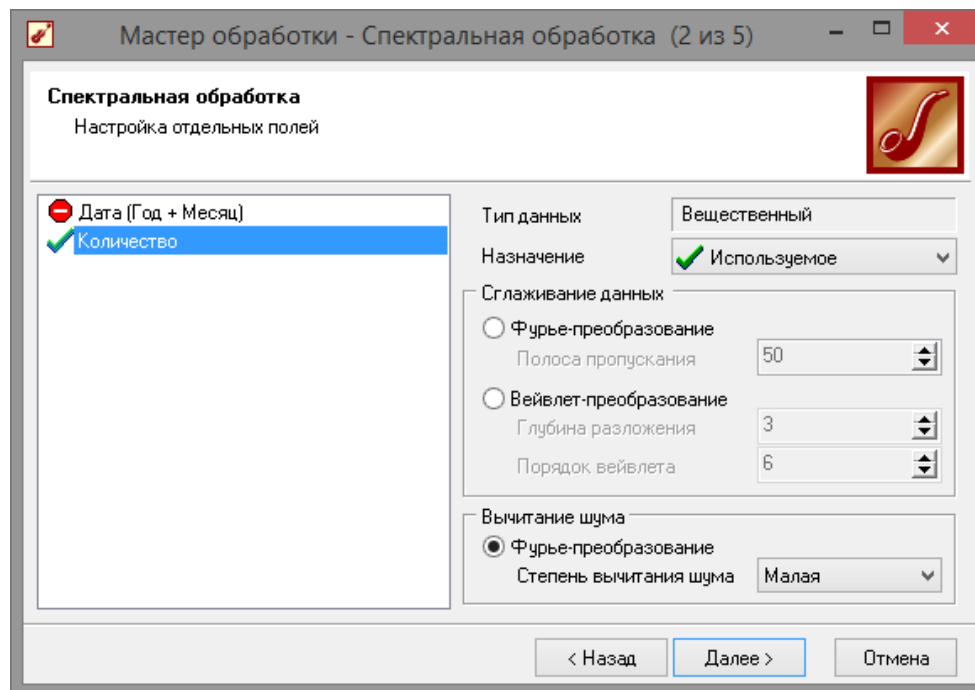


Рис. 3.15 - Выбор спектральной обработки

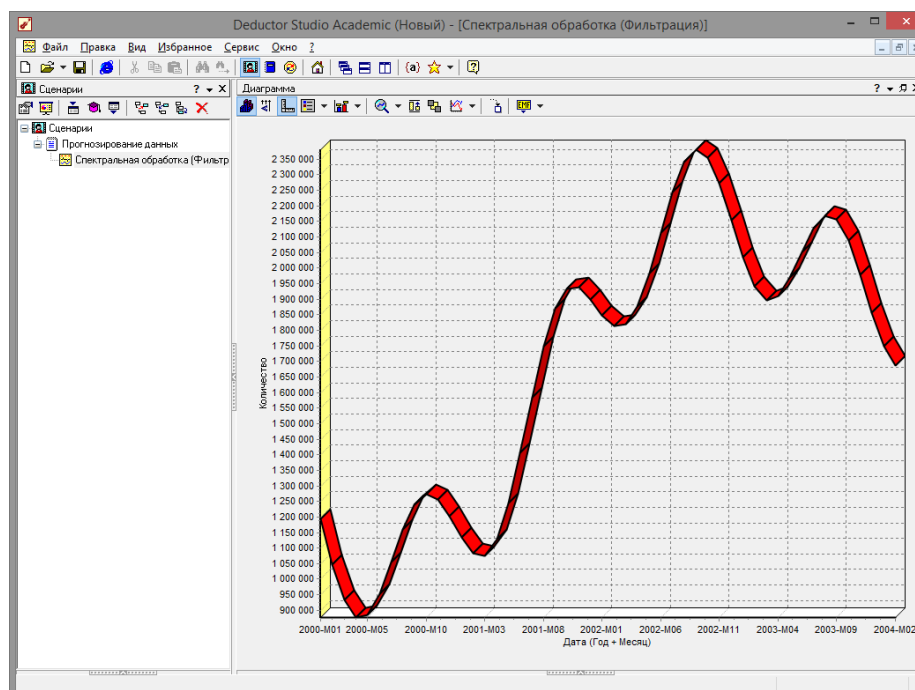
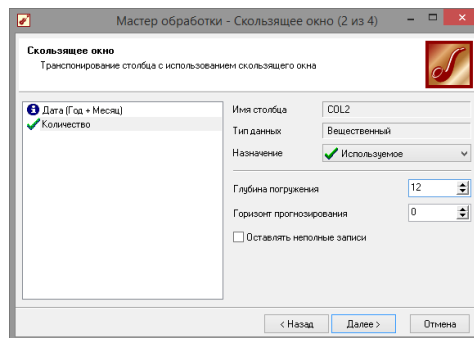


Рис. 3.16 - Результат спектральной обработки



Запустим мастер обработки, выберем в качестве обработчика скользящее окно и перейдем на следующий шаг. Было решено строить прогноз на неделю вперед, основываясь на данных за 12, 11 месяцев назад, два месяца назад и месяц назад. Поэтому необходимо, назначив поле «КОЛИЧЕСТВО» используемым, выбрать глубину погружения 12. Тогда данные трансформируются к скользящему окну так, что аналитику будут доступны все



требуемые факторы для построения прогноза (рис. 3.17).

Рис. 3.17 - Скользящее окно

Просмотреть полученные данные можно в виде таблицы. Как видно, теперь в качестве входных факторов можно использовать «КОЛИЧЕСТВО - 12», «КОЛИЧЕСТВО - 11» - данные по количеству 12 и 11 месяцев назад (относительно прогнозируемого месяца) и остальные необходимые факторы. В качестве результата прогноза будет указан столбец «КОЛИЧЕСТВО».

Перейдем непосредственно к самому построению модели прогноза. Откроем мастер обработки и выберем в нем нейронную сеть (рис. 3.19). На втором шаге мастера, согласно с принятым ранее решением, установим в качестве входных поля «КОЛИЧЕСТВО-12», «КОЛИЧЕСТВО-11», «КОЛИЧЕСТВО-2» и «КОЛИЧЕСТВО-1», а в качестве выходного - «КОЛИЧЕСТВО». Остальные поля сделаем информационными.

Дата (Год + Месяц)	Количество12	Количество11	Количество10	Количество9	Количество8	Количество7
2001.M01	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2001.M02	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2001.M03	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2001.M04	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2001.M05	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2001.M06	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2001.M07	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2001.M08	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2001.M09	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2001.M10	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2001.M11	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2001.M12	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2002.M01	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2002.M02	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2002.M03	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2002.M04	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2002.M05	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2002.M06	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2002.M07	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2002.M08	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2002.M09	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2002.M10	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2002.M11	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2002.M12	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2003.M01	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2003.M02	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2003.M03	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2003.M04	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2003.M05	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2003.M06	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2003.M07	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2003.M08	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2003.M09	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2003.M10	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2003.M11	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344
2003.M12	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344	104415.52428344

Рис. 3.18 - Таблица скользящего окна

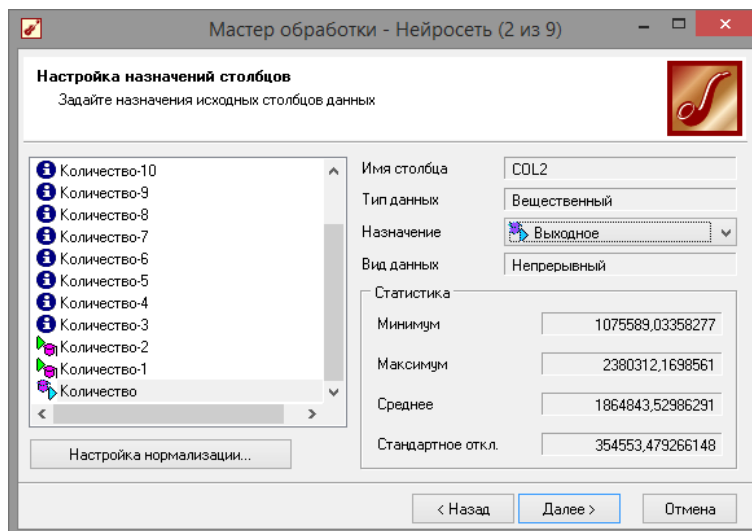


Рис. 3.19 - Настройки мастера спектральной обработки

Оставив остальные параметры построения модели по умолчанию, только количество нейронов скрытого слоя поставим равным 5 (рис. 3.20), обучим нейросеть (см. пример «прогнозирование умножения с помощью нейронной сети») (рис. 3.21). После построения модели для просмотра качества обучения представим полученные данные в виде диаграммы и диаграммы рассеяния. В мастере настройки диаграммы выберем для отображения поля «КОЛИЧЕСТВО» и «КОЛИЧЕСТВО_OUT» - реальное и спрогнозированное значение. Результатом будет два графика, показанные на рис.3.22.

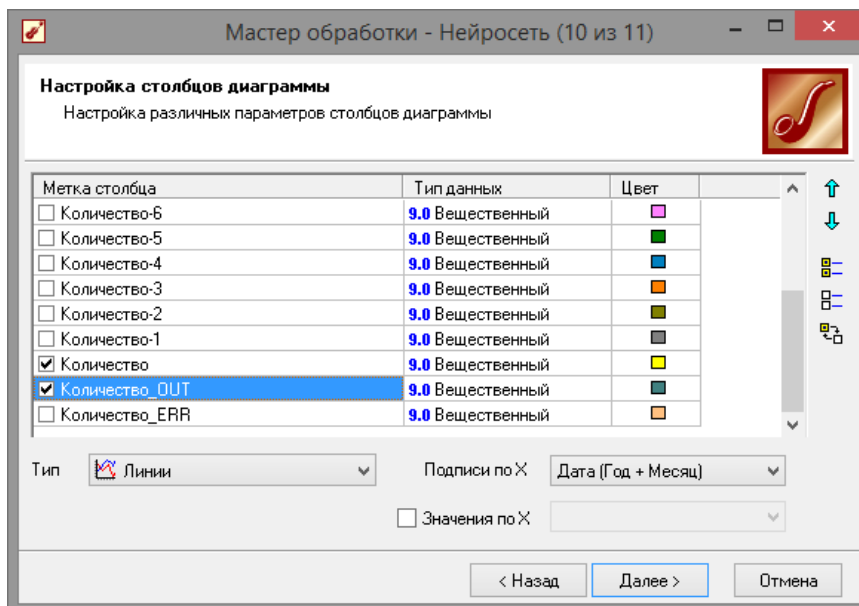


Рис. 3.20 - Настройки нейросети



МИНОБРАЗОВАНИЯ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 53 из 74

Первый экземпляр _____

КОПИЯ № _____

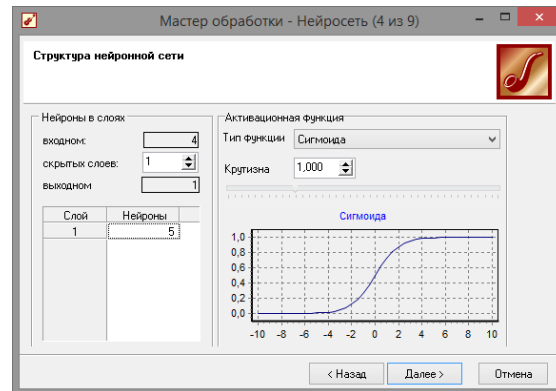


Рис. 3.21 - Обучение нейронной сети

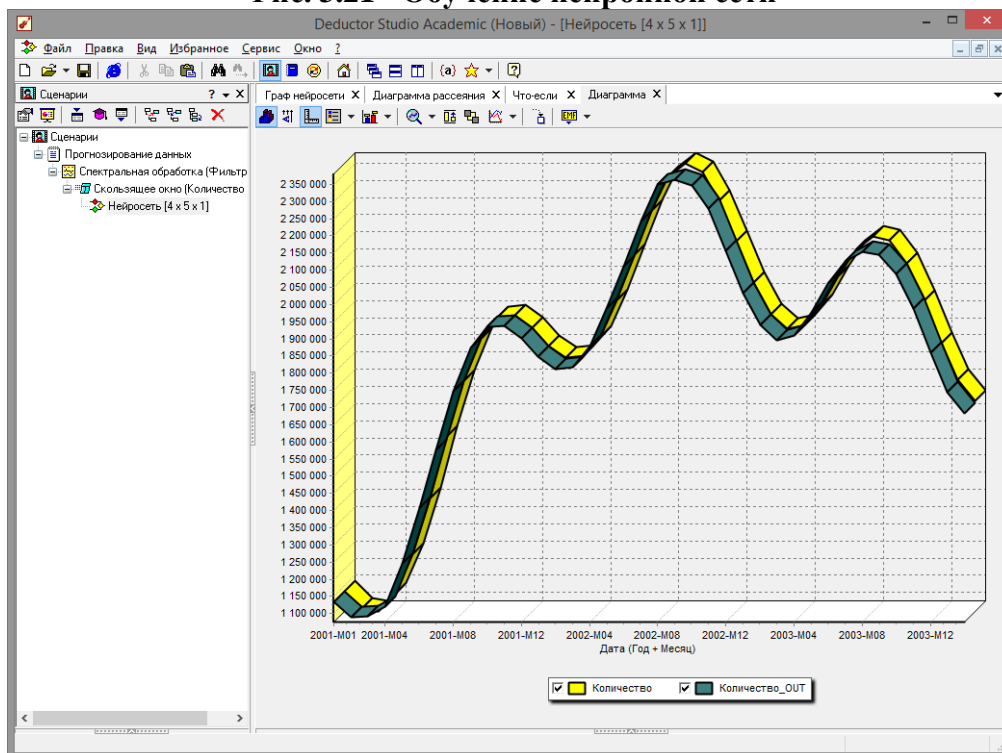
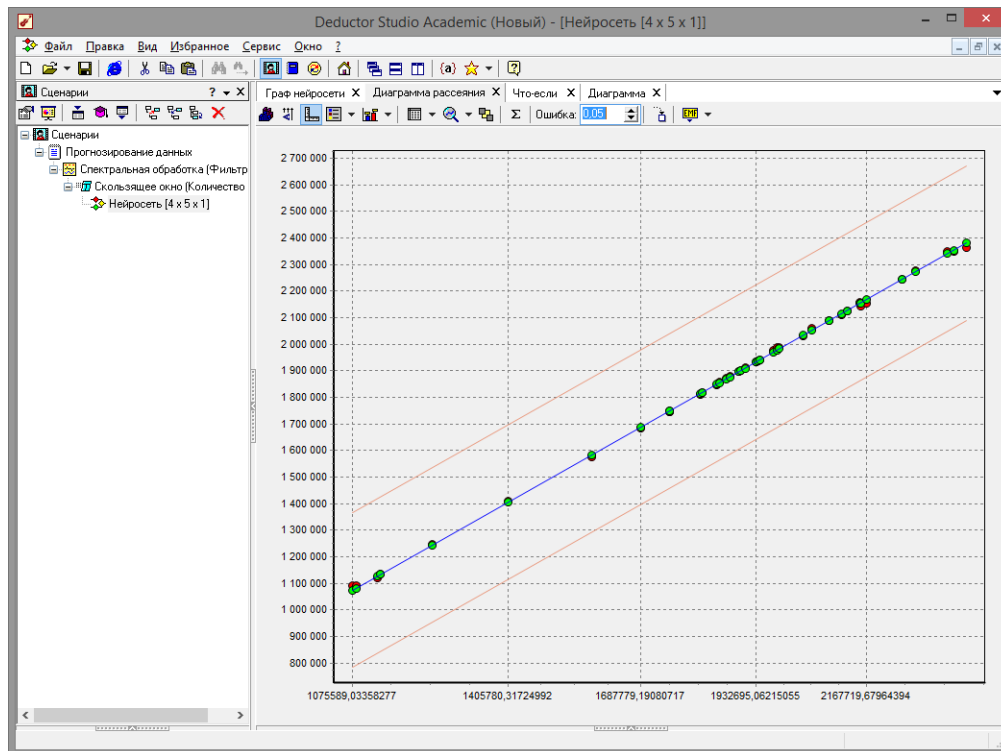


Рис. 3.22 - Сравнение эталонных данных с прогнозом

Диаграмма рассеяния более наглядно показывает качество обучения (рис. 3.23).

**Рис. 3.23 - Диаграмма рассеяния**

Нейросеть обучена, теперь осталось самое главное – построить требуемый прогноз. Для этого открываем мастер обработки (рис. 3.24) и выбираем появившийся теперь обработчик

«Прогнозирование». На втором шаге мастера предлагается настроить связи столбцов для прогнозирования временного ряда – откуда брать данные для столбца при очередном шаге прогноза (рис. 3.25). Мастер сам верно настроил все переходы, поэтому остается только указать горизонт прогноза (на сколько вперед будем прогнозировать) равным трем, а также, для наглядности, необходимо добавить к прогнозу исходные данные, установив в мастере соответствующий флажок.

После этого необходимо в качестве визуализатора выбрать диаграмму прогноза, которая появляется только после прогнозирования временного ряда. В мастере настройки столбцов диаграммы прогноза необходимо указать в качестве отображаемого столбец «КОЛИЧЕСТВО». Теперь аналитик может дать ответ на вопрос, какое количество товаров будет продано в следующем месяце и даже два месяца спустя (рис. 3.26). Масштабировав результат и включив метки, можно увидеть расчетные значения на 3 месяца вперед (рис. 3.27).

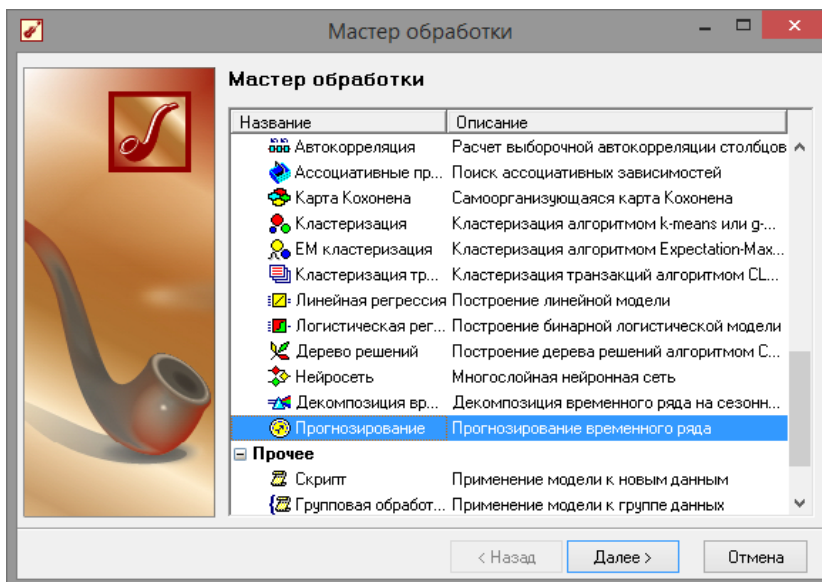


Рис. 3.24 - Мастер прогнозирования

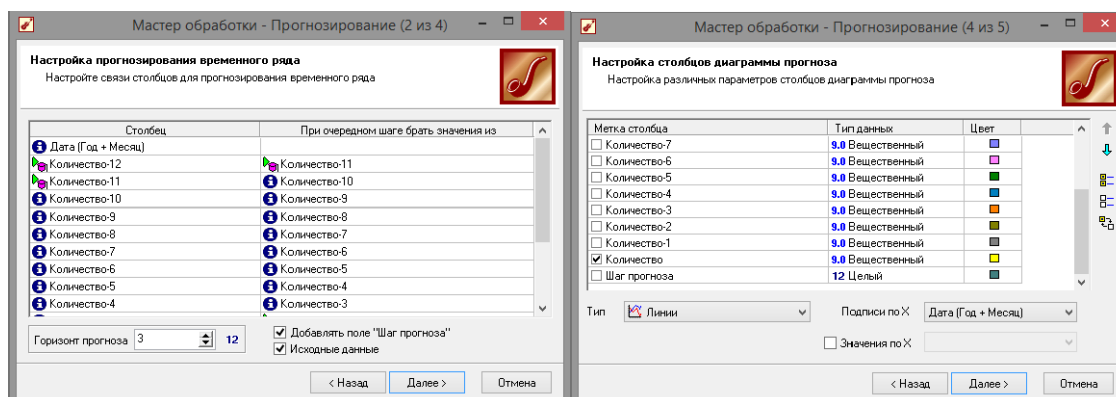


Рис. 3.25 - Мастер прогнозирования

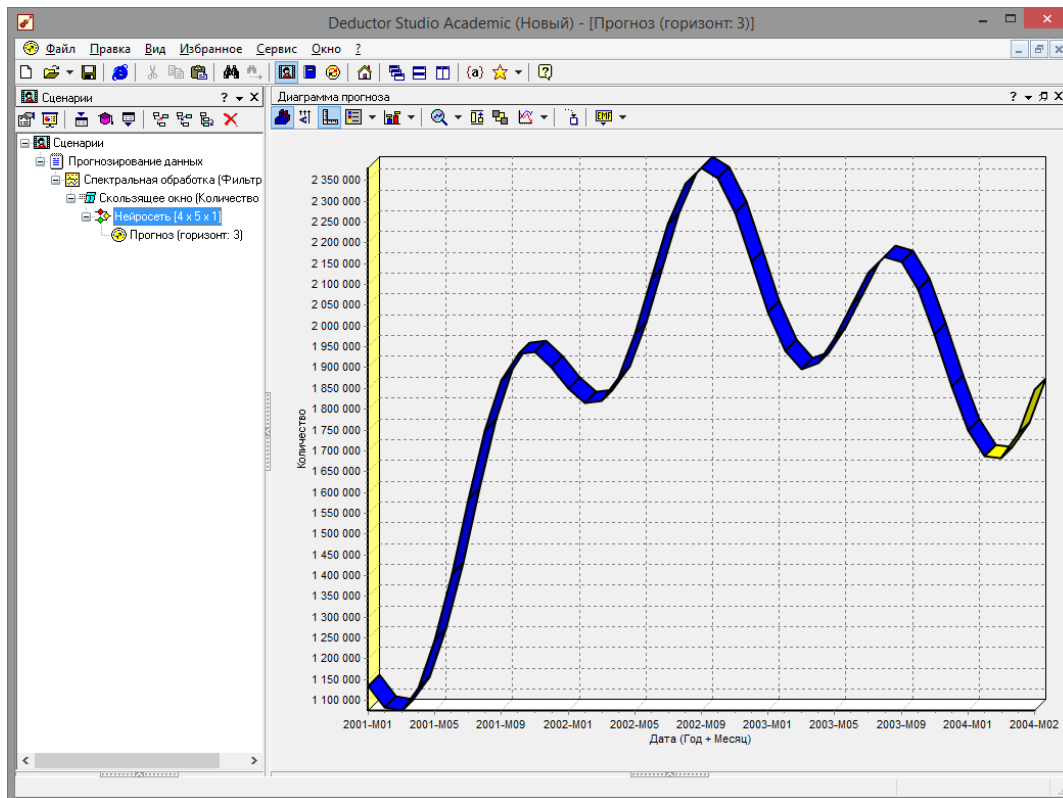


Рис. 3.26 - Расчетные значение на 3 месяца вперед

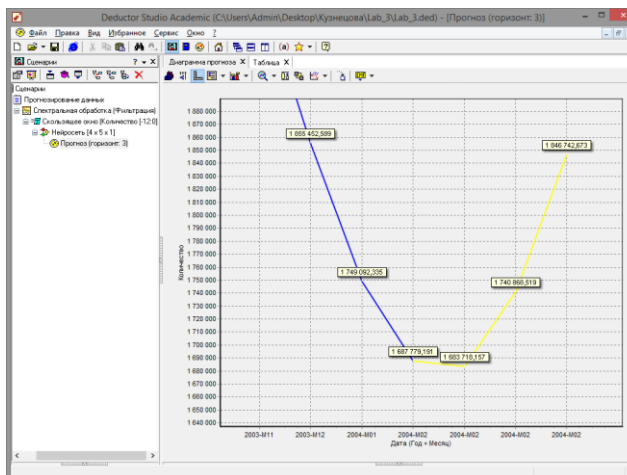


Рис. 3.27. Результаты мастера прогнозирования

После завершения анализа данные можно экспортировать. Так как данная версия является бесплатной для образовательных целей, то данные можно выгрузить либо в текстовый файл, либо в собственный проект программы с расширением «*.ded». Мастер



экспорта имеет точно такие же настройки, как и мастер импорта. Более того, если экспорт данных совершить в текстовый файл, то далее данные можно скопировать в файл табличного процессора *Excel*, и достаточно комфортно с ними работать.

Данный пример показал, как с помощью *Deductor Studio* прогнозировать временной ряд. При решении задачи были применены механизмы очистки данных от шумов, аномалий, которые обеспечили качество построения модели прогноза далее и соответственно достоверный результат самого прогнозирования количества продаж на три месяца вперед. Также был продемонстрирован принцип прогнозирования временного ряда – импорт, выявление сезонности, очистка, сглаживание, построение модели прогноза и собственно построение прогноза временного ряда, а также экспорт результатов во внешний файл.

3.3 Задание на самостоятельную работу

Получить от преподавателя вариант задания для прогнозирования (изменение курса валют, график синусоиды, прогнозирование суммы или разности чисел и др.).

Контрольные вопросы

1. Что такое временной ряд?
2. В какие форматы можно экспортировать данные из *Deductor Academic*?
3. Что такое обучающая и тестовая выборка?
4. Какие инструменты можно использовать для прогнозирования?
5. Для чего служит диаграмма рассеяния?

Библиографический список

1. Официальный сайт *Deductor Academic Studio* [Электронный ресурс] Режим доступа: <http://www.basegroup.ru/>
2. *Deductor 5.3* Руководство аналитика.
3. Встроенная справка приложения.



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 58 из 74

Первый экземпляр _____

КОПИЯ № _____

Практическая работа 4

Нейросетевые технологии в интеллектуальном анализе данных

Продолжительность работы – 8 час.

Цель работы: изучить кластеризацию с помощью самоорганизующихся карт Кохонена в аналитическом пакете Deductor Academic.

Программа работы

1. Произвести импорт данных из подготовленного файла.
2. С помощью Карты Кохонена выполнить кластеризацию на данных контрольного примера.
3. Выполнить задачу кластеризацию для данных поиндивидуальному заданию.

Методические указания по выполнению работы

4.1 Общие понятия о самоорганизующихся картах Кохонена

Самоорганизующаяся карта Кохонена (англ. *Self-organizing map* - *SOM*) – соревновательная нейронная сеть с обучением без учителя, выполняющая задачу визуализации и кластеризации. Идея сети предложена финским учёным Т. Кохоненом. Является методом проецирования многомерного пространства в пространство с более низкой размерностью (чаще всего, двумерное), применяется также для решения задач моделирования, прогнозирования и др.

Самоорганизующаяся карта состоит из компонентов, называемых узлами или нейронами. Их количество задаётся аналитиком. Каждый из узлов описывается двумя векторами. Первый – т. н. вектор веса m , имеющий такую же размерность, что и входные данные. Второй — вектор r , представляющий собой координаты узла на карте. Обычно узлы располагают в вершинах регулярной решётки с квадратными или шестиугольными ячейками.

Самоорганизующаяся карта Кохонена является разновидностью нейронной сети. Она применяется, когда необходимо решить задачу кластеризации, т.е. распределить данные по нескольким кластерам. Алгоритм определяет расположение кластеров в многомерном пространстве факторов. Исходные данные будут относиться к какому-либо кластеру в зависимости от расстояния до него. Многомерное пространство трудно для представления в графическом виде. Механизм же построения карты Кохонена позволяет отобразить многомерное пространство в двумерном, которое более удобно и для визуализации, и для интерпретации результатов аналитиком. Также с помощью построенной карты Кохонена можно решить и задачу прогнозирования. В этом случае результирующее поле (то, которое необходимо спрогнозировать) в построении карты не участвует. После кластеризации,



используя диаграмму «Что-если», можно провести эксперимент. Алгоритм определяет точку пространства, где расположены введенные для прогноза данные и к какому кластеру принадлежит данная точка, и подсчитывает среднее по результирующему полю всех точек этого кластера, что и будет результатом прогноза (для дискретных данных результатом прогноза является значение, больше всего встречающееся в результирующем поле всех ячеек кластера).

4.2. Пример кластеризации данных

Рассмотрим механизм кластеризации путем построения самоорганизующейся карты, основываясь на типичных характеристиках цветков. Исходная таблица находится в файле примеров «Ирисы.txt».

Она содержит следующие параметры цветов: «ДЛИНА ЧАШЕЛИСТИКА», «ШИРИНА ЧАШЕЛИСТИКА», «ДЛИНА ЛЕПЕСТКА», «ШИРИНА ЛЕПЕСТКА», «КЛАСС ЦВЕТКА». Задача состоит в том, чтобы определить по различным параметрам цветка его класс. Предполагается, что цветы одного класса имеют схожие параметры, поэтому они должны находиться в одном кластере.

Для начала необходимо импортировать данные из файла (рис. 4.1). После этого запустим мастер обработки и выберем из списка метод обработки «Карта Кохонена» (рис. 4.2). На втором шаге мастера настроим назначения столбцов (рис. 4.3). Укажем столбцу «КЛАСС ЦВЕТКА» назначение «Выходной», а остальным – «Входной». Т.е. на основе данных о цветке будем относить его к тому или иному классу.

Длина чашелистика	Ширина чашелистика	Длина лепестка	Ширина лепестка	Класс
5.1	3.5	1.4	0.2	Iris-setosa
4.9	3	1.4	0.2	Iris-setosa
4.7	3.2	1.3	0.2	Iris-setosa
4.6	3.1	1.5	0.2	Iris-setosa
5	3.6	1.4	0.2	Iris-setosa
5.4	3.9	1.7	0.4	Iris-setosa
4.6	3.4	1.4	0.3	Iris-setosa
5	3.4	1.5	0.2	Iris-setosa
4.4	2.9	1.4	0.2	Iris-setosa
4.9	3.1	1.5	0.1	Iris-setosa
5.4	3.7	1.5	0.2	Iris-setosa
4.8	3.4	1.6	0.2	Iris-setosa
4.8	3	1.4	0.1	Iris-setosa
4.3	3	1.1	0.1	Iris-setosa
5.8	4	1.2	0.2	Iris-setosa
5.7	4.4	1.5	0.4	Iris-setosa
5.4	3.9	1.3	0.4	Iris-setosa
5.1	3.5	1.4	0.3	Iris-setosa
5.7	3.8	1.7	0.3	Iris-setosa

Рис. 4.1 - Импортированные данные



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 60 из 74

Первый экземпляр _____

КОПИЯ № _____

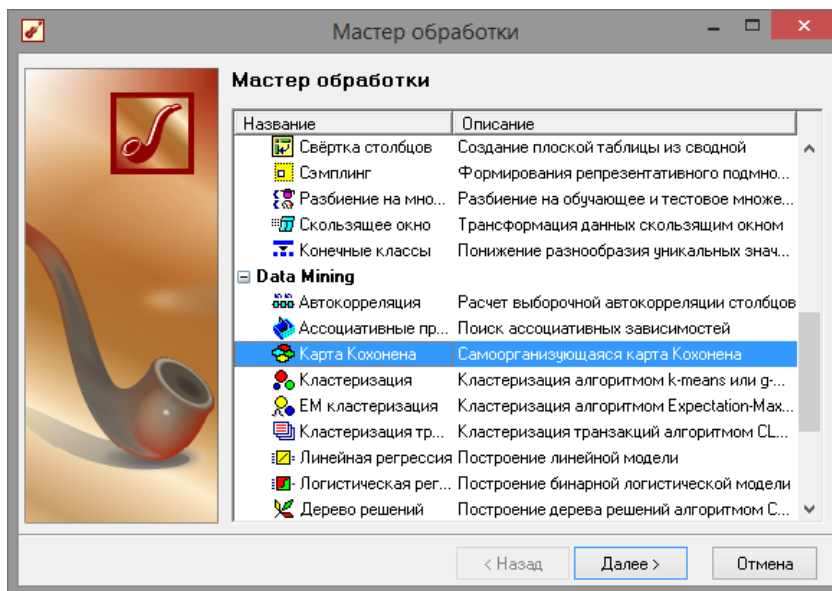


Рис. 4.2 - Мастер обработки «Карта Кохонена»

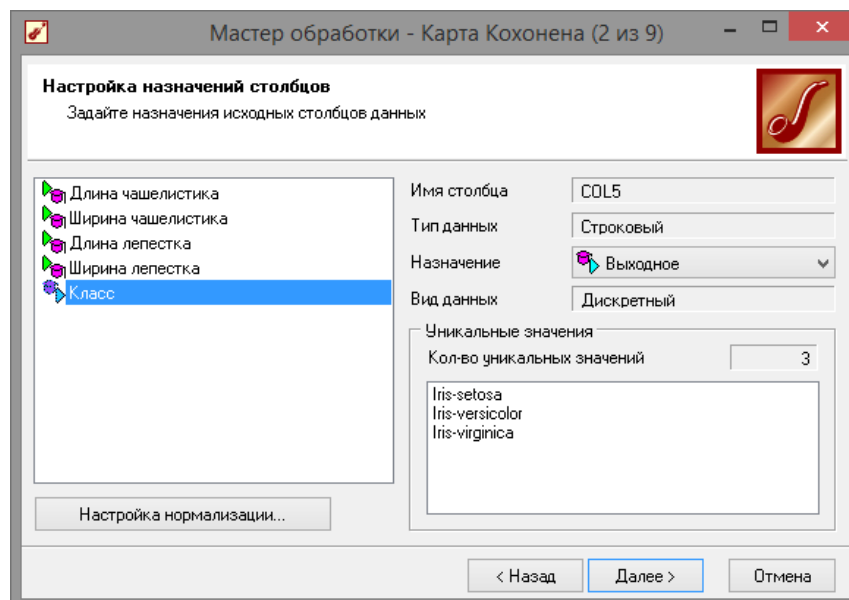


Рис. 4.3. Настройки мастера карт Кохонена

На третьем шаге мастера необходимо настроить способ разделения исходного множества данных на тестовое и обучающее, а также количество примеров в том и другом множестве. Укажем, что данные обоих множеств берутся случайным образом, зададим размер тестового множества равным десяти примерам, путем изменения значения столбца «Размер в строках» строки «Тестовое множество»(рис. 4.4).



Следующий шаг предлагает настроить параметры карты (количество ячеек по X и по Y, их форму) и параметры обучения (способ начальной инициализации, тип функции соседства, перемешивать ли строки обучающего множества и количество эпох, через которые необходимо перемешивание). Значения по умолчанию вполне подходят (рис. 4.5).

На пятом шаге мастера необходимо настроить параметры остановки обучения. Оставим параметры по умолчанию (рис. 4.6).

На шестом шаге настраиваются остальные параметры обучения – способ начальной инициализации, тип функции соседства и также параметры кластеризации – автоматическое определение числа кластеров с соответствующим уровнем значимости либо фиксированное количество кластеров предоставляется возможность настроить интервалы обучения. Каждый интервал задается количеством эпох, радиусом обучения и скоростью обучения. Укажем фиксированное количество кластеров, равное трем (рис. 4.7).

Множество	Размер		Порядок сортировки
	В процентах	В строках	
<input checked="" type="checkbox"/> Обучающее	93,33	140	По возрастанию
<input checked="" type="checkbox"/> Тестовое	6,67	10	По возрастанию
ИТОГО:	100,00	150	

Рис.4.4 - Настройки тестового и обучающего множества



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по
основной профессиональной образовательной программе высшего образования – программе
бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 62 из 74

Первый экземпляр _____

КОПИЯ № _____

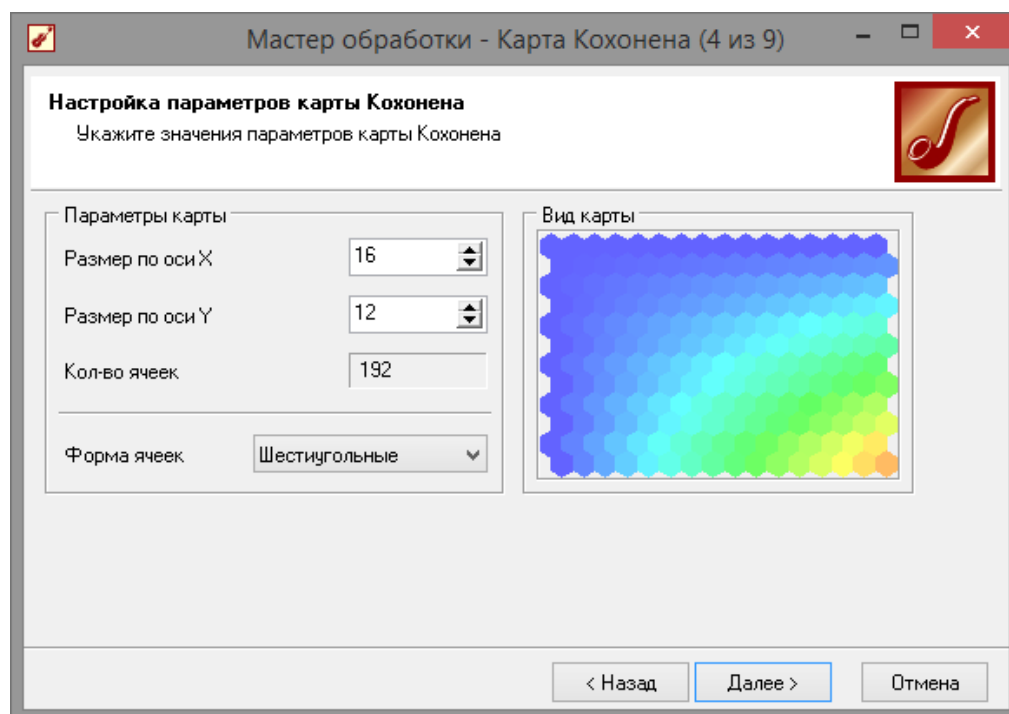


Рис. 4.5 - Настройки значения параметров карт



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 63 из 74

Первый экземпляр _____

КОПИЯ № _____

Рис.4.6 - Настройки параметров остановки обучения

Рис. 4.7 - Настройки параметров обучения карты Кохонена



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 64 из 74

Первый экземпляр _____

КОПИЯ № _____

На седьмом шаге предлагается запустить сам процесс обучения (рис. 4.8). Во время обучения можно посмотреть количество распознанных примеров и текущие значения ошибок. Здесь необходимо нажать на кнопку пуск и дождаться завершения процесса обработки.

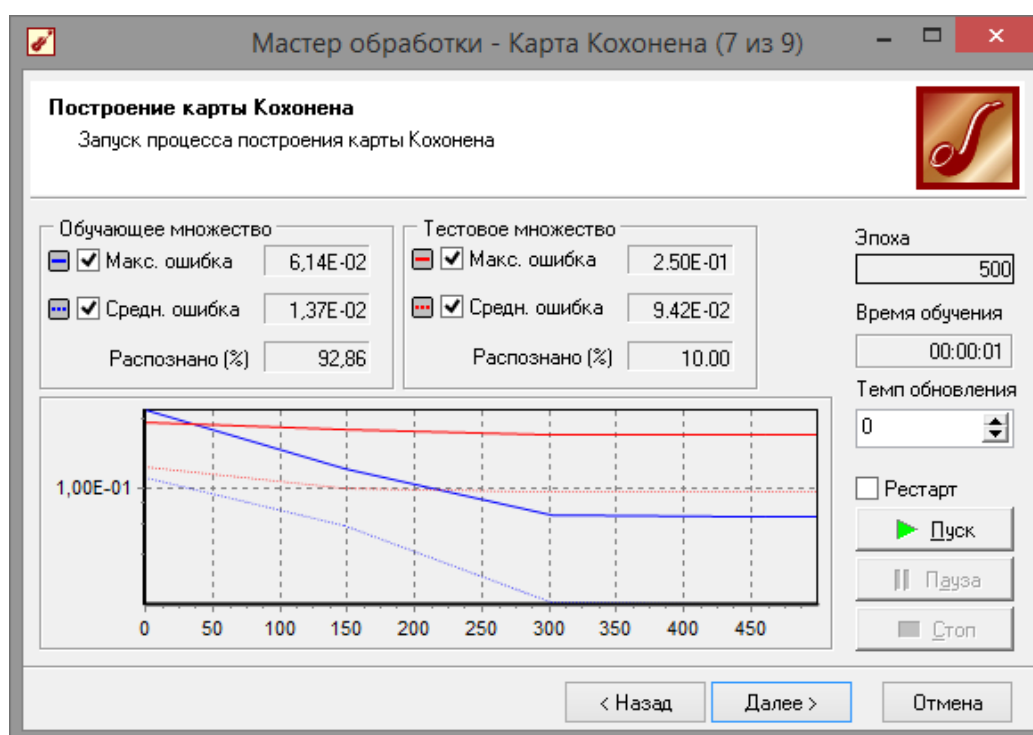


Рис. 4.8 - Процесс обучения

После этого необходимо в списке визуализаторов выбрать появившуюся теперь «Карту Кохонена» для просмотра результатов кластеризации, а также визуализатор «Что-если» для прогнозирования класса цветка (рис. 4.9).

Далее, в мастере настройки отображения карты Кохонена необходимо указать, чтобы отображались все поля, также следует установить количество кластеров равным трем и поставить флажок «Границы кластеров» (рис. 4.10).



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 65 из 74

Первый экземпляр _____

КОПИЯ № _____

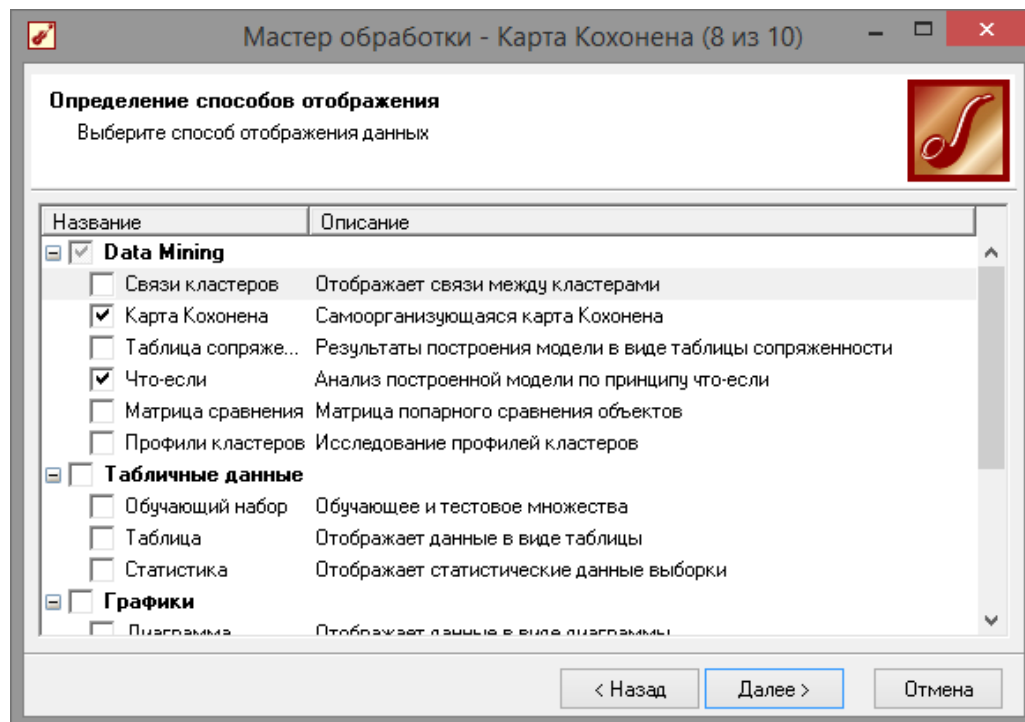


Рис. 4.9 - Способ отображения данных

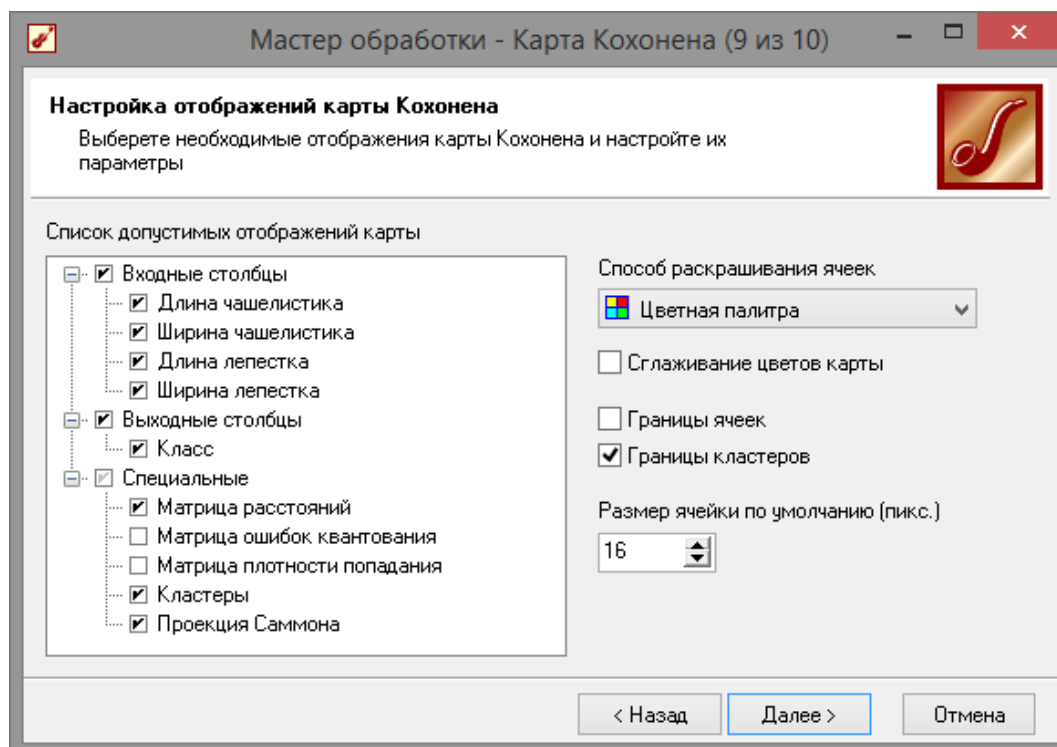


Рис. 4.10 - Настройка отображения кластеров



После этого можно увидеть полученные результаты (рис. 4.11). Качество кластеризации можно оценить, просмотрев карту «КЛАСС ЦВЕТКА». На ней видно, что большинство цветов были классифицированы правильно. Заметим, что все цветы класса *Setosa* попали в один кластер. Это говорит о значительном отличии параметров цветов этого класса от других. Явное различие наблюдается по длине и ширине лепестка. То, что часть примеров *Virginica* попала в класс *Versicolor* и наоборот говорит о меньшем различии этих классов. На картах, в отличие от *Setosa* не видны резкие отличия параметров цветов этих двух классов. Этим как раз и объясняется «проникновение» некоторой части примеров в другой кластер.

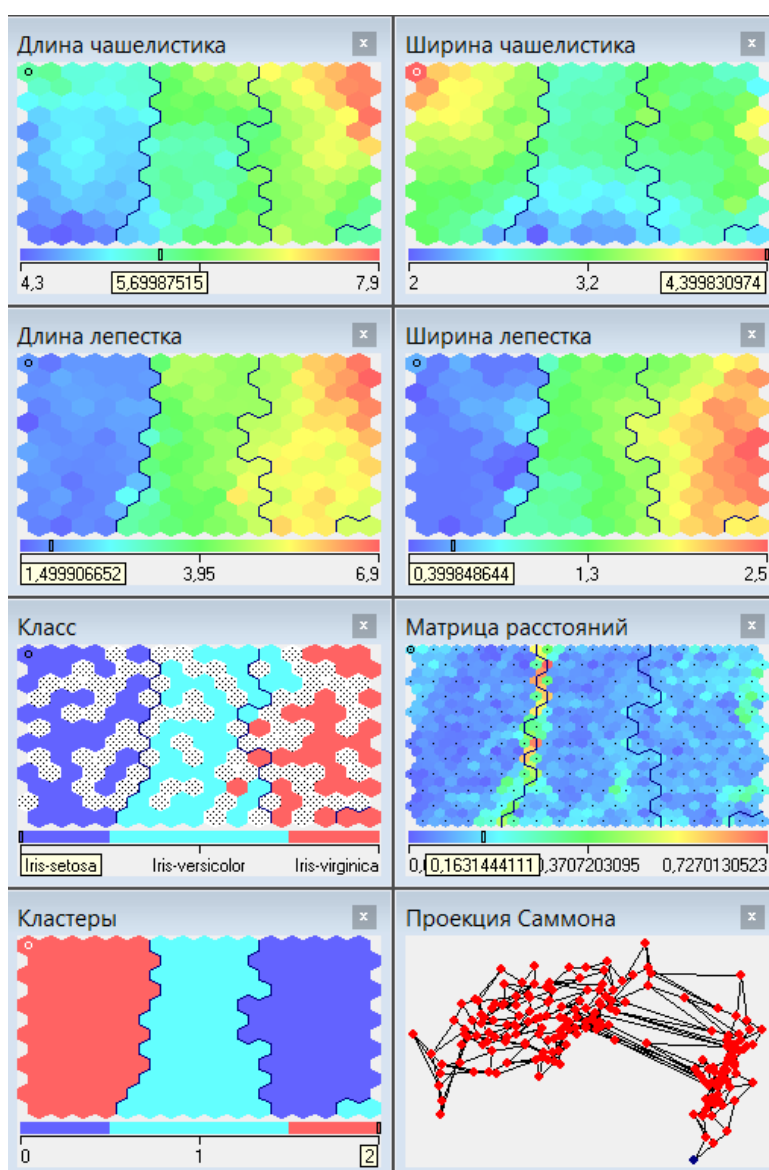


Рис. 4.11 - Карта Кохонена



Рассмотрим построенную таблицу «Что-Если». В верхней части таблицы отображаются входные поля, а в нижней – выходные и расчетные. Изменяя значения входных полей, пользователь дает команду на выполнение расчета и наблюдает за рассчитанными значениями выходов нейронной сети или дерева решений.

Расчетные поля отличаются от выходных тем, что они не существуют в исходном наборе данных и были созданы в ходе обработки. Такими полями являются, например, «Номер ячейки» или «Номер кластера».

Каждое поле таблицы «Что-Если» представлено следующими атрибутами:

- «Тип» – указывается значок, соответствующий типу данных поля;
- «Поле» – имя входного или выходного поля;
- «Значение» – указывается текущее значение поля.

С помощью кнопки «Показать статистику» справа от таблицы можно вывести статистику по выделенному полю. Для непрерывных полей в ней отображается следующая информация:

- «Минимум» – минимальное значение поля в выборке;
- «Максимум» – максимальное значение поля в выборке;
- «Среднее» – среднее по выборке значение поля;
- «Стандартное откл.» – среднеквадратическое отклонение значений поля по выборке.

Знание диапазона входных данных (минимума и максимума), на котором строилась модель, позволит определить область устойчивости системы. Очевидно, что если подать на вход значения, существенно выходящие за диапазон, гарантировать правильную реакцию системы нельзя, и достоверность полученных данных может быть снижена. Если значение, присвоенное полю, выходит за границы диапазона, это поле окрашивается в красный цвет.

Для дискретных полей статистика содержит:

- «Значения» – список уникальных значений;
- «Кол-во» – число вхождений значения в выборку;
- «Итоговая информация» – общее число уникальных значений в выборке.

Для дискретных значений на вход можно подавать только значения, представленные в этом списке.



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по
основной профессиональной образовательной программе высшего образования – программе
бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 68 из 74

Первый экземпляр _____

КОПИЯ № _____

В таблице пользователь может менять лишь содержимое столбца «Значение». Это осуществляется несколькими способами:

- непосредственно ввести данные с клавиатуры;
- заполнить записями из текущей выборки (при этом вводятся записи целиком и заполняются одновременно все поля);
- выбрать значения из статистики, находящейся справа от таблицы.

Чтобы ввести значения входов с клавиатуры, нужно выбрать ячейку «Значение» для соответствующего поля, и только потом вводить данные. Чтобы войти в режим редактирования, достаточно напечатать любой символ с клавиатуры в том числе «Enter» либо дважды "кликнуть" мышкой по соответствующей ячейке. Дискретные значения выбираются из выпадающего списка либо путем циклического перебора в следствии двойного "клика" мышкой. Для перехода к предыдущим или последующим строкам используются клавиши со стрелками. Если введенные вами значения выходят за диапазон значений выборки, соответствующая строка таблицы выделяется красным цветом. Если находясь на ячейке, нажать клавишу "Del", то значение соответствующего входного поля будет очищено.

Для автоматического ввода в таблицу «Что-Если» записей из текущей выборки используются кнопки на панели инструментов:

- Первая запись (Ctrl+PgUp) – позволяет выбрать для загрузки в таблицу «Что-Если» первую запись выборки;
- Предыдущая запись (PgUp) – позволяет загрузить предыдущую запись;
- Загрузить запись – загружает текущую запись в соответствующие входные поля таблицы «Что-Если»;
- Загрузить из исходной выборки – выводит на экран модальное окно с таблицей, из которой можно загрузить необходимую запись;

Как видно из таблицы «Что если» (рис. 4.12), даже данные отсутствующие в изначальной выборке определяются корректно.



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по
основной профессиональной образовательной программе высшего образования – программе
бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 69 из 74

Первый экземпляр _____

КОПИЯ № _____

Поле	Значение
Входные	
9.0 Длина чашелистика	4,7
9.0 Ширина чашелистика	3,8746
9.0 Длина лепестка	1,56
9.0 Ширина лепестка	0,345
Выходные	
ab Класс	Iris-setosa
Расчетные	
12 Номер ячейки	48
9.0 Расстояние до центра ячейки	0,116971545667132
12 Номер кластера	2
9.0 Расстояние до центра кластера	0,162006332142713

Параметр	Значение
Минимум	2
Максимум	4,4
Среднее	3,054
Стандартное откл.	0,4335943114

Рис. 4.12 - Таблица «Что-Если»

Данный пример показал область применения самоорганизующихся карт. Изначально имелось многомерное (четырёхмерное) пространство входных факторов. Алгоритм представил его в двумерном виде, которое удобнее анализировать. Также исходные данные были отнесены к трем кластерам, по типу цветка – «*Setosa*», «*Versicolor*», «*Virginica*». Основным визуализатором после построения является «Самоорганизующаяся карта». Здесь в первую очередь следует обратить внимание на матрицу расстояний и проекцию Саммона. На них явно видны расстояния между отдельными ячейками карты, т.е. четкие границы различных скоплений данных. Мастер предоставляет широкий набор настройки параметров обучения: настройка нормализации столбцов, настройка разбиения на тестовое и обучающее множество, настройка условий останковки обучения, настройка параметров карты и параметров обучения, настройка интервалов обучения. информация о задаче, то качество очистки данных можно увеличить на порядки.



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 70 из 74

Первый экземпляр _____

КОПИЯ № _____

4.3 Задание на самостоятельную работу

Получить от преподавателя вариант задания на кластеризации (классификация помидоров по диаметру плода, весу плода, количеству плодов на кусте, высоте куста, или задача классификации призывников по категориям на основе их параметров и др.).

Решить задачу при помощи карт Кохонена.

Контрольные вопросы

1. Что такое карта Кохонена?
2. Что решают задачи кластеризации?
3. Для чего служит таблица «Что-Если»?

Библиографический список

1. Официальный сайт Deductor Academic Studio [Электронный ресурс] Режим доступа: <http://www.basegroup.ru/>
2. Deductor 5.3 Руководство аналитика.
3. Встроенная справка приложения.

Самостоятельная работа студента

СРС № 1

Задание Составить опорный конспект по теме: "Методы нечеткой логики в задачах анализа данных"

СРС № 2

Задание

Многомерные задачи классификации и распознавания. Основные технологии статистической классификации многомерных случайных объектов средствами ИАД.

СРС № 3

Задание

Сети прямого распространения ошибки. Карты Кохонена

СРС № 4

Генетические алгоритмы анализа данных



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 71 из 74

Первый экземпляр _____

КОПИЯ № _____

3.3 Критерии оценивания

Виды текущего контроля	Высокий уровень 90-100%	Продвинутый уровень 75-89%	Пороговый уровень 50-74%	Недопустимый 0-49%
Конспект, структурно-логические схемы, практическая работа, презентация, ментальная карта, проектная работа и т.д.	27-30	23-26	15-22	0-14
Задание самостоятельной работы	27-30	23-26	15-22	0-14

В соответствии с Положением о балльно-рейтинговой системе оценки результатов обучения студентов посещение оценивается следующим образом:

- менее 50% занятий – 0 баллов;
- 50 – 74% занятий – 15 баллов;
- 75 – 89% занятий – 18 баллов;
- 90 – 100% занятий – 20 баллов

Критерии оценивания конспектов

«отлично» (90-100%) выставляется студенту, если демонстрируются полнота использования учебного материала, логика изложения (наличие схем, количество смысловых связей между понятиями), наглядность (наличие рисунков, символов и пр.; аккуратность выполнения, читаемость конспекта, грамотность (терминологическая и орфографическая).

«хорошо» (75-89%) выставляется студенту, если демонстрируются использование учебного материала неполное, недостаточно логично изложено (наличие схем, количество смысловых связей между понятиями), наглядность (наличие рисунков, символов и пр.; аккуратность выполнения, читаемость конспекта, грамотность (терминологическая и орфографическая), отсутствие связанных предложений.

«удовлетворительно»(50-74%) выставляется студенту, если демонстрируются использование учебного материала неполное, недостаточно логично изложено (наличие схем, количество смысловых связей между понятиями), наглядность (наличие рисунков, символов, и пр.; аккуратность выполнения, читаемость конспекта, грамотность (терминологическая и орфографическая), прослеживается несамостоятельность при составлении.

«неудовлетворительно» (0-49%) выставляется студенту, если демонстрируются использование учебного материала неполное, отсутствуют схемы, количество смысловых



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)
Костанайский филиал
Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по
основной профессиональной образовательной программе высшего образования – программе
бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 72 из 74

Первый экземпляр _____

КОПИЯ № _____

связей между понятиями, отсутствует наглядность (наличие рисунков, символов, и пр.; аккуратность выполнения, читаемость конспекта, допущены ошибки терминологические и орфографические, несамостоятельность при составлении.

Критерии оценивания устного опроса

«отлично» (90-100%) ставится, если студент полно излагает материал (отвечает на вопрос), дает правильное определение основных понятий; обнаруживает понимание материала, может обосновать свои суждения, применить знания на практике, привести необходимые примеры не только из учебника, но и самостоятельно

составленные; излагает материал последовательно и правильно с точки зрения норм литературного языка.

«хорошо» (75-89%) ставится, если студент дает ответ, удовлетворяющий тем же требованиям, что и для «отлично», но допускает 1–2 ошибки, которые сам же исправляет, и 1–2 недочета в последовательности и языковом оформлении излагаемого.

«удовлетворительно» (50-74%) ставится, если студент обнаруживает знание и понимание основных положений данной темы, но излагает материал неполно и допускает неточности в определении понятий или формулировке правил; не умеет достаточно глубоко и доказательно обосновать свои суждения и привести свои примеры; излагает материал непоследовательно и допускает ошибки в языковом оформлении излагаемого.

«неудовлетворительно» (0-49%) ставится, если студент обнаруживает незнание большей части соответствующего вопроса, допускает ошибки в формулировке определений и правил, искажающие их смысл, беспорядочно и неуверенно излагает материал.

Критерии оценки тестовых заданий

«отлично» (А, А-, 4.0-3.67, 90-100%) - выставляется, если студент выполнил работу без ошибок и недочётов, допустил не более 10% ошибок от общего объема заданий.

«хорошо» (В+, В, В-, 3.33-2.67, 75-89%) - выставляется, если студент выполнил работу, допустив от 11% до 25% ошибок от общего объема заданий.

«удовлетворительно» (С+, С, С-, D+,D, 2.33-1.0, 50-74%) - выставляется, если студент выполнил работу, допустив от 26% до 50% ошибок от общего объема заданий.

«неудовлетворительно» (F, 0, 0-49%) - выставляется, если студент допустил число ошибок и недочётов превосходящее норму, при которой может быть выставлено «удовлетворительно», или если правильно выполнил менее половины работы.

Критерии оценивания самостоятельных работ

«отлично» (90-100%) выставляется, если студент выполнил работу без ошибок и недочётов, допустил не более одного недочёта.

«хорошо» (75-89%), если студент выполнил работу полностью, но допустил в ней не более одной негрубой ошибки и одного недочёта, или не более двух недочётов.

«удовлетворительно» (50-74%), если студент правильно выполнил не менее половины работы или допустил не более двух грубых ошибок, или не более одной грубой и одной негрубой ошибки и одного недочета, или не более двух-трех негрубых ошибок, или одной



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)
Костанайский филиал
Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по
основной профессиональной образовательной программе высшего образования – программе
бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 73 из 74

Первый экземпляр _____

КОПИЯ № _____

негрубой ошибки и трех недочётов, или при отсутствии ошибок, но при наличии четырех-пяти недочётов, допускает искажение фактов.

«неудовлетворительно» (0-49%) , если студент допустил число ошибок и недочётов превосходящее норму, при которой может быть выставлено «удовлетворительно», или если правильно выполнил менее половины работы.

Критерии оценивания практических работ:

«отлично» (90-100%) – задание по работе выполнено в полном объёме. Студент точно ответил на контрольные вопросы, свободно ориентируется в предложенном решении, может его модифицировать при изменении условия задачи. Ответил на все дополнительные вопросы на защите. Отчёт выполнен аккуратно и в соответствии с предъявляемыми требованиями. Показал отличное владения навыками применения полученных знаний и умений при решении профессиональных задач в рамках усвоенного учебного материала.

«хорошо»(75-89%) – задание по работе выполнено в полном объёме с небольшими неточностями. Студент ответил на теоретические вопросы, испытывая небольшие затруднения. Ответил на большинство дополнительных вопросов на защите. Качество оформления отчёта к работе не полностью соответствует требованиям. Показал хорошие владения навыками применения полученных знаний и умений при решении профессиональных задач в рамках усвоенного учебного материала.

«удовлетворительно» (50-74%) – студент выполнил задание с существенными неточностями, не может полностью объяснить полученные результаты. Составил отчёт в установленной форме, представил решения большинства заданий, предусмотренных в работе. При ответах на дополнительные вопросы на защите допустил много неточностей. Показал удовлетворительное владение навыками применения полученных знаний и умений при решении профессиональных задач в рамках усвоенного учебного материала.

«неудовлетворительно» (0-49%)– студент не выполнил все задания работы и не может объяснить полученные результаты. При ответах на дополнительные вопросы на защите было допущено множество ошибок. Продемонстрировал недостаточный уровень владения умениями и навыками при решении профессиональных задач в рамках усвоенного учебного материала.

Критерии оценки структурно-логической схемы(блок-схемы):

Работа выполнена на "отлично" (А, А-, 4.0-3.67, 90-100%) содержание схемы полностью соответствует заданной теме

Работа выполнена на ""хорошо" (В+, В, В-, 3.33-2.67, 75-89%) содержание материала в схеме соответствует заданной теме, но есть недочеты и незначительные ошибки

Работа выполнена на ""удовлетворительно" обучающийся работу выполнил небрежно, содержание блоков схемы по некоторым аспектам не раскрывает вопросы заданной темы

Иначе, студент получает оценку «неудовлетворительно»(F, 0, 0-49%).

Критерии оценки презентации

Работа выполнена на "отлично" (А, А-, 4.0-3.67, 90-100%) если: презентация соответствует теме самостоятельной работы; оформлен титульный слайд с заголовком (тема, цели, план и



МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования
«Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ»)
Костанайский филиал
Кафедра социально-гуманитарных и естественнонаучных дисциплин

Фонд оценочных средств по дисциплине (модулю) «Интеллектуальный анализ данных» по
основной профессиональной образовательной программе высшего образования – программе
бакалавриата «Бизнес логистика» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 74 из 74

Первый экземпляр _____

КОПИЯ № _____

т.п.); сформулированная тема ясно изложена и структурирована; использованы графические изображения (фотографии, картинки и т.п.), соответствующие теме; выдержан стиль, цветовая гамма, использована анимация, звук; работа оформлена и предоставлена в установленный срок.

Работа выполнена на ""хорошо" (В+, В, В-, 3.33-2.67, 75-89%) если: презентация соответствует теме самостоятельной работы; оформлен титульный слайд с заголовком (тема, цели, план и т.п.); сформулированная тема ясно изложена и структурирована; использованы графические изображения (фотографии, картинки и т.п.), соответствующие теме; работа оформлена и предоставлена в установленный срок.

«неудовлетворительно» (F, 0, 0-49%) выставляется студенту, если работа не выполнена или содержит материал не по вопросу.

Во всех остальных случаях работа оценивается на «удовлетворительно» (С+, С, С-, D+,D, 2.33-1.0, 50-74%).

Требования к оформлению презентации:

Презентация создается по указанной теме. Объем презентации не менее 10 слайдов. Фон слайдов – однотонный. Выравнивание текста слева, заголовки – по центру. Шрифт текста на слайде – 28-30 пт. Рекомендуется на слайде располагать рисунки или иллюстрации. При создании презентации, можно использовать рекомендуемую литературу, так и ресурсы Интернет. При защите учитывается наглядность презентации, содержание и соответствие материала.

Отметки о продлении срока действия

Фонд оценочных средств пролонгирован на 2024 / 2025 учебный год решением учёного совета Костанайского филиала ФГБОУ ВО «ЧелГУ» от «18» апреля 2024 г. Протокол № 09

Фонд оценочных средств пролонгирован на 20__ / 20__ учебный год решением учёного совета Костанайского филиала ФГБОУ ВО «ЧелГУ» от «__» ____ 20__ г. Протокол № ____

Фонд оценочных средств пролонгирован на 20__ / 20__ учебный год решением учёного совета Костанайского филиала ФГБОУ ВО «ЧелГУ» от «__» ____ 20__ г. Протокол № ____