



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 2 из 81

Первый экземпляр _____

КОПИЯ № _____

Методические рекомендации приняты

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

Методические рекомендации рекомендованы

Учебно-методическим советом Костанайского филиала ФГБОУ ВО «ЧелГУ»

Протокол заседания № 10 от 18 мая 2023 г.

Председатель
Учебно-методического совета

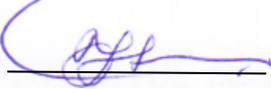
Н.А. Нализко

Методические рекомендации разработаны и рекомендованы кафедрами 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

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 3 из 81

Первый экземпляр _____

КОПИЯ № _____

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ СТУДЕНТАМ ПО ПОДГОТОВКЕ К ЗАНЯТИЯМ ЛЕКЦИОННОГО ТИПА

В ходе занятий лекционного типа обучающийся должен вести конспектирование учебного материала. Обращать внимание на категории, формулировки, раскрывающие содержание тех или иных явлений и процессов, научные выводы и практические рекомендации, положительный опыт в ораторском искусстве. Желательно оставить в рабочих конспектах поля, на которых делать пометки из рекомендованной литературы, дополняющие материал прослушанной лекции, а также подчеркивающие особую важность тех или иных теоретических положений. Задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций.

Конспект (от лат. conspectus – обзор) – письменный текст, в котором кратко и последовательно изложено содержание основного источника информации.

Выделяют две разновидности конспектирования:

- конспектирование письменных текстов;
- конспектирование устных сообщений (например, лекций).

Конспект может быть кратким или подробным.

В основе процесса конспектирования лежит систематизация прочитанного или услышанного материала. Целью процесса служит приведение в единый порядок сведений, полученных из научной статьи, учебной и методической литературы.

Конспектирование является неотъемлемой формы работы обучаемого в силу того, что в учебном процессе студенты сталкиваются с необходимостью краткого изложения большого объема учебного материала.

Цели конспектирования:

- развитие у обучающегося навыков переработки информации полученной в устном или письменном виде и придание ей сжатой формы;
- выработка умений выделить основную идею, мысль из первоисточника информации;
- формирование навыков составления грамотных, логичных, кратких тезисов;
- облегчение процесса запоминания текста.

Обучающимся следует обратить внимание, на то, что дословная запись текста не является конспектом. Только структурированный тезисный текст может называться таковым. При конспектировании письменных источников необходимо обращать внимание на абзацы, их существование призвано облегчить восприятие текста и начало новой мысли автора. Важно учитывать также и то, что одна мысль может быть изложена в нескольких абзацах.

Высокую скорость конспектирования могут обеспечить сокращения (общепринятые, аббревиатуры, стрелочки, указывающие на логические связи, опорные слова, ключевые слова, схемы и т.д.).

Составление конспекта призвано облегчить запоминание текста. Обучающимся рекомендуется после его составления прочесть зафиксированные тезисы несколько раз для полного их усвоения. Допускается подчеркивание тезисов, содержащих основные мысли, выделение их цветным маркером.



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 4 из 81

Первый экземпляр _____

КОПИЯ № _____

Пересказ не является конспектированием статьи, перефразирование возможно с сохранением ее содержания, логических связей с целью сокращения объема.

При использовании в конспекте цитат автора требуется выделение их в кавычки и указание на источник публикации и автора высказывания.

Виды конспектов:

1 Плановый конспект. Его выполнение возможно при составлении предварительного плана научной статьи, главы учебного или методического пособия. Каждый вопрос плана раскрывается в соответствующей части конспекта

2 Текстуальный конспект. Прекрасно подходит для конспектирования научных статей, монографий, когда необходимо в полной мере отразить точку зрения автора, облачив ее в форму цитаты.

3 Тематический конспект. Его основная задача – найти ответ на поставленный вопрос, решить научную задачу. Число конспектируемых источников в этом случае не ограничивается, имеет значение достижение конечной цели.

4 Свободный конспект может содержать в себе элементы вышперечисленных конспектов. Составление такого конспекта способствует лучшему усвоению и осмыслению материала.

Порядок составления конспекта письменного материала:

1 Предварительное знакомство с текстом с целью определения степени его сложности, научности и выбора вида конспектирования.

2 Повторное прочтение и анализ исследуемого материала позволит сосредоточиться на главной мысли текста и определиться с разбивкой текста на смысловые единицы.

3 Выделение тезисов, передающих краткое содержание материала.

Ясность, краткость, логическая связанность тезисов – основные характеристики конспекта. При конспектировании письменного материала необходимо обращать внимание не только на основные положения текста, но и на доказательства, примеры, цитаты, мнения ученых и практиков.

Указания по конспектированию лекций:

- не нужно стараться записать весь материал, озвученный преподавателем. Как правило, лектором делаются акценты на ключевых моментах лекции для начала конспектирования;

- конспектирование необходимо начинать после оглашением главной мысли лектором, перед началом ее комментирования;

- выделение главных мыслей в конспекте другим цветом целесообразно производить вне лекции с целью сокращения времени на конспектирование на самой лекции;

- применение сокращений приветствуется;

- нужно избегать длинных и сложных рассуждений;

- дословное конспектирование отнимает много времени, поэтому необходимо опускать фразы, имеющие второстепенное значение;

- если в лекции встречаются неизвестные термины, лучше всего отметить на полях их существование, оставить место для их пояснения и в конце лекции задать уточняющий вопрос лектору.

Лекция выступает в качестве ведущего звена всего курса обучения и представляет собой способ изложения объемного теоретического материала, обеспечивающий целостность и



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 5 из 81

Первый экземпляр _____

КОПИЯ № _____

законченность его восприятия студентами. Лекция дает систематизированные основы научных знаний по дисциплине, раскрывает состояние и перспективы развития соответствующей области науки и техники, концентрирует внимание студентов на наиболее сложных, узловых вопросах, стимулирует их активную познавательную деятельность и способствует формированию творческого мышления. Основными функциями лекции выступают познавательная, развивающая, воспитательная и организующая.

Подготовка к занятиям лекционного типа включает в себя: осознание необходимости ее выполнения; целенаправленную познавательно-практическую деятельность непосредственно перед лекцией (просмотр материала предыдущей лекции для восстановления в памяти основных моментов; ознакомление с новой информацией по рекомендуемой учебной литературе для установления связей между изученной и изучаемой информацией; подбор необходимой дополнительной литературы; выполнение заданий, предложенных на самостоятельную проработку).

Самостоятельная работа студентов на уровне занятий лекционного типа заключается в следующем: осознание студентами целей и задач лекции; понимание смысла сообщаемой преподавателем информации; понимание новых знаний; понимание особенностей подходов к понятиям различных авторов, оценивание их достоинств и недостатков; участие в решении поставленных проблем. Внеаудиторная самостоятельная работа студентов после прослушивания лекции заключается в обработке, закреплении и углублении знаний по изученной теме; перечитывании своих конспектов; выяснения непонятных вопросов, знакомство с полученным материалом по рекомендованной учебной литературе, внесение дополнений в конспект; изучение дополнительной литературы.

Рекомендации по оформлению лекций.

Внимательно ознакомьтесь с предлагаемыми ниже рекомендациями по работе с текстами лекций.

Слушание и записывание лекций является одной из решающих форм самообучения студентов. С ней, с этой формой, связана и работа с литературой, и составление планов, тезисов, конспектов, и подготовка к коллоквиуму, зачету, экзамену, к написанию докладов, рефератов, курсовых работ.

Рекомендации по работе над текстами лекции сводятся к конкретным советам, с первого взгляда кажущимся до примитивности простыми.

1 На обороте обложки тетради записывается фамилия, имя, отчество преподавателя, его ученая степень и ученое звание

2 При конспектировании рекомендуется вести записи на одной стороне листа, оставляя вторую сторону для размышлений, разборов, вопросов, ответов на них, для фиксирования деталей темы или связанных с ней фактов, которые припоминаются самим студентом в ходе слушания лекционного материала. Иногда для этого оставляются лишь широкие поля. На первых лекциях стоит специально упражняться в использовании полей: фиксировать вопросы, вызывающие личный интерес, варианты ответов на них, сомнения, проблемы, спорные положения – то есть превратить тетрадные поля в поля размышлений, бесед с собой, диалогов с товарищами, с преподавателем



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 6 из 81

Первый экземпляр _____

КОПИЯ № _____

3 Запись лекций ведется в произвольной форме. Это может быть стиль учебной программы (назывные предложения); можно важнейшие мысли выделять цветными фломастерами или применять боковые «фонарики», выделяющие подтемы. При этом форма записи материала зависит от специфики изучаемого предмета; уместны и свои краткие пояснения к записям. Запись лекции лучше вести в сжатой форме, короткими и четкими фразами.

4 Запись лекции помогает сосредоточить внимание на главном, в ходе самой лекции продумать и осмыслить услышанное, осознать план и логику изложения материала преподавателем. Такая работа нередко, особенно поначалу, вызывает трудности у студентов: некоторые стремятся записывать все дословно, другие пишут отрывочно, у третьих запись получается хаотическая.

Чтобы избежать этих ошибок, целесообразно придерживаться следующих правил:· после записи ориентирующих и направляющих внимание данных (тема, цель, план лекции, рекомендованная литература) важно попытаться проследить, как они раскрываются в содержании, подкрепляются формулировками, доказательствами, а затем и выводами;· записывать следует основные положения и доказывающие их аргументы, наиболее яркие примеры и факты, поставленные преподавателем вопросы для самостоятельной проработки;· стремиться к четкости записи, ее последовательности, выделяя темы и подтемы, вопросы и подвопросы, используя цифровую и буквенную нумерацию (римские и арабские цифры, большие и малые буквы), красные строки, выделение абзацев, подчеркивание главного;· каждому студенту полезно выработать свою систему сокращений, в которой он мог бы разобраться легко и безошибочно;· необходимо выполнять систематические возвращения к предыдущим текстам. Даже отлично записанная лекция предполагает дальнейшую самостоятельную работу над ней (глубокое осмысление ее содержания, логической структуры, выводов). Особенно важно в процессе самостоятельной работы над лекцией выделить новый понятийный аппарат, уяснить суть новых понятий, при необходимости обратиться к словарям и другим источникам, заодно устранив неточности в записях.

Работа над лекцией стимулирует самостоятельный поиск ответов на самые различные вопросы: над какими понятиями следует поработать, какие обобщения сделать, какой дополнительный материал привлечь.

Начните использовать предложенные приемы для написания лекции по дисциплине с сегодняшнего дня. Через некоторое время (примерно через месяц) ответьте на вопросы: – помогает ли предложенная методика вам в написании лекций, обоснуйте свой ответ; – какие недостатки предложенной методики вы выявили; – подкорректируйте методику с учетом ваших замечаний.

При подготовке и усвоении материала занятий лекционного типа следует обращаться к рекомендациям преподавателя, а также осуществлять самостоятельный подбор необходимых источников.

Рекомендуемая литература приводится в рабочей программе соответствующей дисциплины и должна быть использована только из библиотеки Костанайского филиала ФГБОУ «ЧелГУ» и электронных библиотечных систем.

Студентам рекомендуется получить в библиотеке учебную литературу по дисциплине, а также использовать электронные системы, имеющиеся в доступе на сайте филиала.



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 7 из 81

Первый экземпляр _____

КОПИЯ № _____

При работе с литературой следует учитывать, что в большинстве учебников и книг делаются ссылки, сноски на другие литературные источники, приводится список литературы по раскрываемой в источнике проблеме.

Однако найти нужную статью или книгу – это только одна проблема профессиональной подготовки студента, другая заключается в том, чтобы суметь быстро прочитать и осмыслить содержащуюся в них информацию.

Работа над литературой состоит из трёх этапов – чтения работы, её конспектирования, заключительного обобщения сути изучаемой работы. Прежде, чем браться за конспектирование следует её хотя бы однажды прочитать, чтобы составить о ней предварительное мнение, постараться выделить основную мысль или несколько базовых точек, опираясь на которые можно будет в дальнейшем работать с текстом. Конспектирование в общем виде может быть определено как фиксация основных положений и отличительных черт рассматриваемого труда.

Если программа занятия предусматривает работу с электронным источником, то этой стороне подготовки к занятию следует уделить пристальное внимание. В сущности, разбор электронного источника не отличается от работы с литературой – то же чтение, конспектирование, обобщение. Тщательная подготовка к лекциям, имеет определяющее значение: занятие пройдёт так, как аудитория подготовилась к его проведению.

В случае возникновения неясных вопросов студент может получить индивидуальную консультацию преподавателя согласно графику консультаций, утвержденному на кафедре.

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/

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



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 8 из 81

Первый экземпляр _____

КОПИЯ № _____

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.

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ПОДГОТОВКЕ И К ВЫПОЛНЕНИЮ ПРАКТИЧЕСКИХ РАБОТ

Практические занятия направлены на расширение и детализацию знаний, на выработку и закрепление навыков профессиональной деятельности. Подготовка к практическим занятиям предполагает предварительную самостоятельную работу студентов в соответствии с методическими разработками по каждой запланированной теме.

Практические занятия позволяют интегрировать теоретические знания и формировать практические умения и навыки студентов в процессе учебной деятельности.

Цели практических занятий:

1. Закрепление теоретического материала путем систематического контроля за самостоятельной работой студентов;
2. Формирование умений использования теоретических знаний в процессе выполнения практических работ;
3. Развитие аналитического мышления путем обобщения результатов практических работ;
4. Формирование навыков оформления результатов практических работ в виде таблиц, графиков, выводов.

Структура и последовательность занятий: на первом, вводном, занятии проводится инструктаж студентов по охране труда, технике безопасности и правилам работы в кабинете ВТ по инструкциям с фиксацией результатов в журнале инструктажа. Студенты также знакомятся с основными требованиями преподавателя по выполнению учебного плана, с расписанием практических занятий, с графиком прохождения рубежных контролей, с основными формами отчетности по выполненным работам и заданиям.

Студентам для выполнения практических работ необходимо индивидуальное рабочее место и компьютер.

Структура практического занятия

1. В начале занятия называется его тема, цель и этапы проведения.
2. По теме занятия проводится беседа, что необходимо для осознанного выполнения практической работы.
3. Практическая работа выполняется в соответствии с методическими указаниями.
4. Перед уходом из компьютерного класса должны навести порядок на своем рабочем месте столе.

Практические занятия – вид самостоятельной практической работы обучающихся (индивидуальной и в группах), направленной на закрепление и углубление, практическое



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 9 из 81

Первый экземпляр _____

КОПИЯ № _____

подтверждение теоретических концепций курса (дисциплины), а также на формирование и развитие умений и навыков планирования и проведения эксперимента.

Раздел 1. Системные задачи

Практическая работа 1

Анализ признаков и оценка их информативности

Продолжительность работы – 8 час.

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

Программа работы

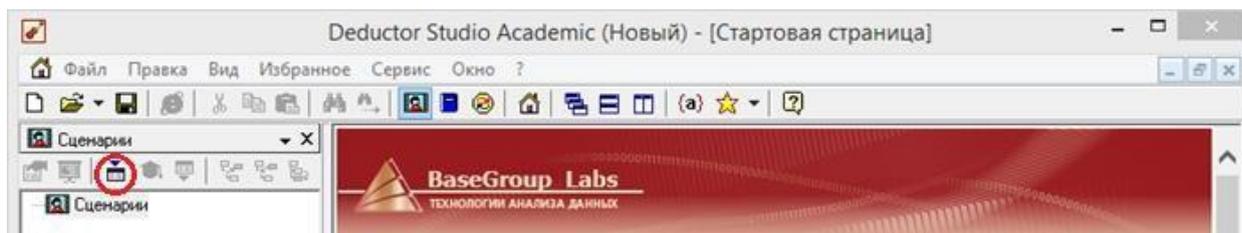
1. Выполнить импорт данных в программный комплекс *Deductor*.
2. Выполнить задание по предварительной парциальной обработке данных.
3. Выполнить задание по предварительной обработке путем удаления аномалий в данных.
4. Выполнить задание по предварительной обработке путем сглаживания данных методом спектральной 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).



| | | | |
|--|---------------|------------------------|---------------|
|  МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ») Костанайский филиал Кафедра социально-гуманитарных и естественнонаучных дисциплин | | | |
| Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент | | | |
| Версия документа - 1 | стр. 10 из 81 | Первый экземпляр _____ | КОПИЯ № _____ |

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.

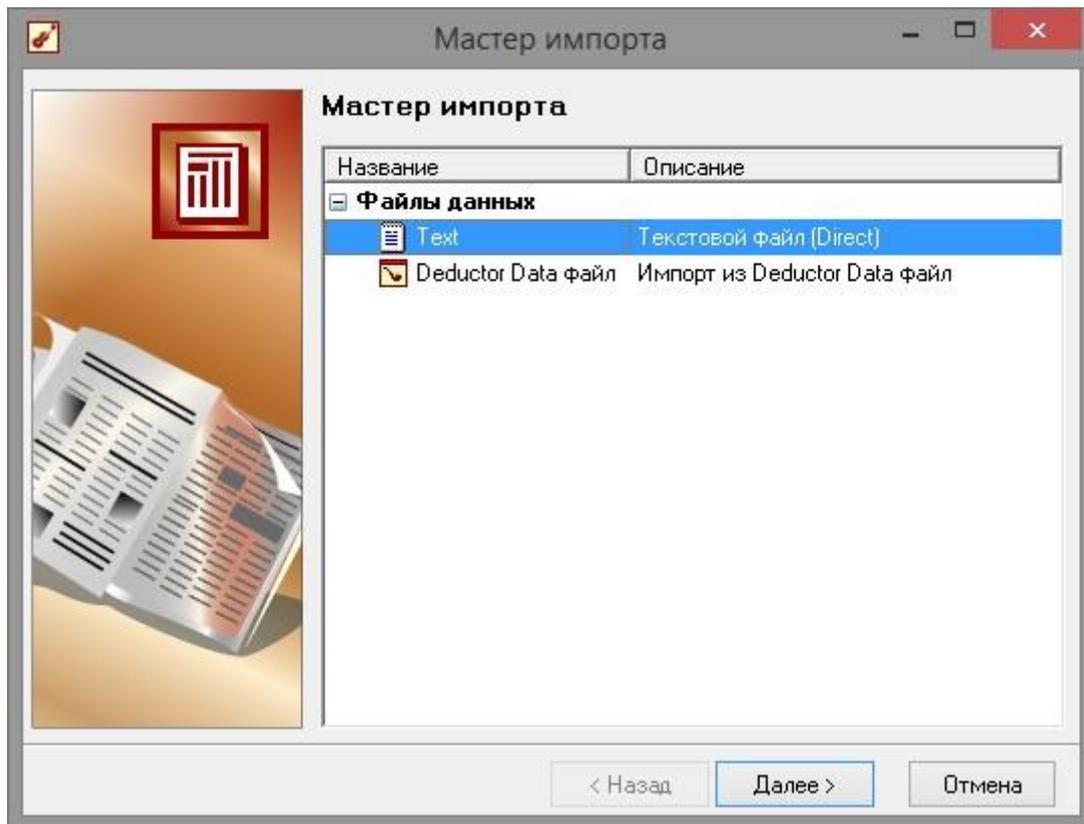


Рис. 1.2 - Мастер импорта

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

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 11 из 81

Первый экземпляр _____

КОПИЯ № _____

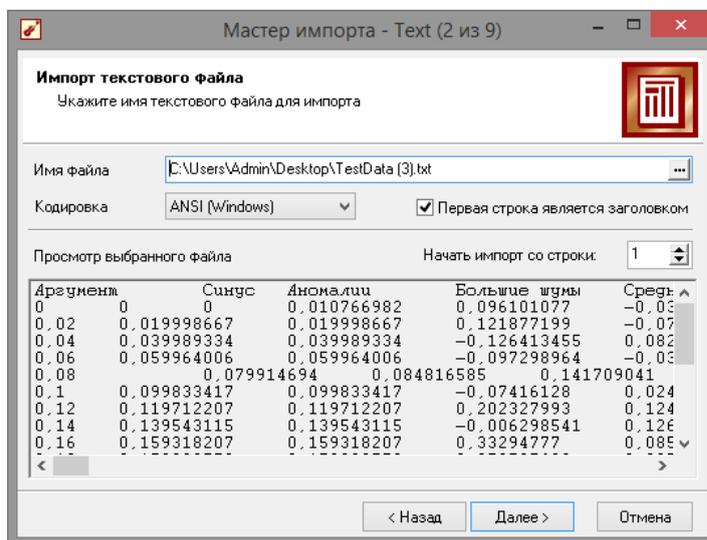


Рис. 1.3 - Выбор файла

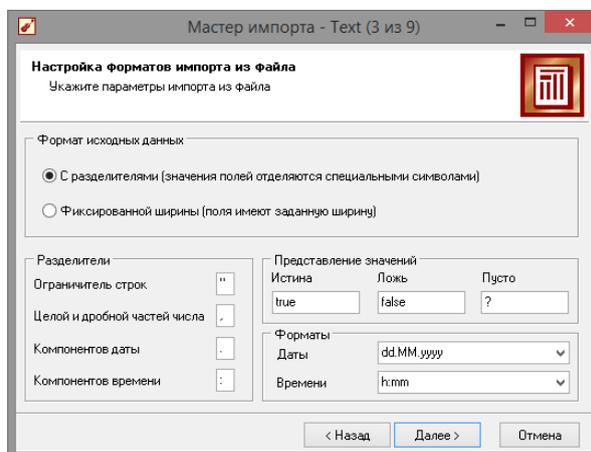


Рис. 1.4 - Параметры импорта

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

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 12 из 81

Первый экземпляр _____

КОПИЯ № _____

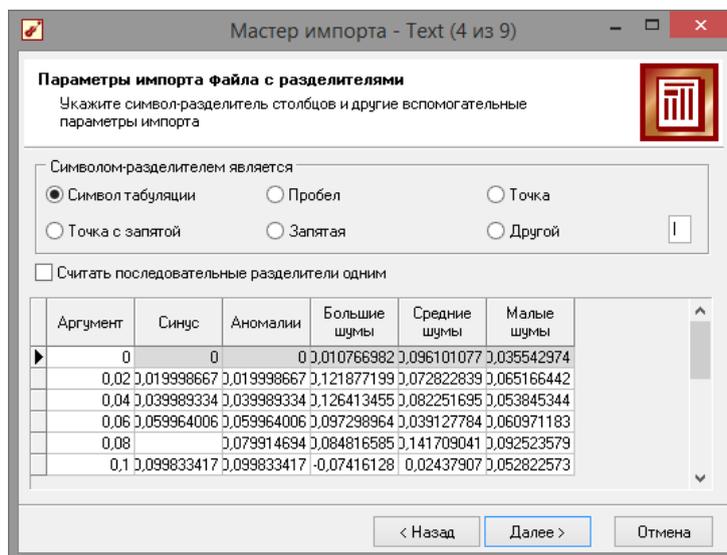


Рис. 1.5 - Параметры разделителей

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.



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 13 из 81

Первый экземпляр _____

КОПИЯ № _____

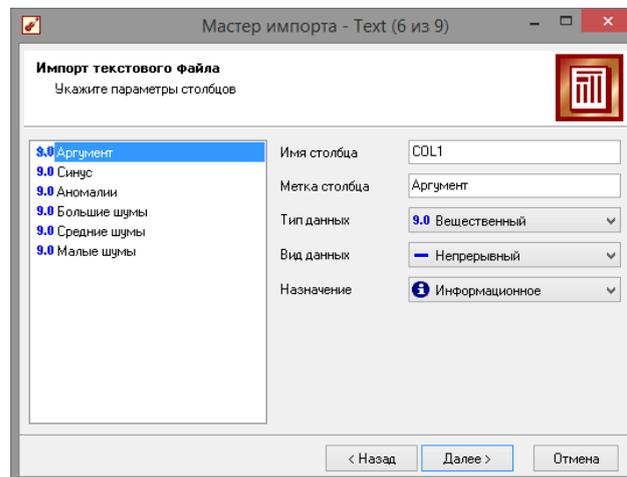


Рис. 1.6 - Параметры столбцов

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.

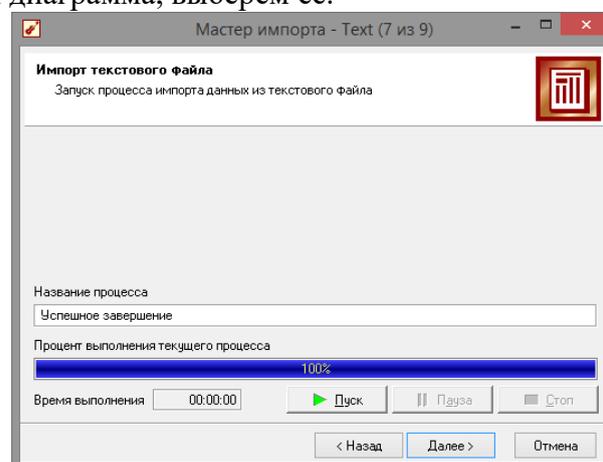


Рис. 1.7 - Импорт файла



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 14 из 81

Первый экземпляр _____

КОПИЯ № _____

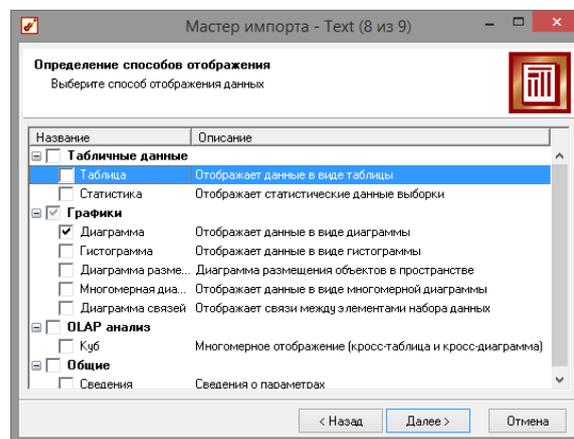


Рис. 1.8 - Способ отображения

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).

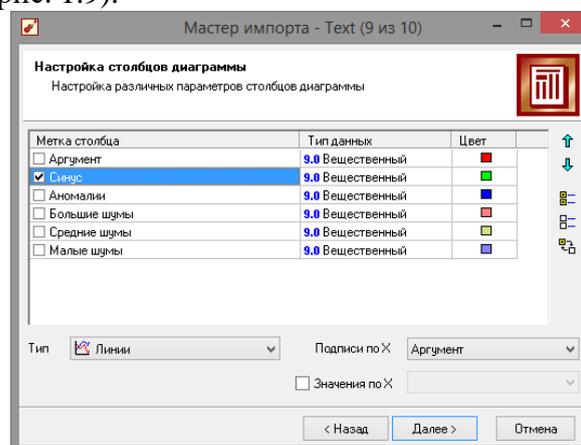


Рис. 1.9 - Настройка столбцов

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



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 15 из 81

Первый экземпляр _____

КОПИЯ № _____

«З-х мерный вид» в левом верхнем углу панели «Диаграмма». А для просмотра другой диаграммы, нажать на значок лупы «Отображать поля» (рис. 1.11).

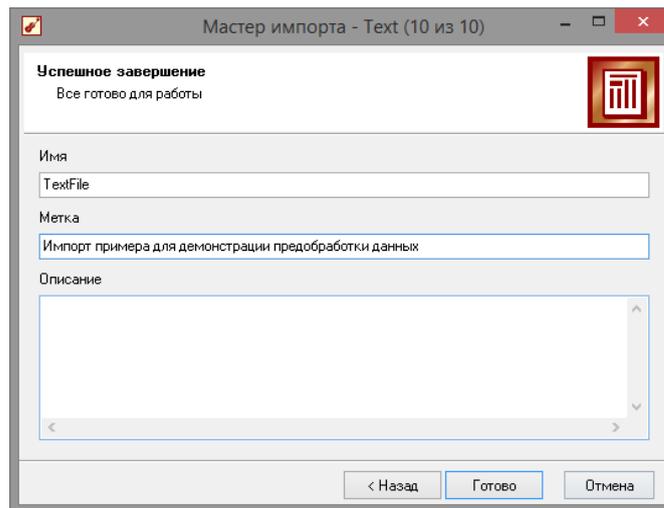


Рис. 1.10 – Завершение импорта

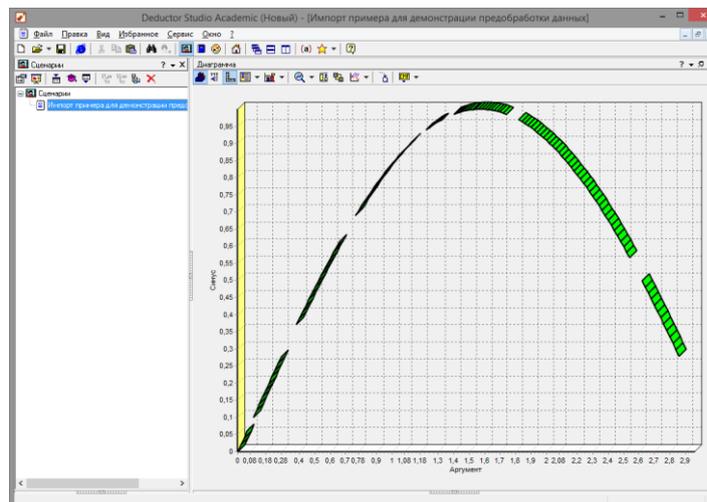


Рис. 1.11 - Диаграмма синуса

1.1 Предварительная парциальная обработка



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 16 из 81

Первый экземпляр _____

КОПИЯ № _____

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.



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 17 из 81

Первый экземпляр _____

КОПИЯ № _____

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).

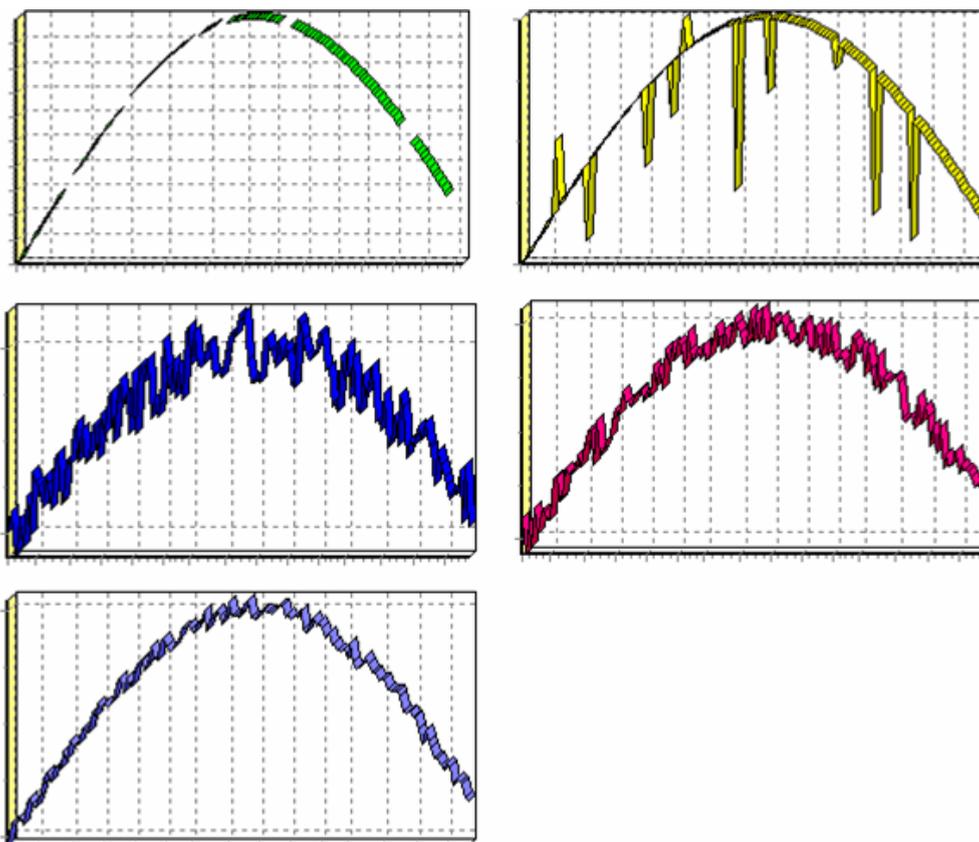


Рисунок 1.12 - Варианты столбцов



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 18 из 81

Первый экземпляр _____

КОПИЯ № _____

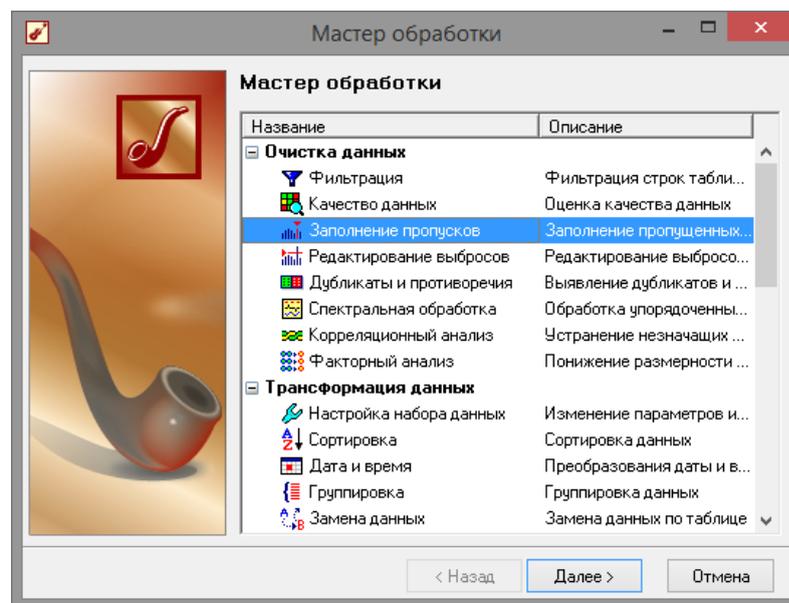


Рисунок 1.13 - Мастер обработки

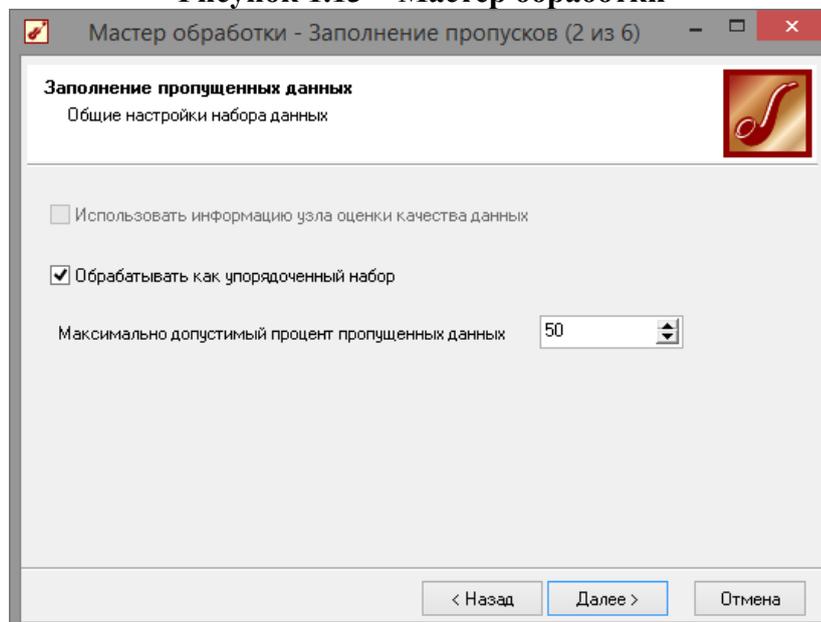


Рис. 1.14 - Мастер заполнения пропусков



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 19 из 81

Первый экземпляр _____

КОПИЯ № _____

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).

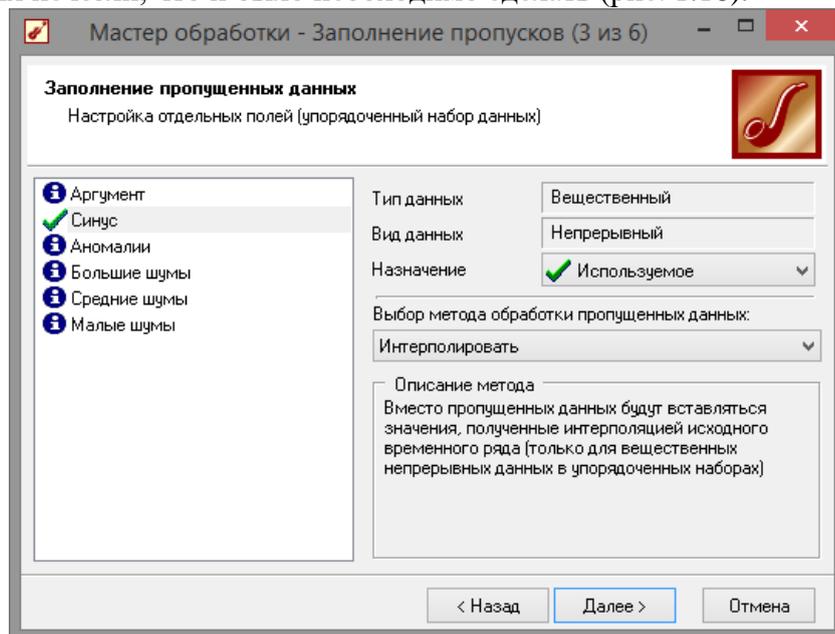


Рис. 1.15 - Настройки мастера заполнения пропусков

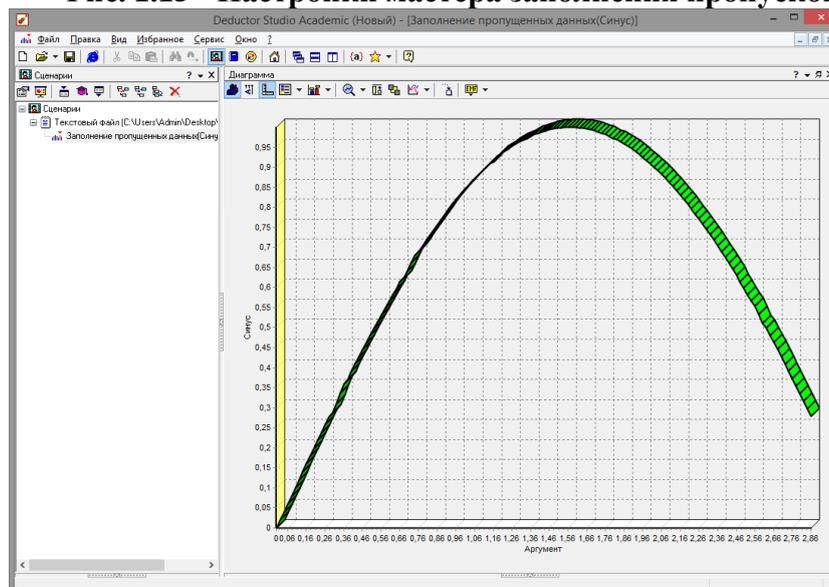


Рис.1.16 - Заполнение пропусков методом интерполяции



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).

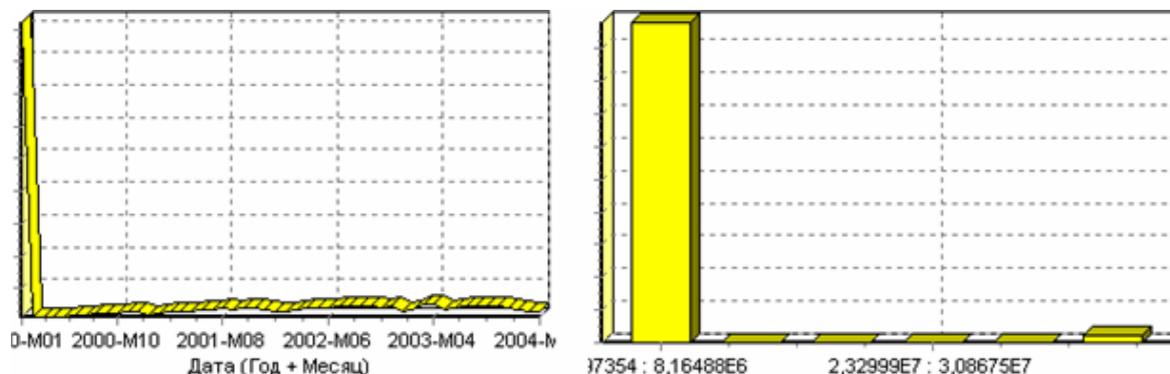


Рис. 1.17 - Гистограмма с аномалиями

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

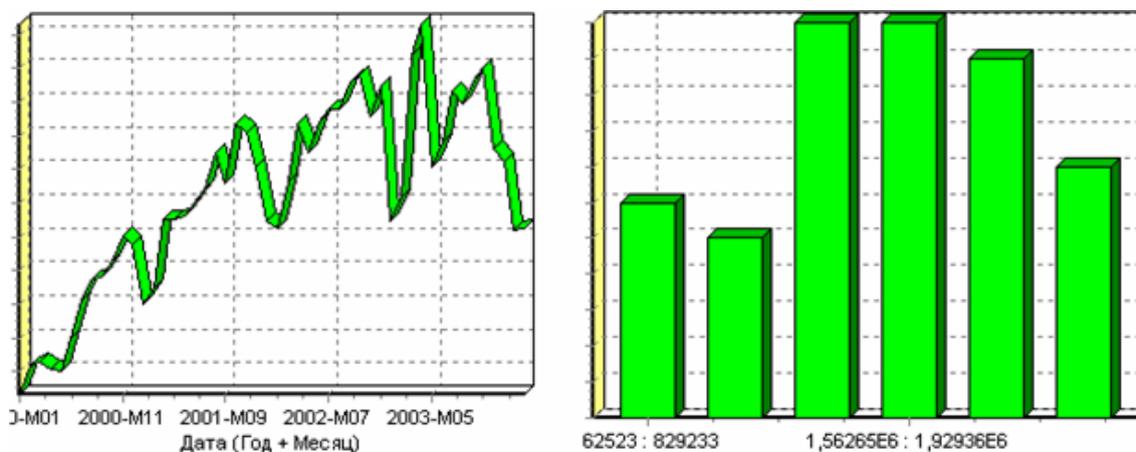


Рис. 1.18 - Гистограмма без аномалий



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 21 из 81

Первый экземпляр _____

КОПИЯ № _____

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

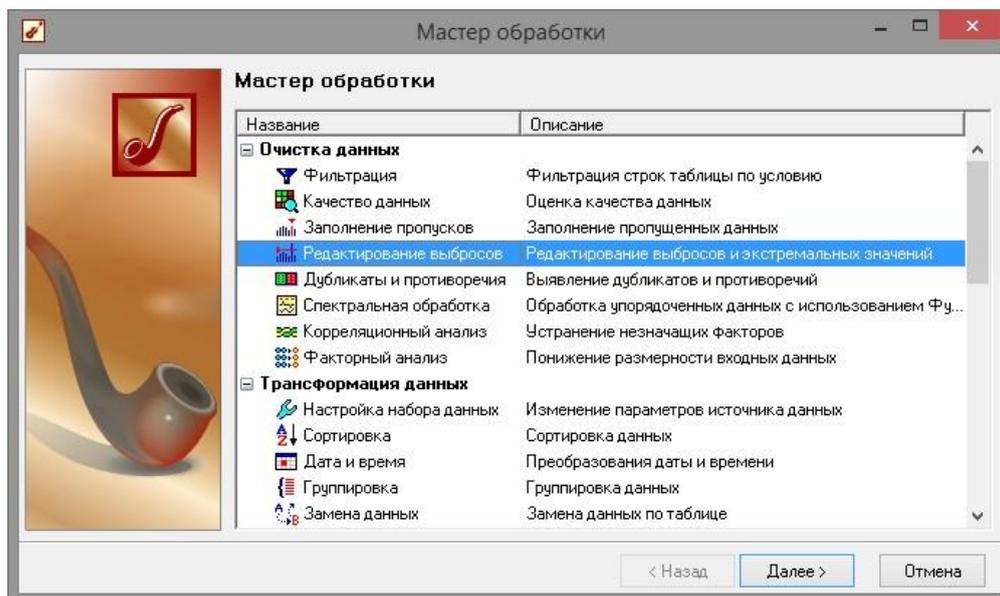


Рис. 1.19 - Мастер редактирования выбросов



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 22 из 81

Первый экземпляр _____

КОПИЯ № _____

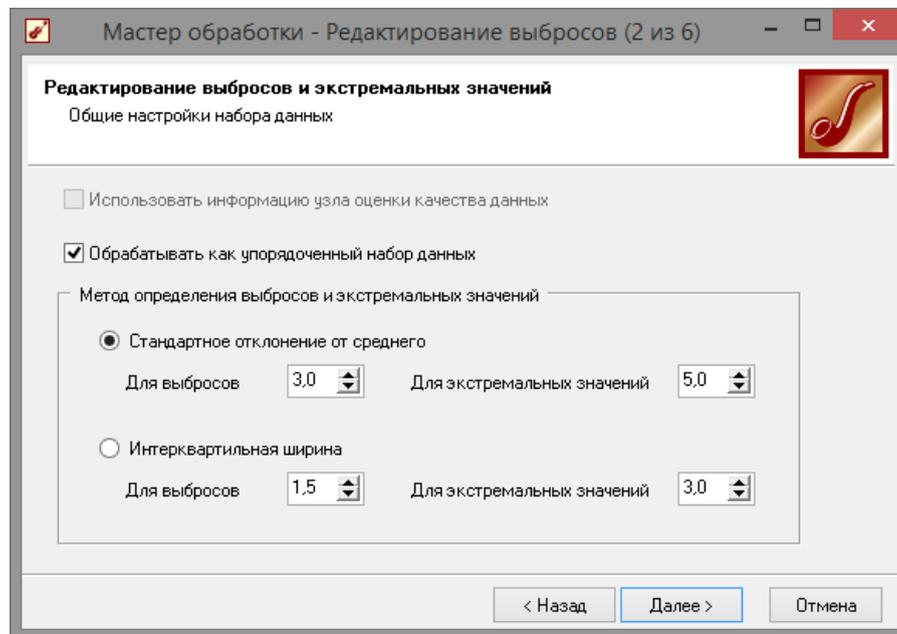


Рис. 1.20 - Настройки мастера редактирования выбросов

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*.



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 23 из 81

Первый экземпляр _____

КОПИЯ № _____

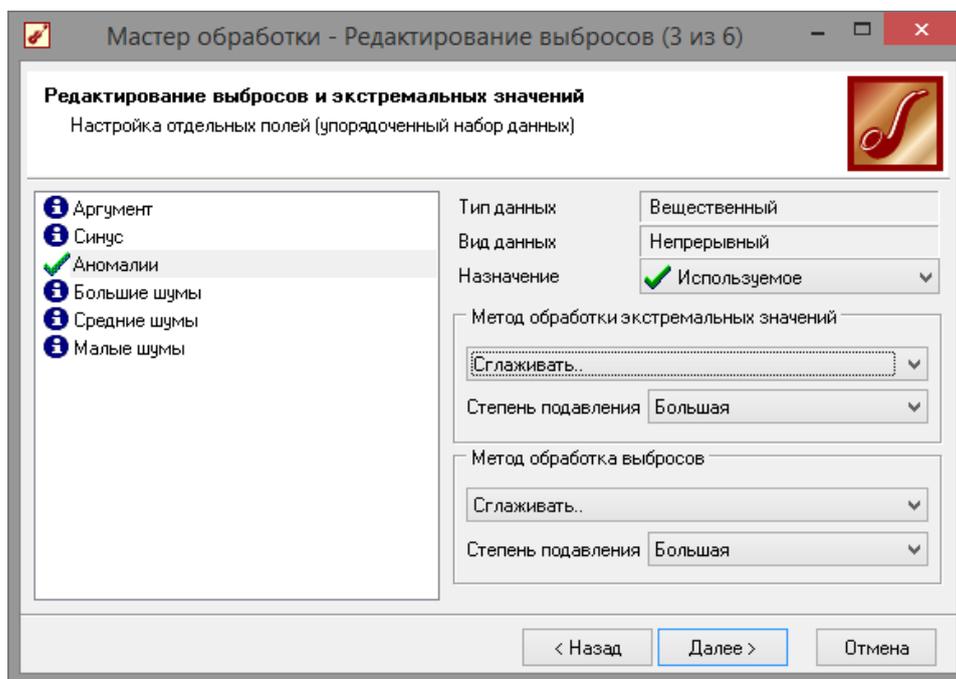


Рис. 1.21 - Настройки отдельных полей мастера редактирования выбросов

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).



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 24 из 81

Первый экземпляр _____

КОПИЯ № _____

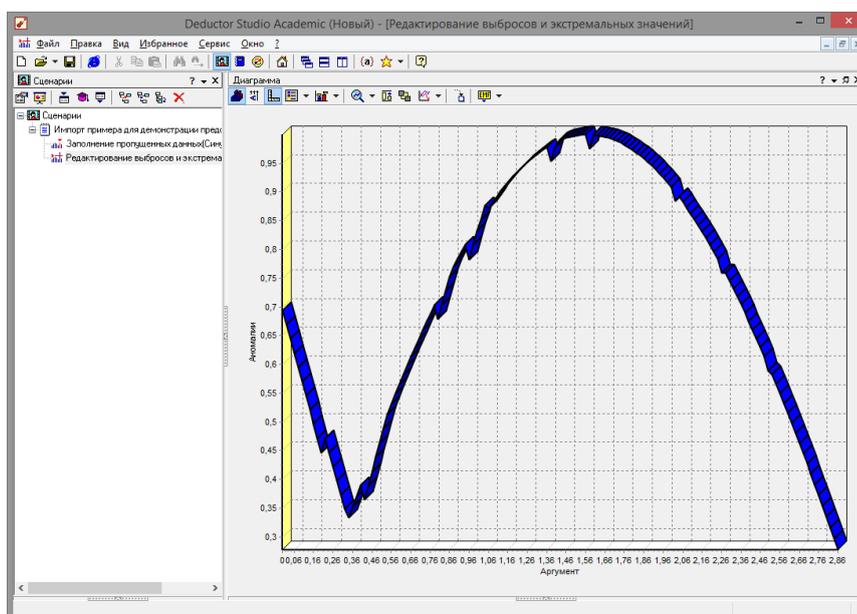


Рис. 1.22 - Диаграмма после удаления аномалий

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

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 25 из 81

Первый экземпляр _____

КОПИЯ № _____

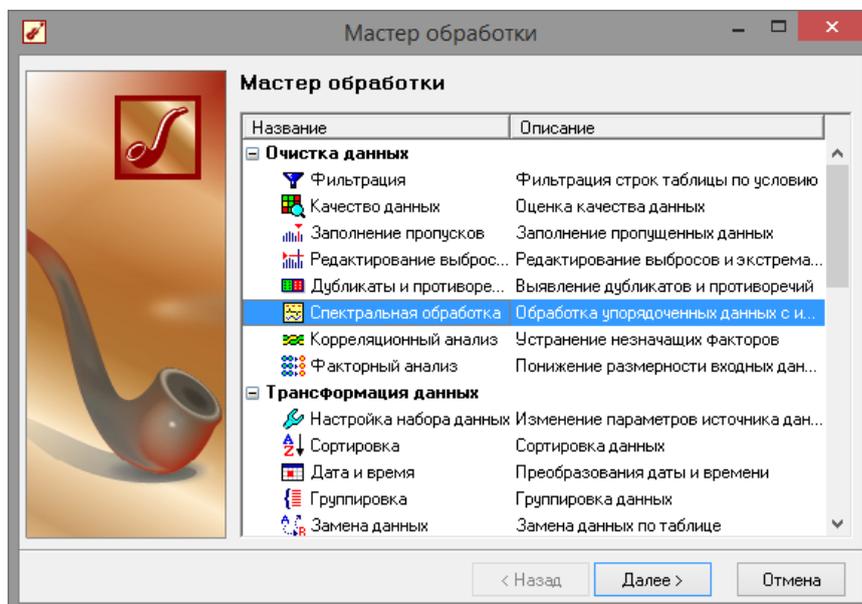


Рис. 1.23 - Мастер обработки

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



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 26 из 81

Первый экземпляр _____

КОПИЯ № _____

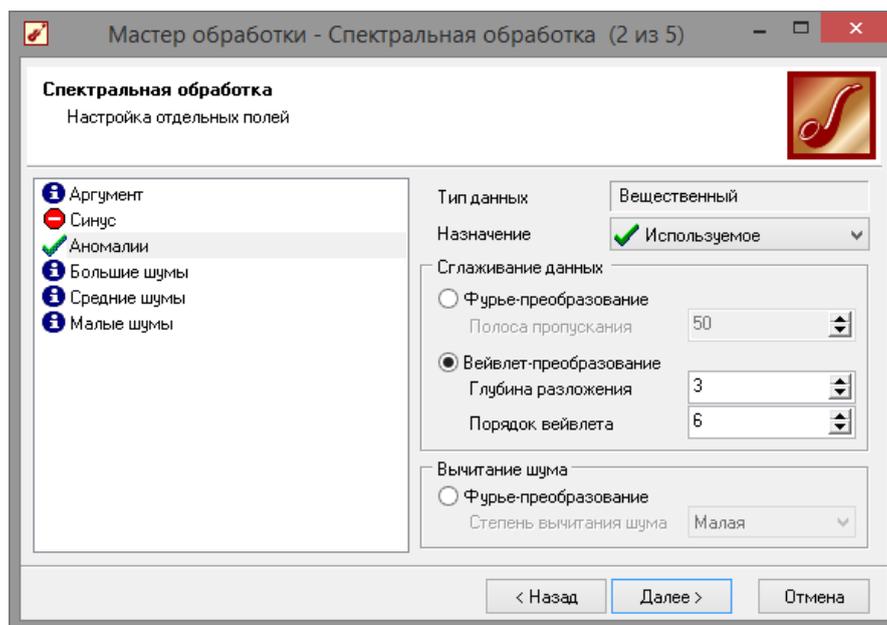


Рис. 1.24 - Мастер спектральной обработки

Далее выбрать визуализацию «**Диаграмма**» и столбец

«**Аномалии**». После обработки можно убедиться на диаграмме в отсутствии выбросов (рис. 1.25).

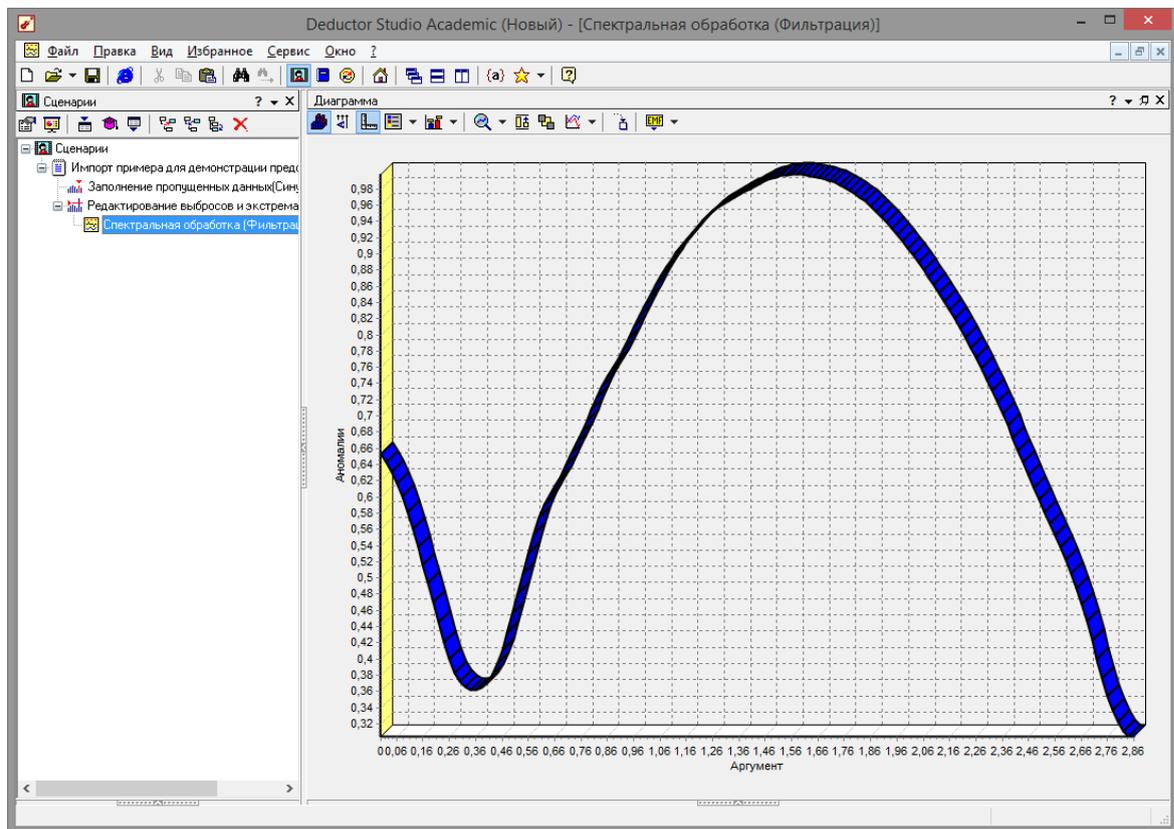


Рис. 1.25 - Диаграмма после применения спектральной обработки

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.



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 28 из 81

Первый экземпляр _____

КОПИЯ № _____

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 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).

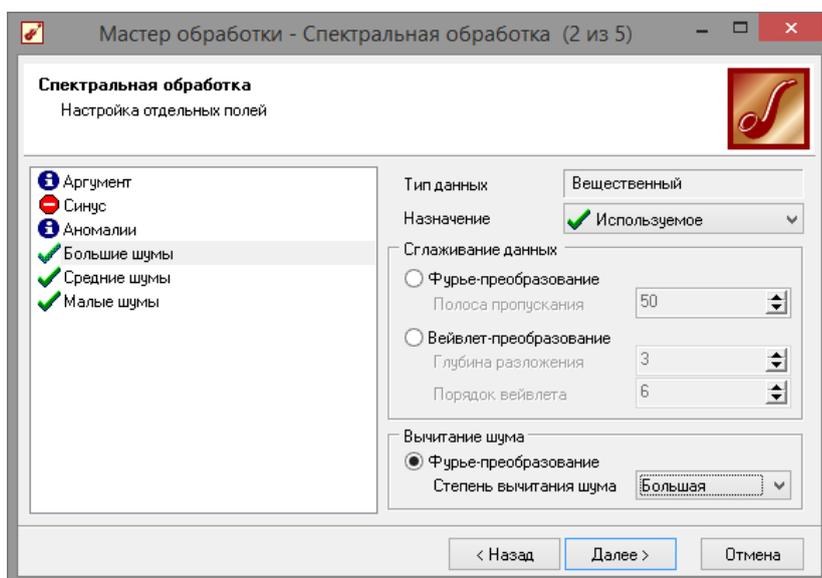


Рис. 1.26 - Настройки мастера спектральной обработки



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 29 из 81

Первый экземпляр _____

КОПИЯ № _____

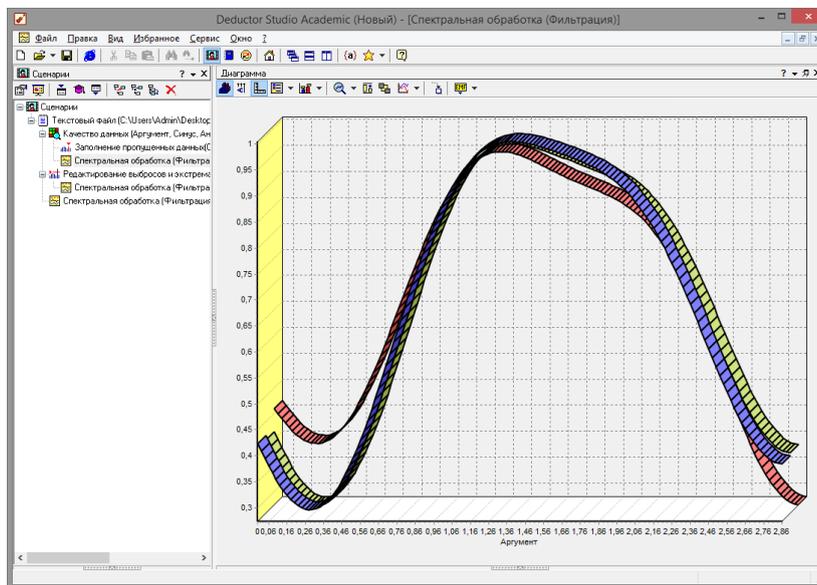


Рис 1.27 - Диаграмма после применения спектральной обработки

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

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

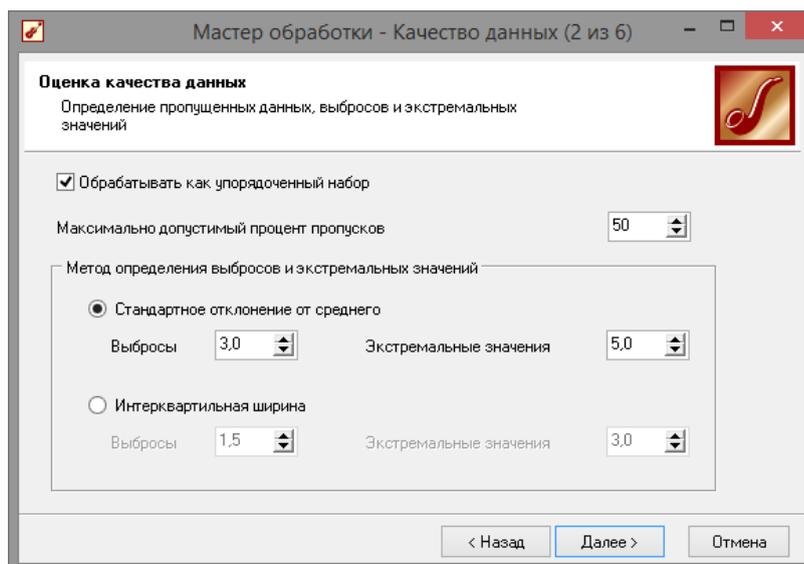


Рис. 1.28 - Мастер качества данных



Версия документа - 1

стр. 30 из 81

Первый экземпляр _____

КОПИЯ № _____

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 | | Колво уникальн. | Качество данных | Рекоменд. |
|---|--------------|---------------|------------|----------|----------|---------|--------------|---------------|------------|-----------------|-----------------|-----------|
| | | | | Колво | Действие | Колво | Действие | Колво | Действие | | | |
| 1 | Аргумент | 9.0 Вещест... | Непер... | | | 21 | Игнорировать | 6 | Сглаживает | | 1.0000 | Пригоден |
| 2 | Синус | 9.0 Вещест... | Непер... | | | | | 5 | Сглаживает | | 0.8963 | Пригоден |
| 3 | Аномалии | 9.0 Вещест... | Непер... | | | | | 5 | Сглаживает | | 0.3120 | Пригоден |
| 4 | Большие ш... | 9.0 Вещест... | Непер... | | | | | 5 | Сглаживает | | 0.3408 | Пригоден |
| 5 | Средние ш... | 9.0 Вещест... | Непер... | | | | | 7 | Сглаживает | | 0.3076 | Пригоден |
| 6 | Малые ш... | 9.0 Вещест... | Непер... | | | | | 6 | Сглаживает | | 0.3206 | Пригоден |

Рис. 1.29 - Результаты мастера качества до обработки данных

| № | Столбец | Тип данных | Вид данных | Пропуски | | Выбросы | | Экстремальные | | Колво уникальн. | Качество данных | Рекоменд. |
|---|--------------|---------------|------------|----------|----------|---------|----------|---------------|----------|-----------------|-----------------|-----------|
| | | | | Колво | Действие | Колво | Действие | Колво | 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... | | | | | | | | 0.3120 | Пригоден |
| 4 | Большие ш... | 9.0 Вещест... | Непер... | | | | | | | | 0.3408 | Пригоден |
| 5 | Средние ш... | 9.0 Вещест... | Непер... | | | | | | | | 0.3076 | Пригоден |
| 6 | Малые ш... | 9.0 Вещест... | Непер... | | | | | | | | 0.3206 | Пригоден |

Рис. 1.30 - Вывод мастера после всех внесенных нами изменений



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 31 из 81

Первый экземпляр _____

КОПИЯ № _____

«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.



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 32 из 81

Первый экземпляр _____

КОПИЯ № _____

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



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 33 из 81

Первый экземпляр _____

КОПИЯ № _____

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 четыре характеристики (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.

| 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).



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 34 из 81

Первый экземпляр _____

КОПИЯ № _____

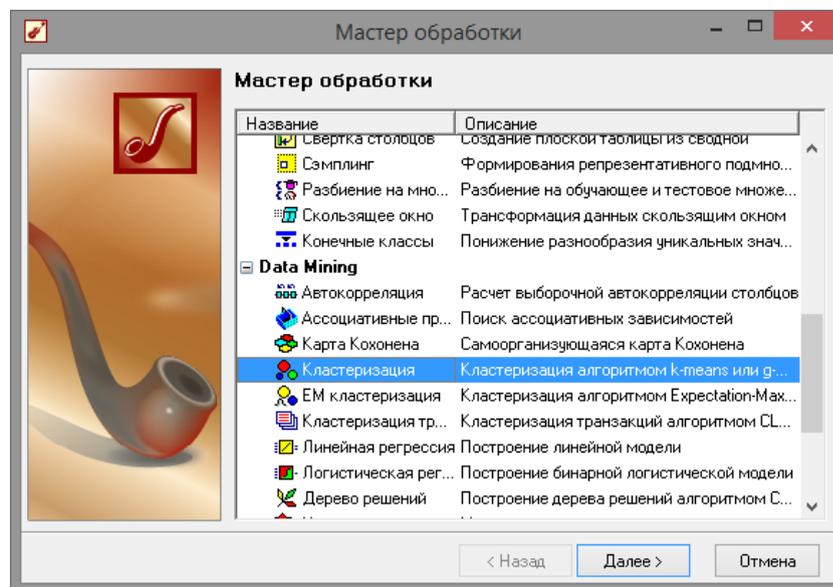


Рис. 2.2 - Мастер кластеризации

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

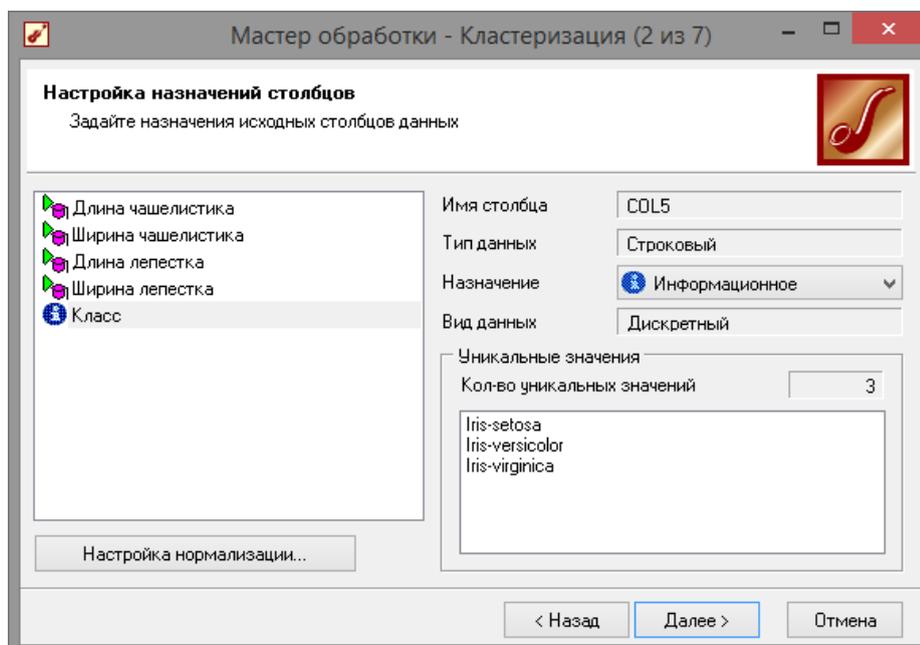


Рис. 2.3 - Настройка мастера кластеризации



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 35 из 81

Первый экземпляр _____

КОПИЯ № _____

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

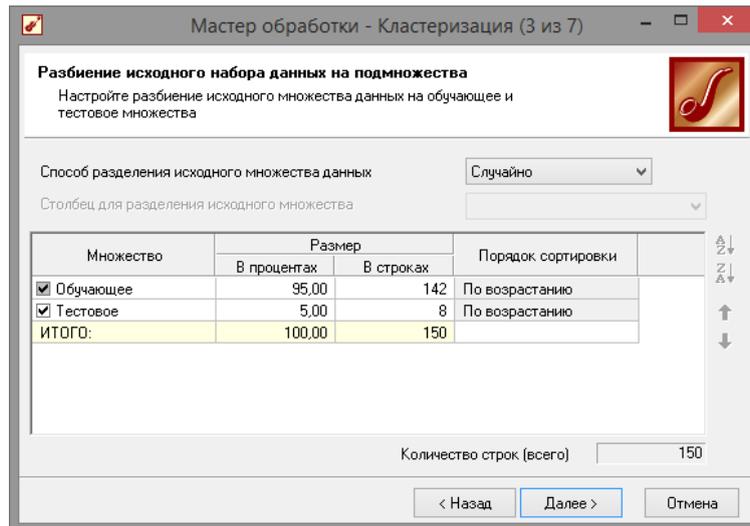


Рис. 2.4 - Настройка параметров обучения

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).

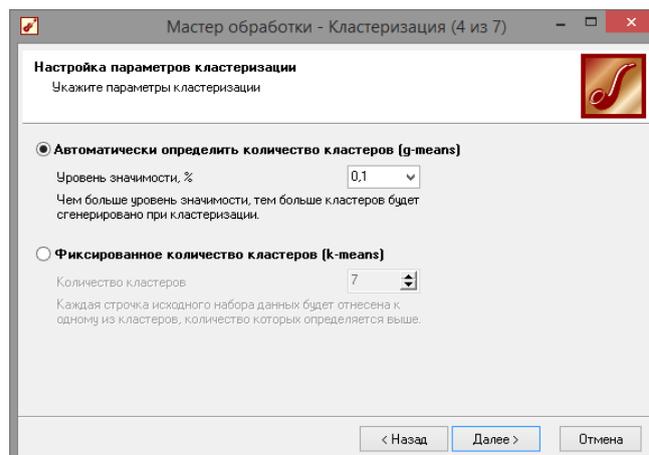


Рис. 2.5 - Выбор алгоритма кластеризации



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 36 из 81

Первый экземпляр _____

КОПИЯ № _____

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).

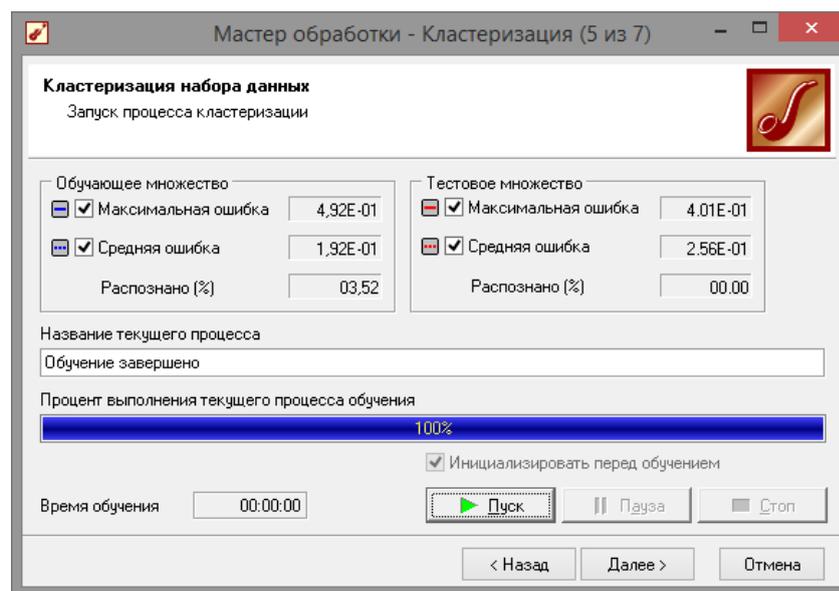


Рис. 2.6 - Обучение сети

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

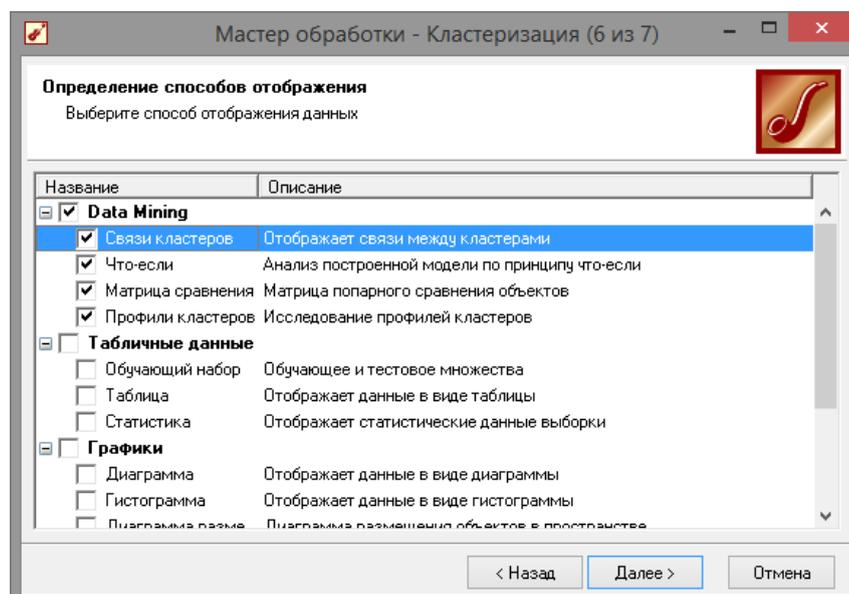




Рис. 2.7 - Визуализация данных

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

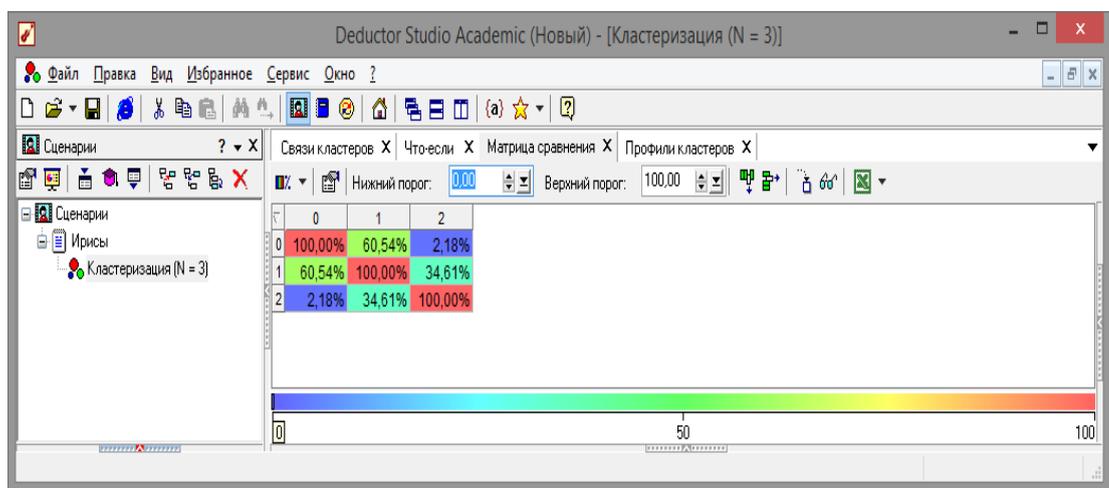


Рис. 2.8 - Профили кластеров

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

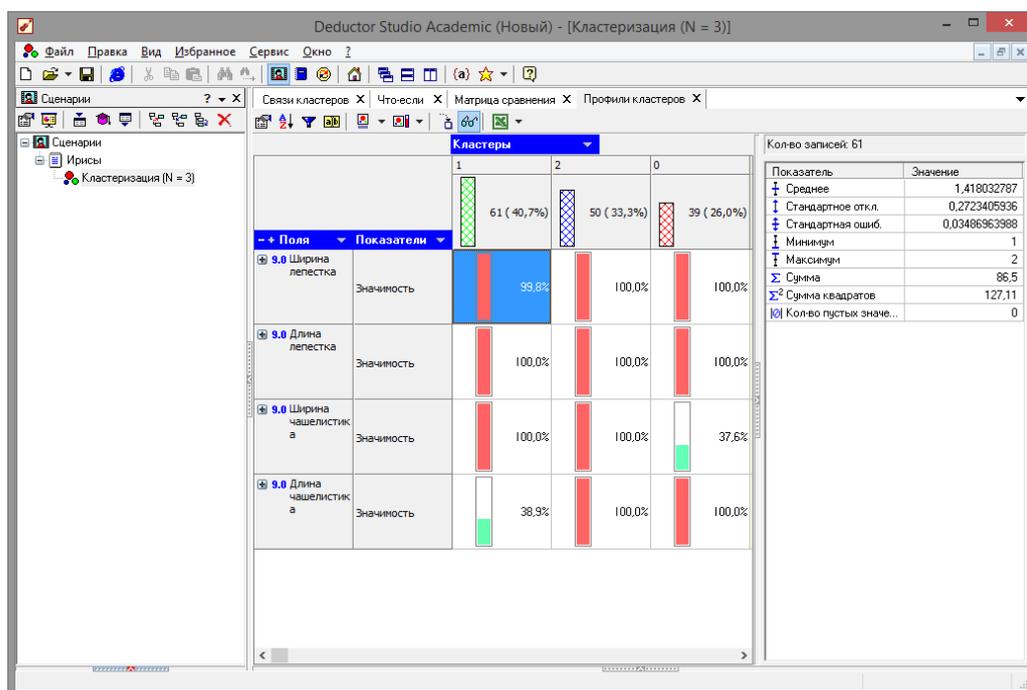




Рис. 2.9 - Матрица сравнения

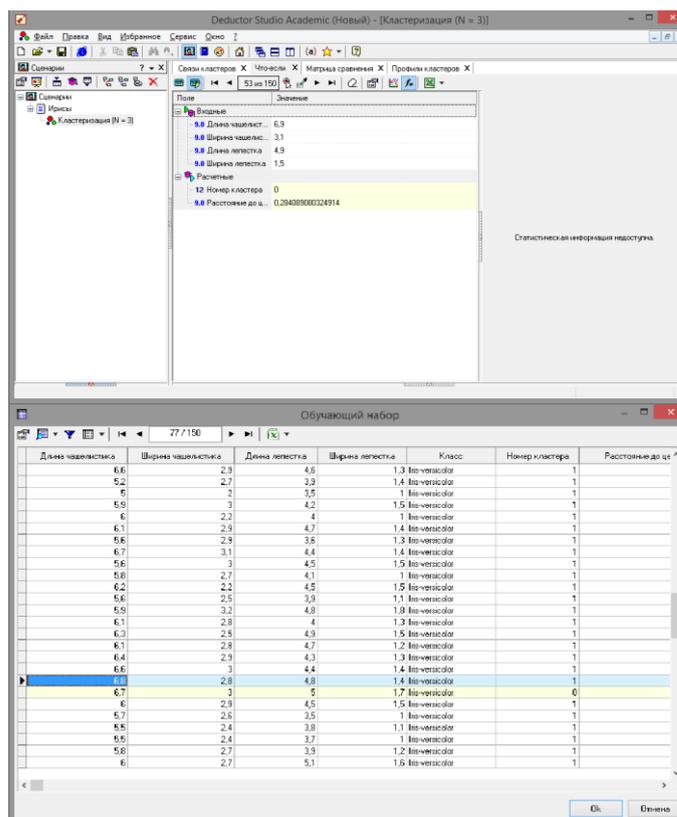


Рис. 2.10 - Инструмент «Что-Если»

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 опробуем 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).



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 39 из 81

Первый экземпляр _____

КОПИЯ № _____

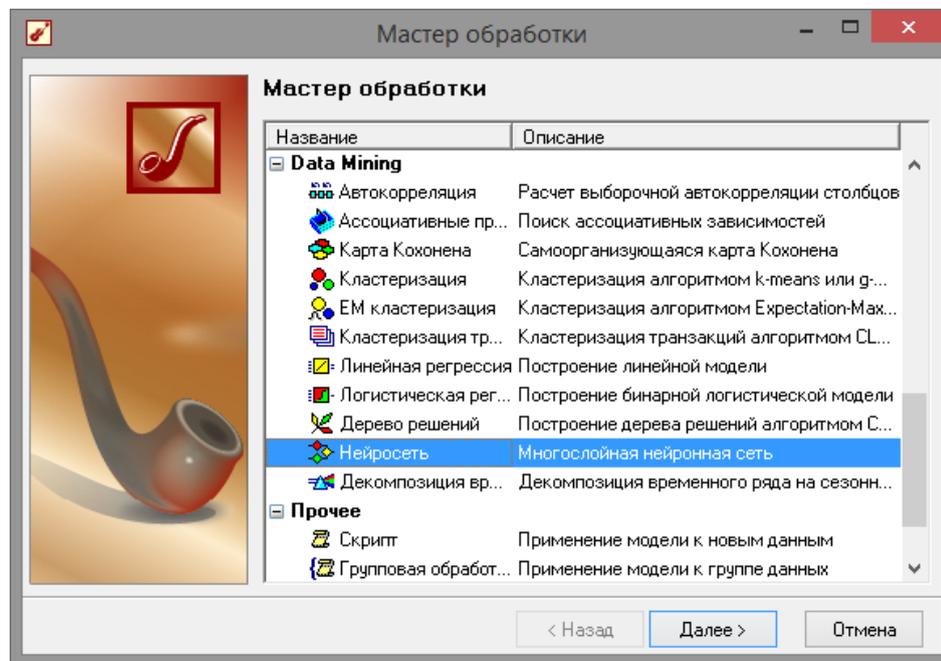


Рис. 2.11 - Мастер обработок

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

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 40 из 81

Первый экземпляр _____

КОПИЯ № _____

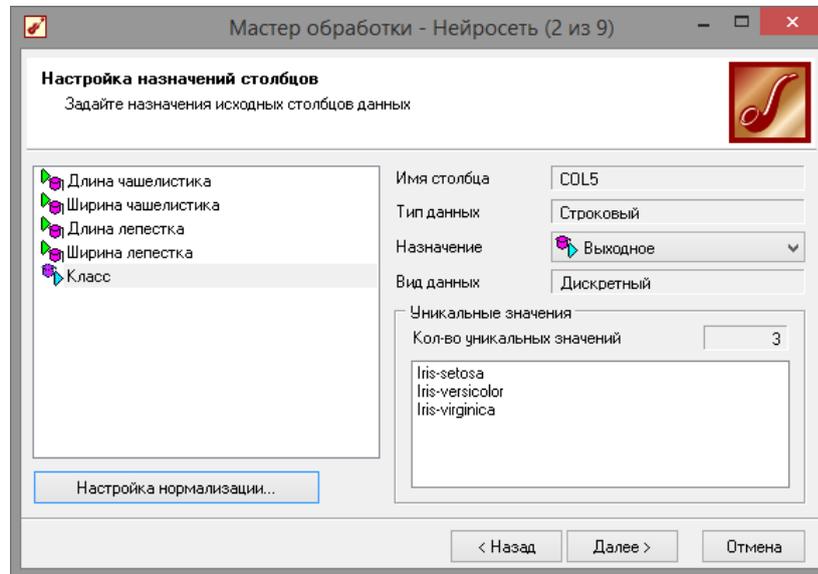


Рис. 2.12 - Назначение входов и выходов

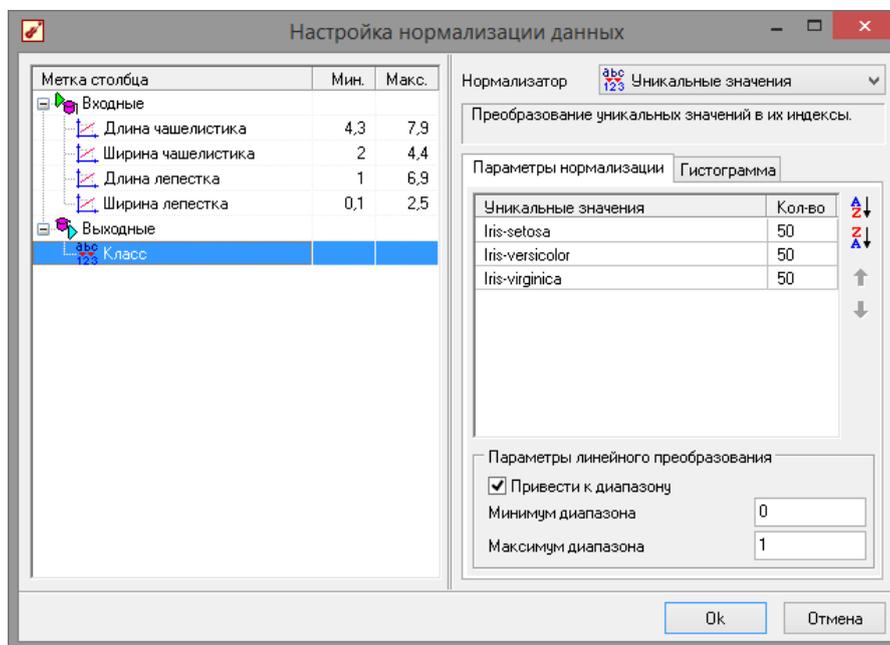


Рис. 2.13 - Настройка нормализации

Следующим шагом указываем параметры обучения (рис. 2.14).



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 41 из 81

Первый экземпляр _____

КОПИЯ № _____

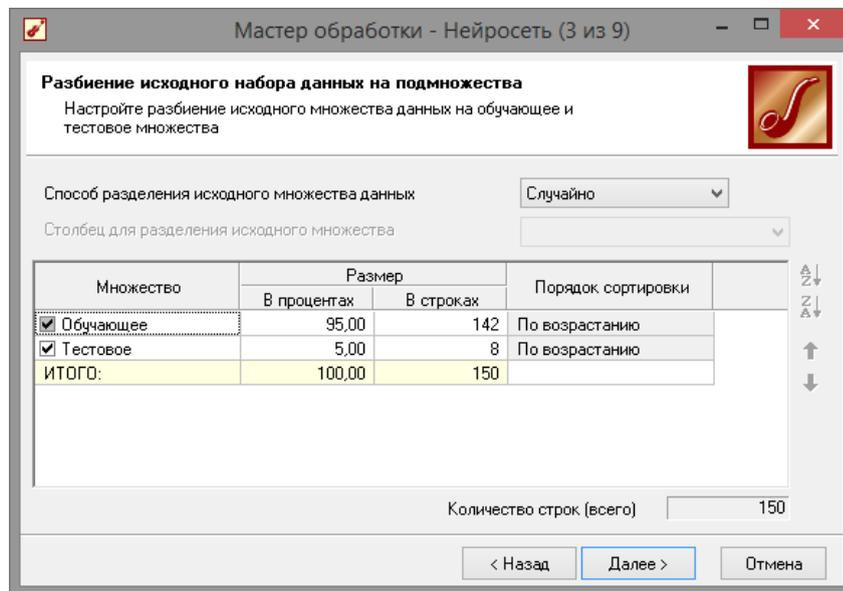


Рис. 2.14 - Параметры обучения

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

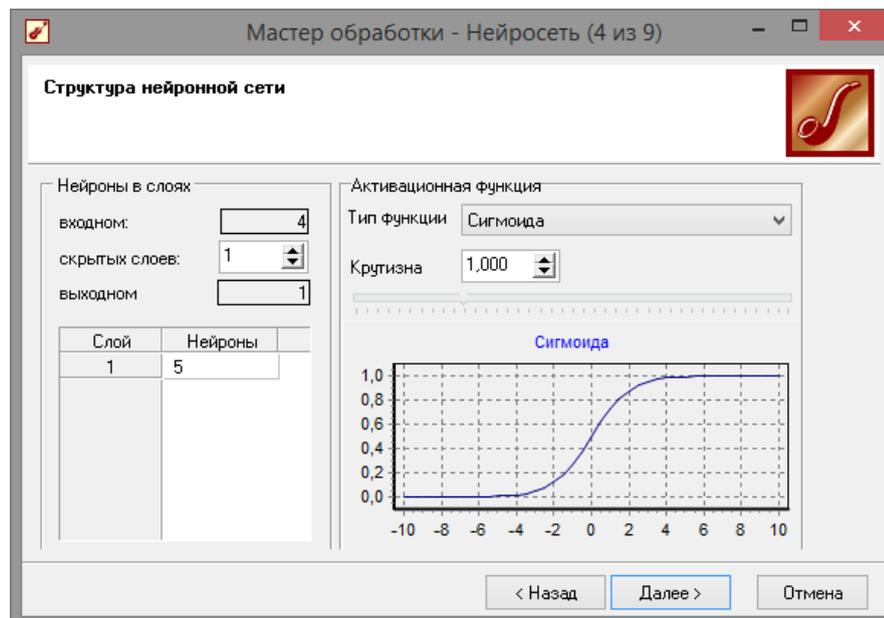


Рис. 2.15 - Параметры обучения

| | | | |
|--|---------------|------------------------|---------------|
|  МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ») Костанайский филиал Кафедра социально-гуманитарных и естественнонаучных дисциплин | | | |
| Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент | | | |
| Версия документа - 1 | стр. 42 из 81 | Первый экземпляр _____ | КОПИЯ № _____ |

Далее выбираем алгоритм обучения (рис. 2.16).

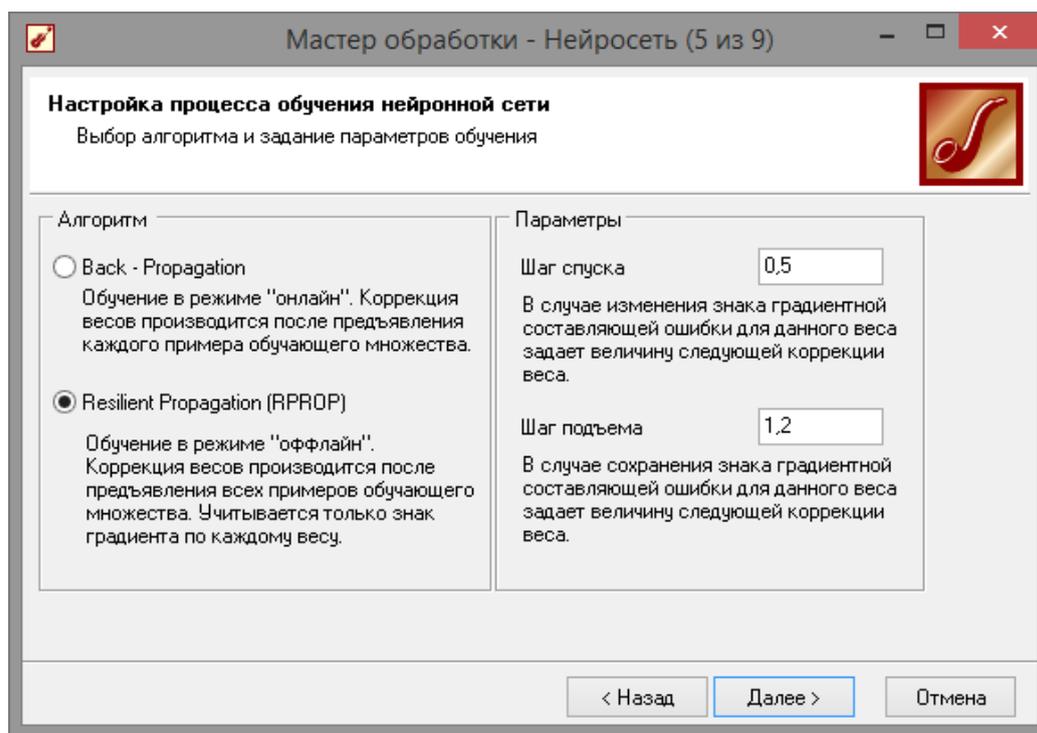
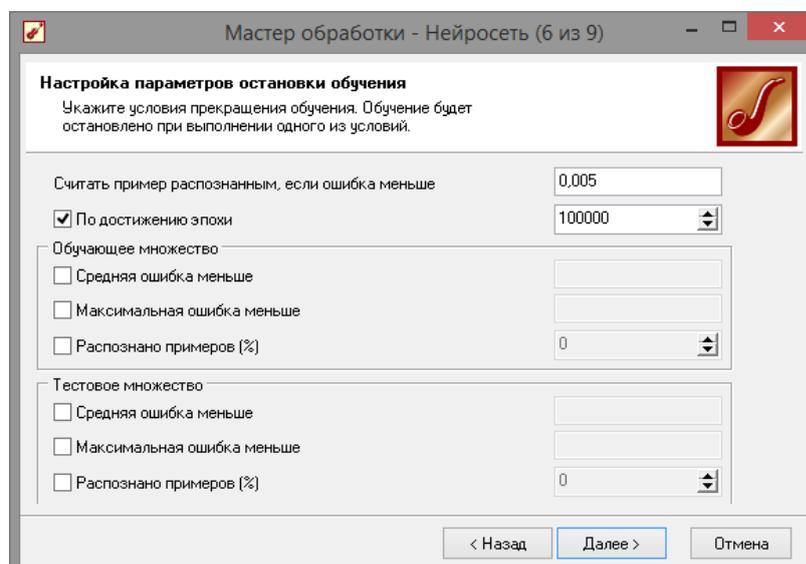


Рис. 2.16 - Алгоритм обучения

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





МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 43 из 81

Первый экземпляр _____

КОПИЯ № _____

Рис. 2.17 - Установка погрешности и числа эпох обучения

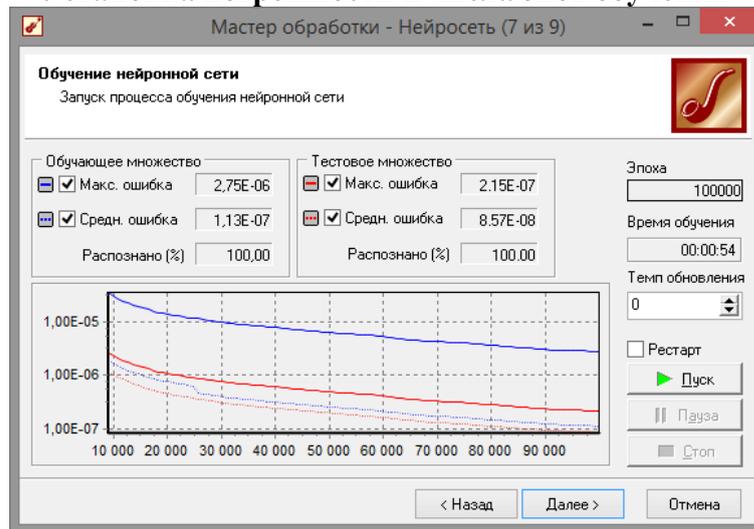
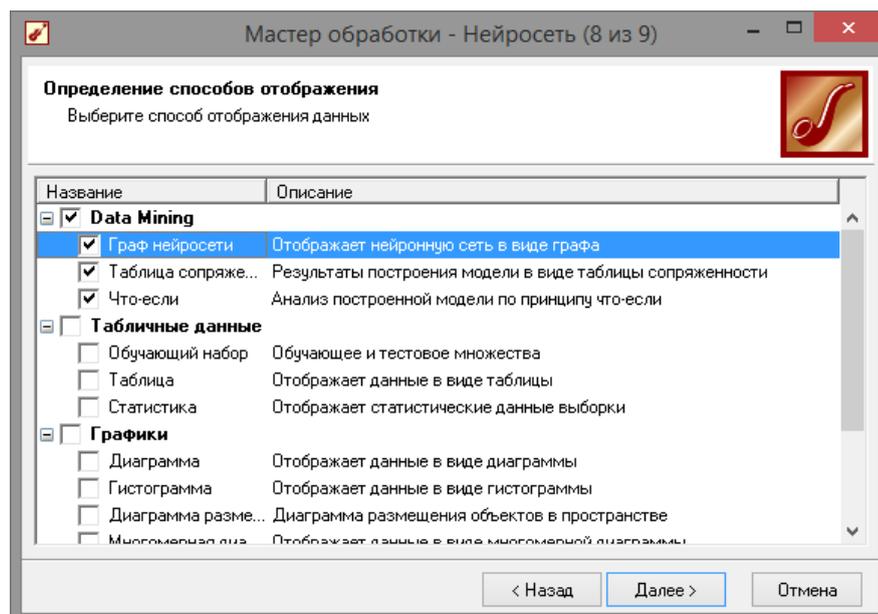


Рис. 2.18 - Обучение сети

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).





МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 44 из 81

Первый экземпляр _____

КОПИЯ № _____

Рис. 2.19 – Визуализаторы

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

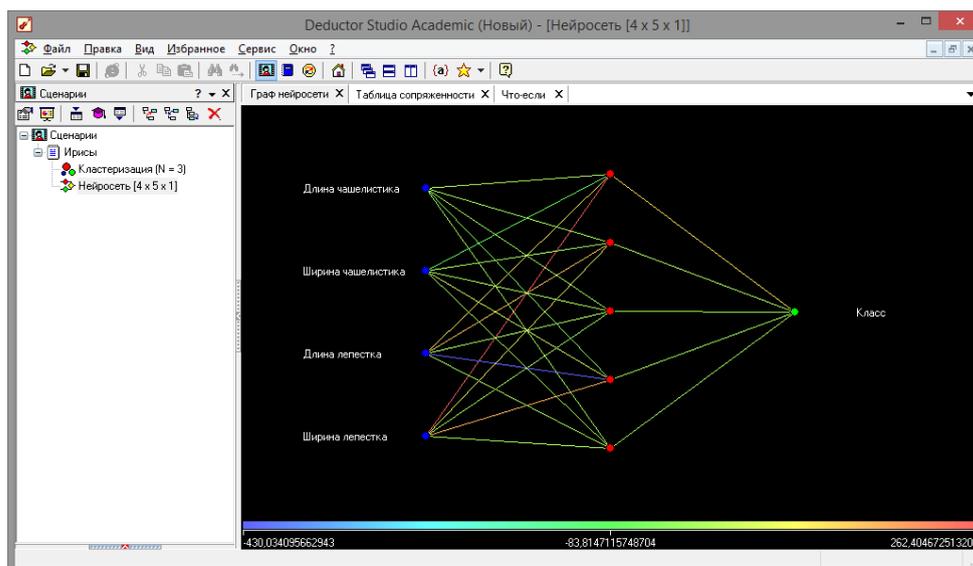


Рис. 2.20 - Граф нейросети

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



| Фактически | Классифицировано | | | Итого |
|-----------------|------------------|-----------------|----------------|-------|
| | Iris-setosa | Iris-versicolor | Iris-virginica | |
| Iris-setosa | 50 | | | 50 |
| Iris-versicolor | | 50 | | 50 |
| Iris-virginica | | | 50 | 50 |
| Итого | 50 | 50 | 50 | 150 |

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.

| Поле | Значение |
|------------------------|-------------|
| Выводные | |
| 9.0 Длина чашелист... | 5,12 |
| 9.0 Ширина чашелист... | 3,556 |
| 9.0 Длина лепестка | 1,4564 |
| 9.0 Ширина лепестка | 0,246541 |
| Выводные | |
| Класс | Iris-setosa |

Рис. 2.22 - Инструмент «Что-Если»



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 46 из 81

Первый экземпляр _____

КОПИЯ № _____

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 больший 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?



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 47 из 81

Первый экземпляр _____

КОПИЯ № _____

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



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 48 из 81

Первый экземпляр _____

КОПИЯ № _____

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 результатумножения, 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).

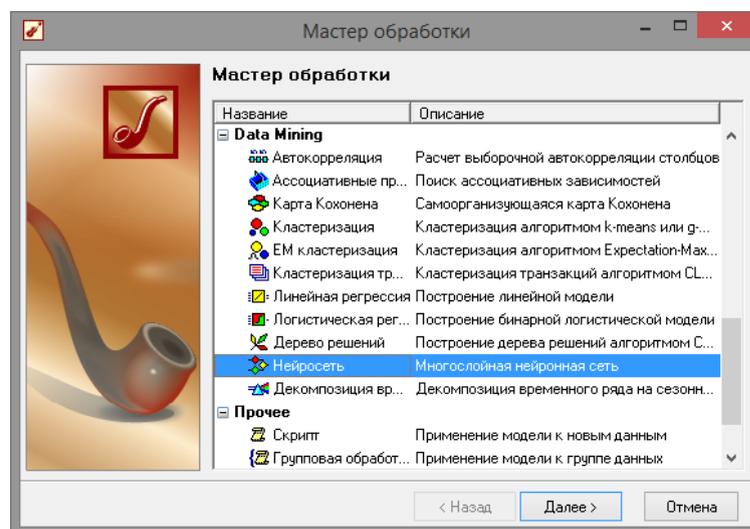


Рис. 3.1 - Мастер обработки

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



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 49 из 81

Первый экземпляр _____

КОПИЯ № _____

большое количество нейронов и слоев, может сильно замедлить процесс обучения. Эмпирически были выставлены настройки, показанные на рис. 3.4.

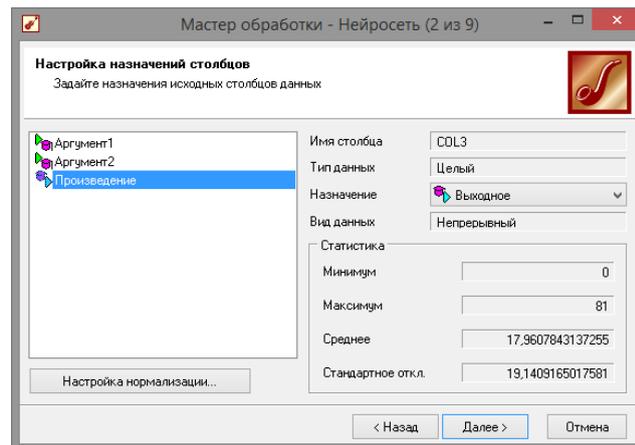


Рис. 3.2 - Мастер нейросети

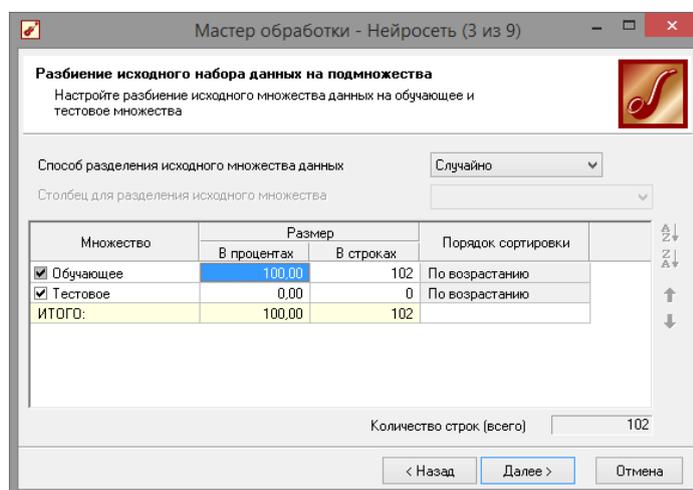
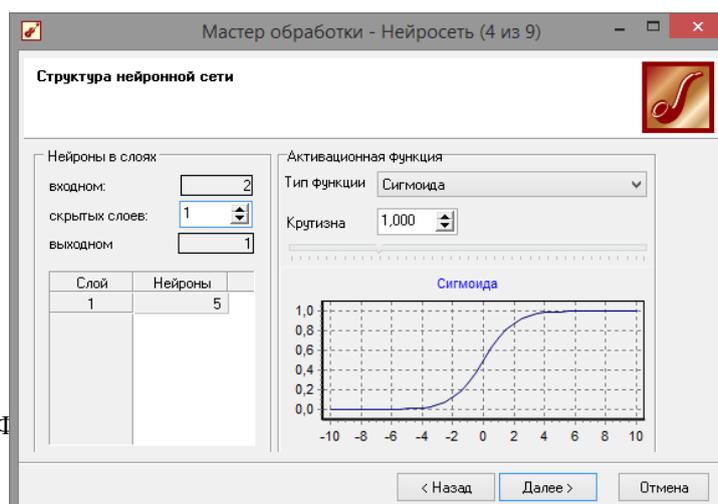


Рис. 3.3 - Настройки параметров обучения



| | | | |
|--|---------------|------------------------|---------------|
|  МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ») Костанайский филиал Кафедра социально-гуманитарных и естественнонаучных дисциплин | | | |
| Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент | | | |
| Версия документа - 1 | стр. 50 из 81 | Первый экземпляр _____ | КОПИЯ № _____ |

Рис. 3.4 - Структура нейронной сети

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

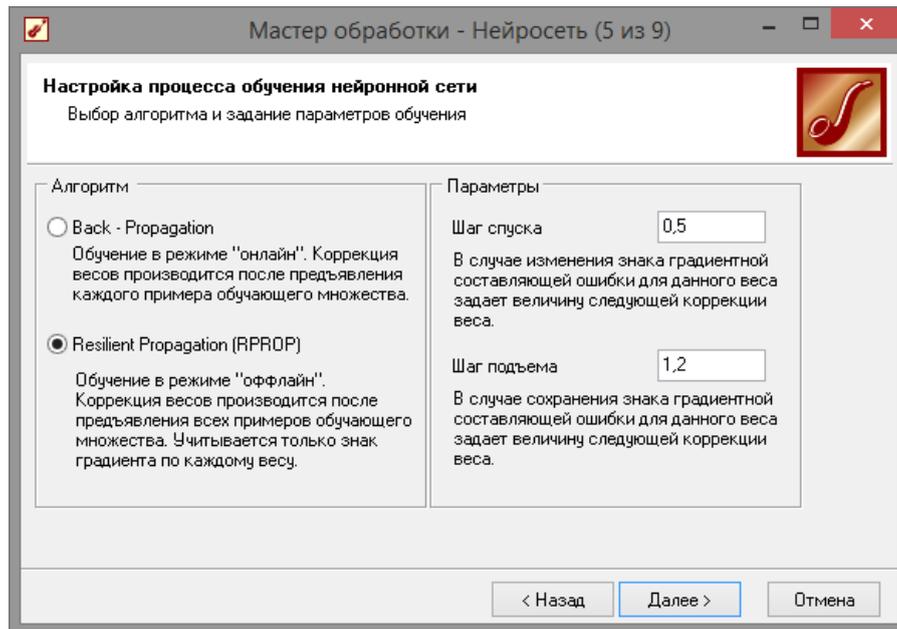


Рис. 3.5 - Алгоритм обучения

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

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 51 из 81

Первый экземпляр _____

КОПИЯ № _____

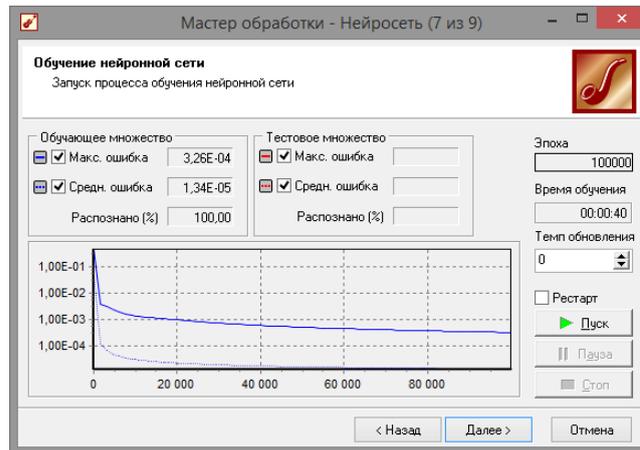


Рис. 3.6 - Настройка параметров остановки обучения

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.

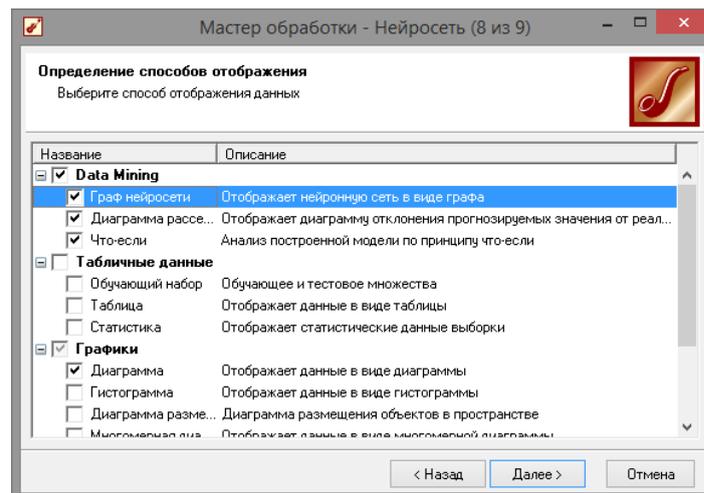


Рис. 3.7 - Обучение сети



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 52 из 81

Первый экземпляр _____

КОПИЯ № _____

Рис. 3.8 - Визуализация данных

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

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

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 53 из 81

Первый экземпляр _____

КОПИЯ № _____

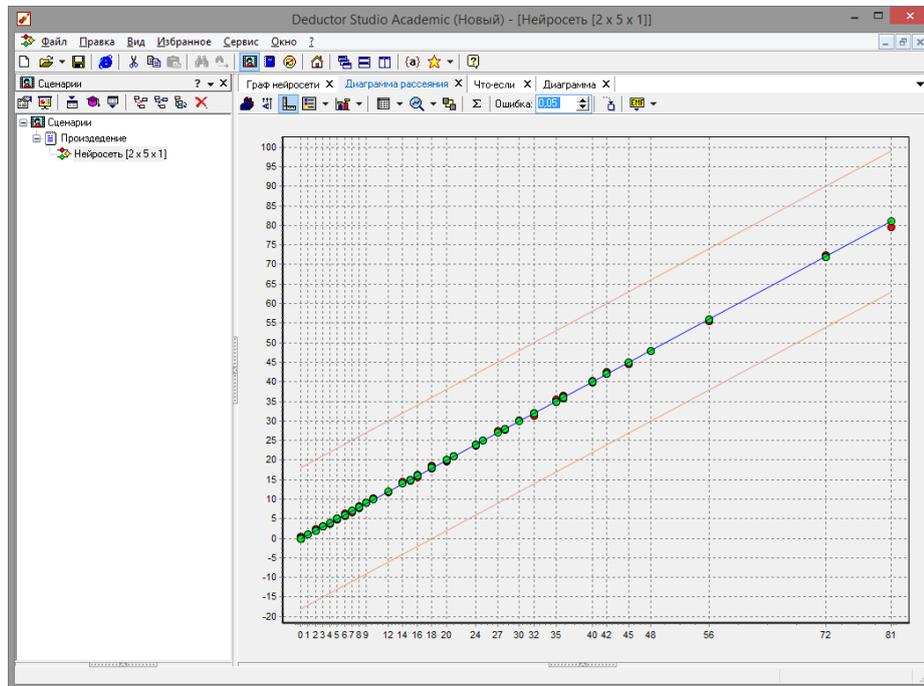


Рис. 3.9 - Диаграмма рассеяния

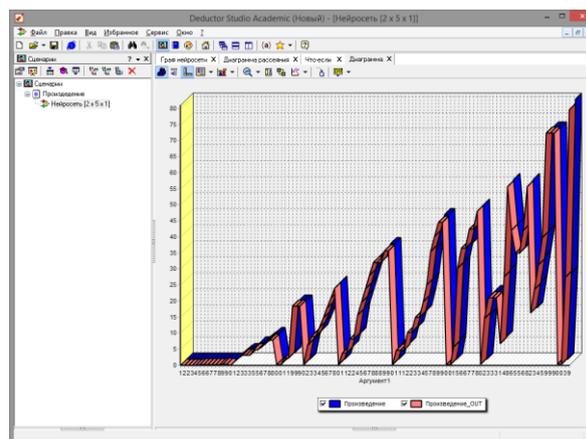


Рис. 3.10 - Сравнение эталонных данных с прогнозируемыми



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 54 из 81

Первый экземпляр _____

КОПИЯ № _____

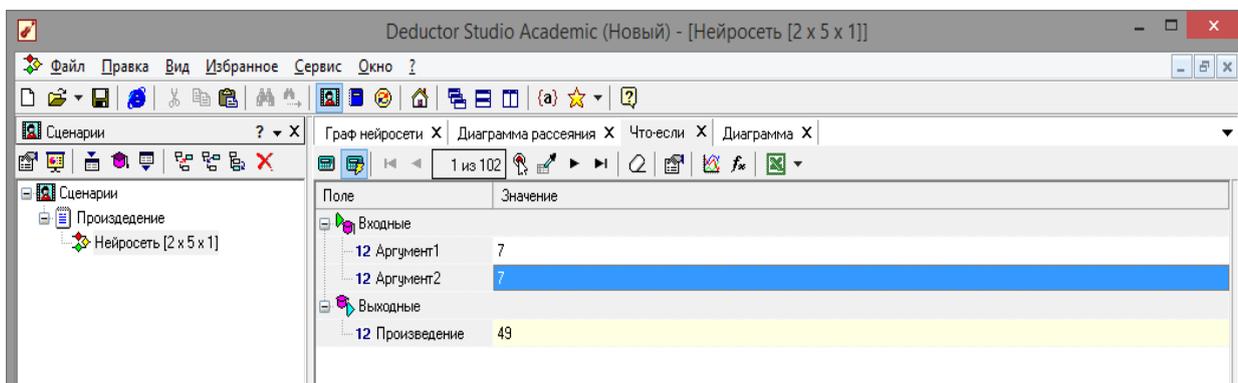


Рис. 3.11 - Инструмент «Что-Если»

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

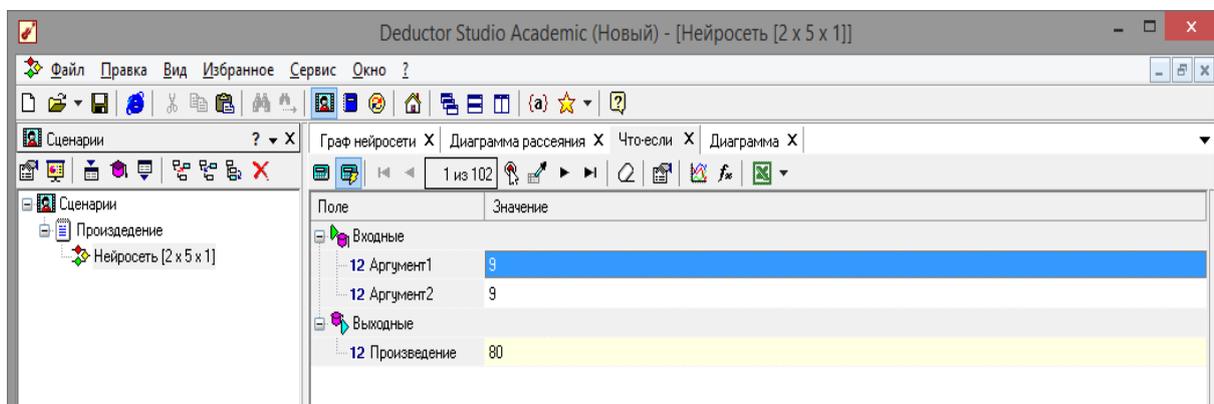


Рис. 3.12 - Ошибки прогнозирования

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



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 55 из 81

Первый экземпляр _____

КОПИЯ № _____

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.

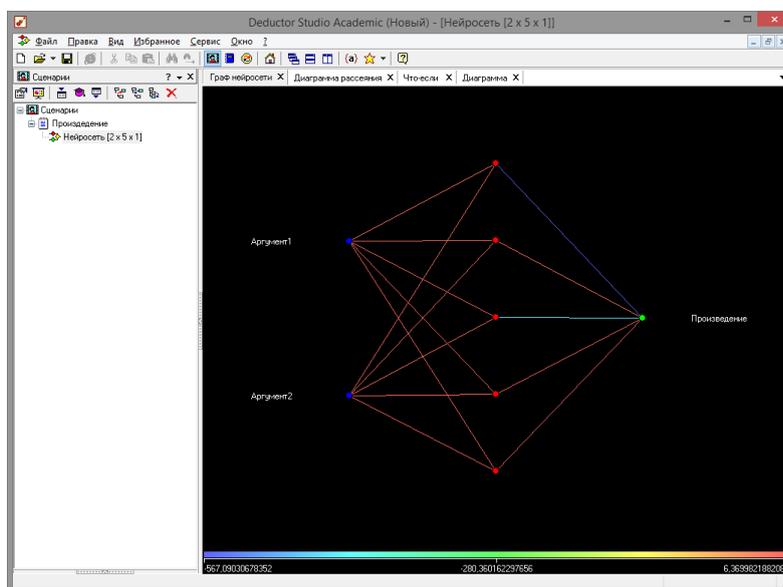


Рис. 3.13 - Граф нейронной сети

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



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 56 из 81

Первый экземпляр _____

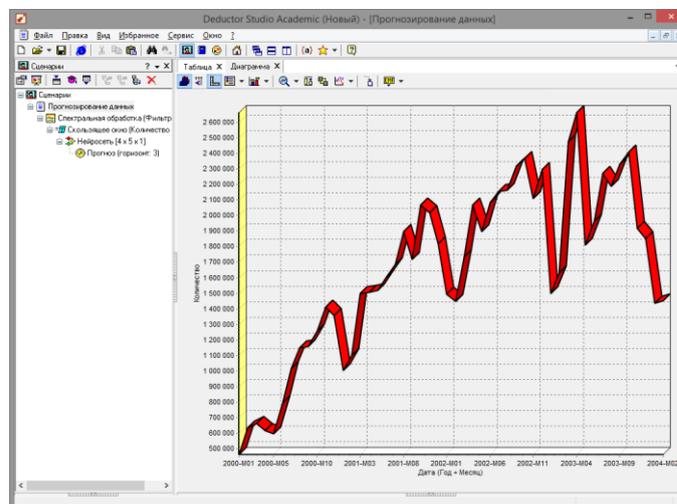
КОПИЯ № _____

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).





МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 57 из 81

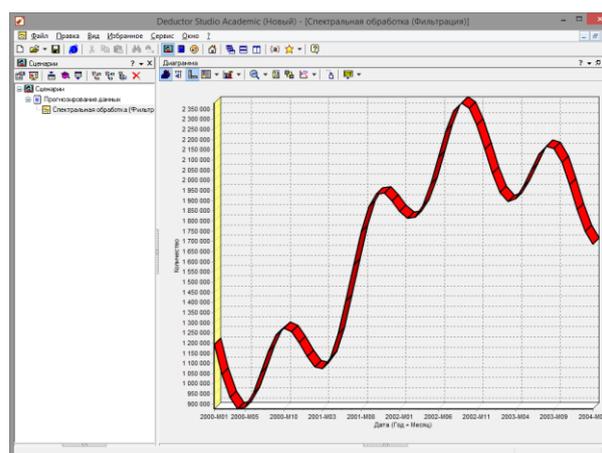
Первый экземпляр _____

КОПИЯ № _____

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.

Рис. 3.15 - Выбор спектральной обработки



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 58 из 81

Первый экземпляр _____

КОПИЯ № _____

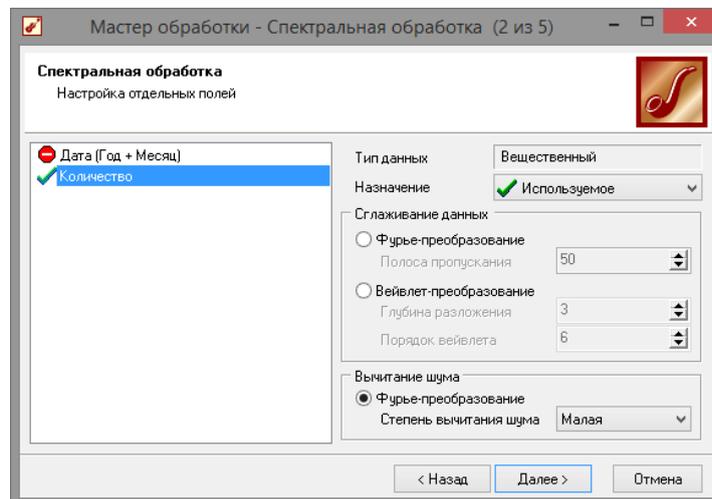


Рис. 3.16 - Результат спектральной обработки

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).

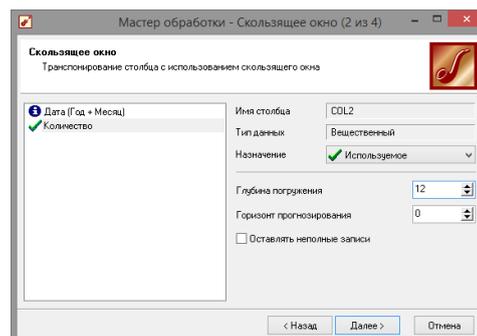


Рис. 3.17 - Скользящее окно

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».



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 59 из 81

Первый экземпляр _____

КОПИЯ № _____

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 (Гг + Мгс) | Количество 12 | Количество 11 | Количество 10 | Количество 9 | Количество 8 | Количество 7 |
|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| 2001.001 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 |
| 2001.002 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 |
| 2001.003 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 |
| 2001.004 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 |
| 2001.005 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 |
| 2001.006 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 |
| 2001.007 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 |
| 2001.008 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 |
| 2001.009 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 |
| 2001.010 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 |
| 2001.011 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 |
| 2001.012 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 |
| 2002.001 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 |
| 2002.002 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 |
| 2002.003 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 |
| 2002.004 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 |
| 2002.005 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 |
| 2002.006 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 |
| 2002.007 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 |
| 2002.008 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 |
| 2002.009 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 |
| 2002.010 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 |
| 2002.011 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 |
| 2002.012 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 |
| 2003.001 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 |
| 2003.002 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 |
| 2003.003 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 |
| 2003.004 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 |
| 2003.005 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 |
| 2003.006 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 |
| 2003.007 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 |
| 2003.008 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 |
| 2003.009 | 932948.6907043 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 |
| 2003.010 | 876011.6829453 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 |
| 2003.011 | 886303.9541079 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 |
| 2003.012 | 951010.8593006 | 1154171.0324344 | 1046415.5743264 | 932948.6907043 | 876011.6829453 | 886303.9541079 |

Рис. 3.18 - Таблица скользящего окна

Мастер обработки - Нейросеть (2 из 9)

Настройка назначений столбцов
Задайте назначения исходных столбцов данных

| Имя столбца | Тип данных | Назначение | Вид данных | Статистика |
|---------------|------------|------------|-------------|-------------------------------------|
| Количество-10 | COL2 | Выходное | Непрерывный | Минимум: 1075589,0358277 |
| Количество-9 | | | | Максимум: 2380312,1698561 |
| Количество-8 | | | | Среднее: 1864843,52986291 |
| Количество-7 | | | | Стандартное откл.: 354553,479266148 |
| Количество-6 | | | | |
| Количество-5 | | | | |
| Количество-4 | | | | |
| Количество-3 | | | | |
| Количество-2 | | | | |
| Количество-1 | | | | |
| Количество | | | | |

Настройка нормализации...

< Назад Далее > Отмена

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



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 60 из 81

Первый экземпляр _____

КОПИЯ № _____

«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.

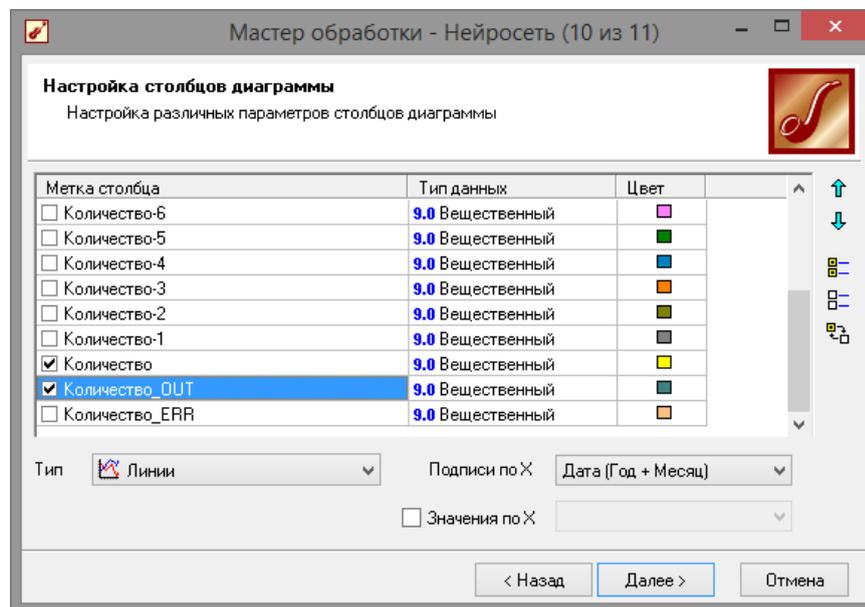


Рис. 3.20 - Настройки нейросети

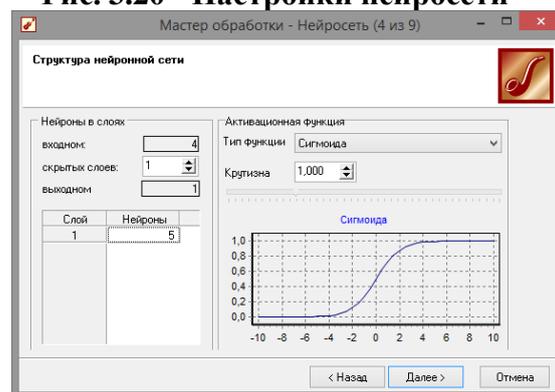


Рис. 3.21 - Обучение нейронной сети



МИНОБНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 61 из 81

Первый экземпляр _____

КОПИЯ № _____

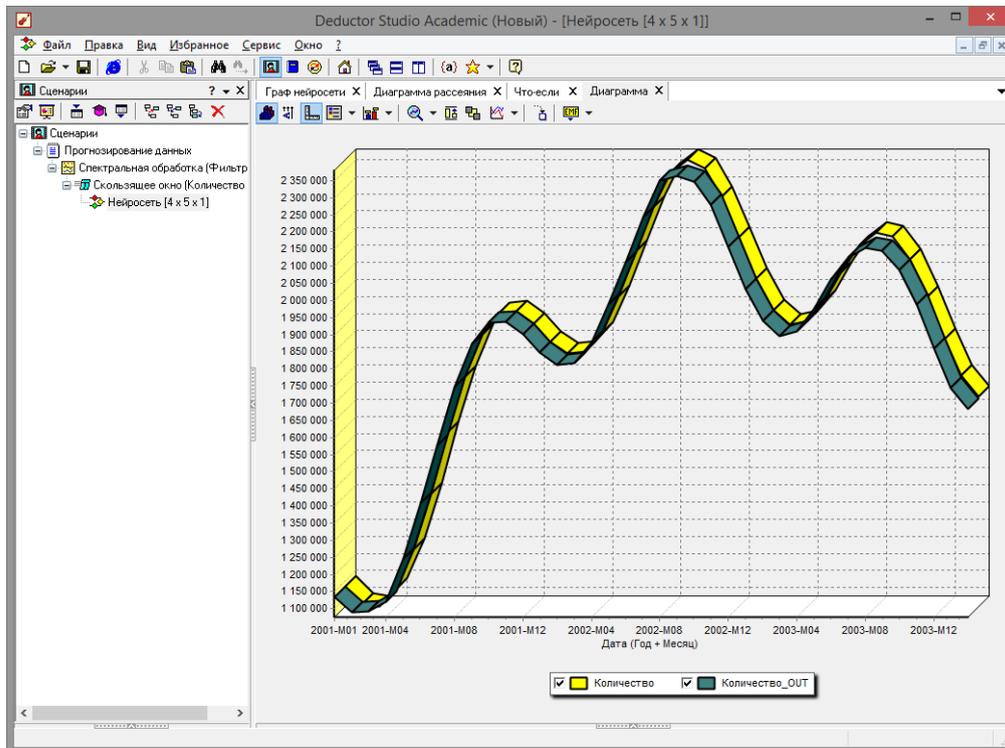


Рис. 3.22 - Сравнение эталонных данных с прогнозом

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

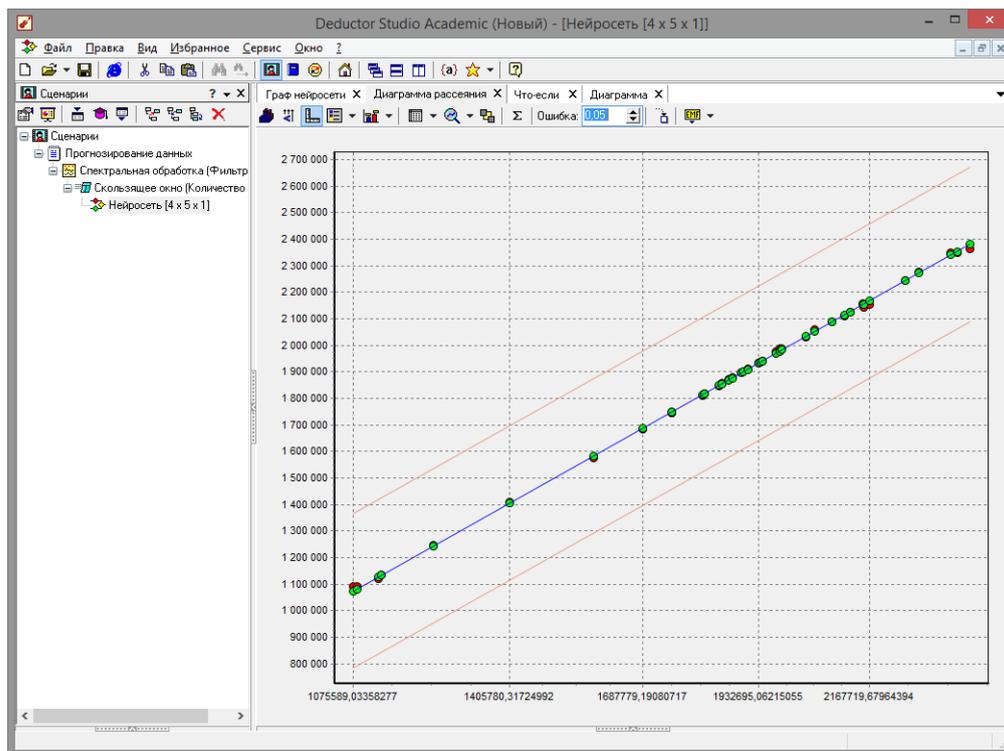


Рис. 3.23 - Диаграмма рассеяния

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).



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 63 из 81

Первый экземпляр _____

КОПИЯ № _____

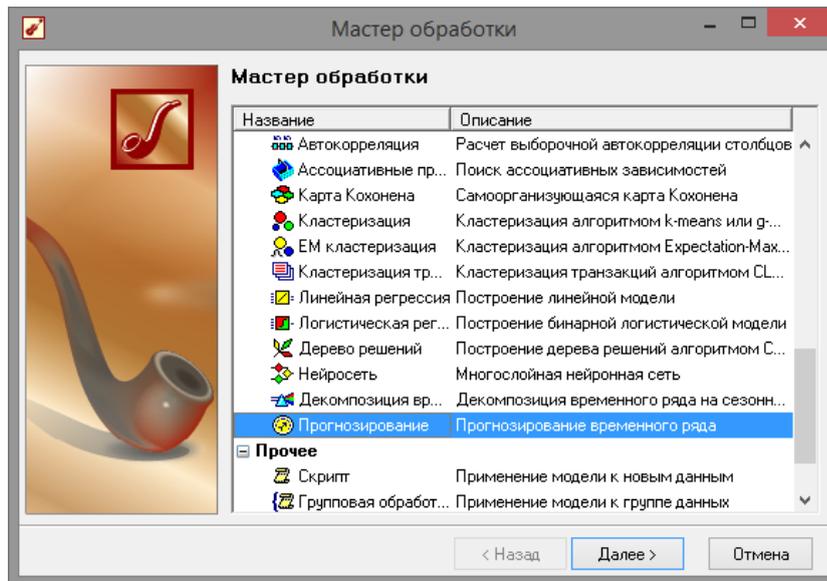


Рис. 3.24 - Мастер прогнозирования

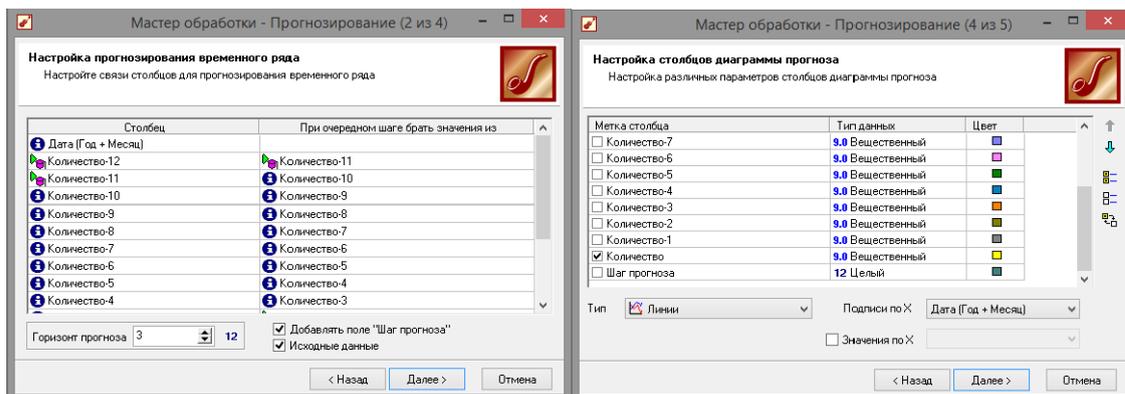


Рис. 3.25 - Мастер прогнозирования

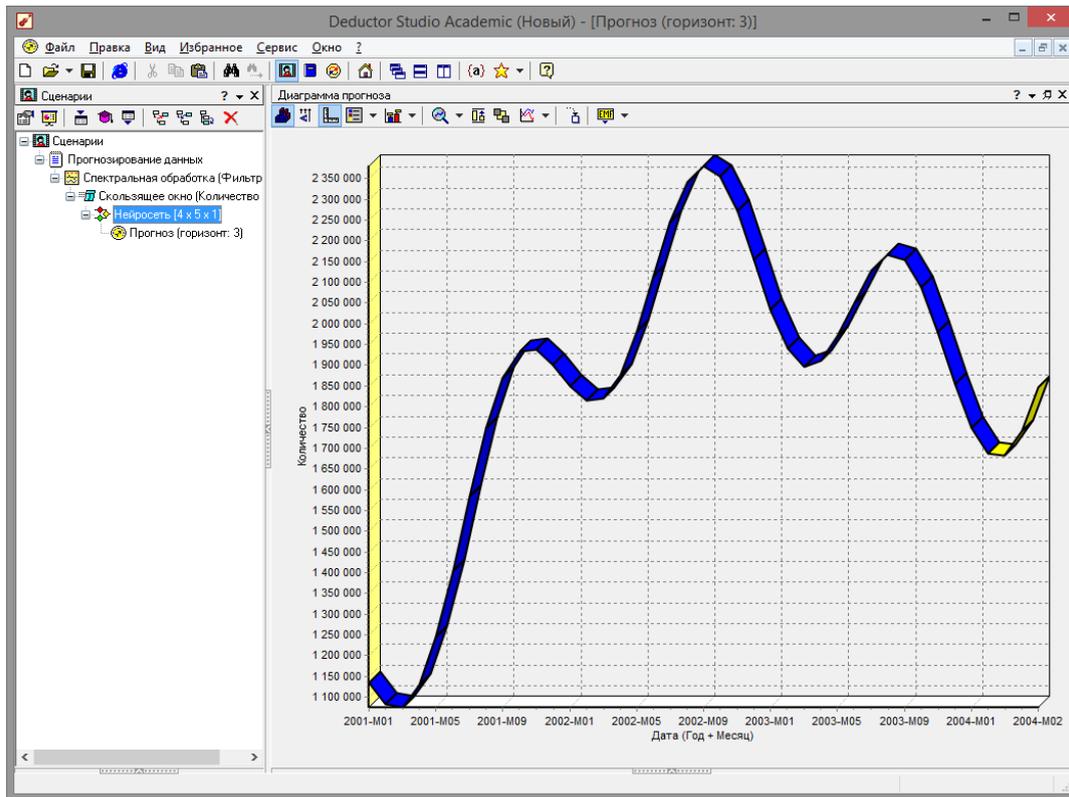


Рис. 3.26 - Расчетные значение на 3 месяца вперед

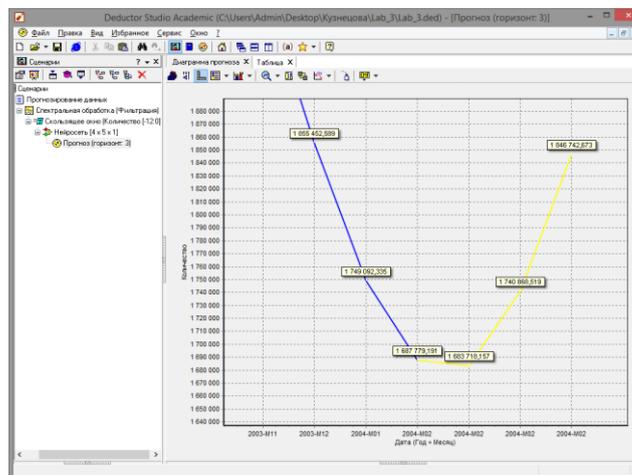


Рис. 3.27. Результаты мастера прогнозирования



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 65 из 81

Первый экземпляр _____

КОПИЯ № _____

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

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 66 из 81

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 кластеризации 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



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 67 из 81

Первый экземпляр _____

КОПИЯ № _____

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.



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 68 из 81

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 |
| 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

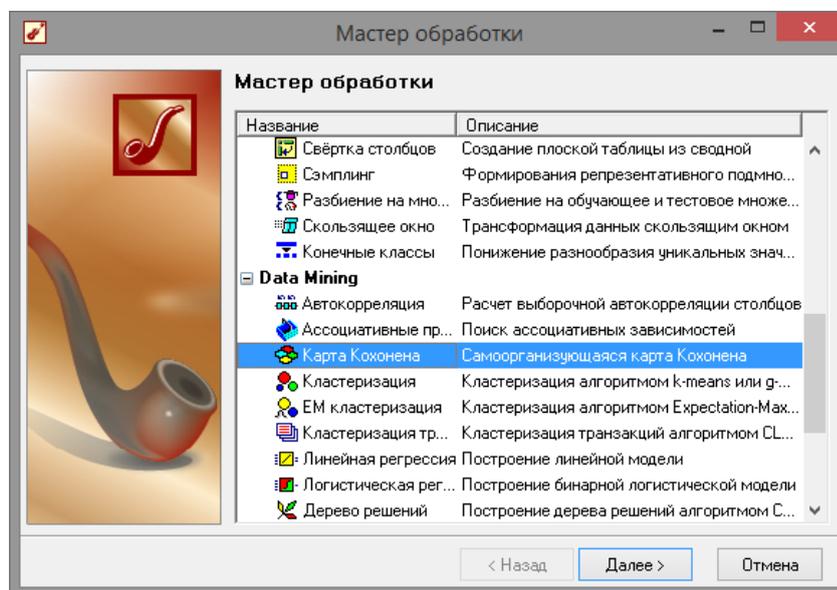


Рис. 4.2 - Мастер обработки «Карта Кохонена»



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 69 из 81

Первый экземпляр _____

КОПИЯ № _____

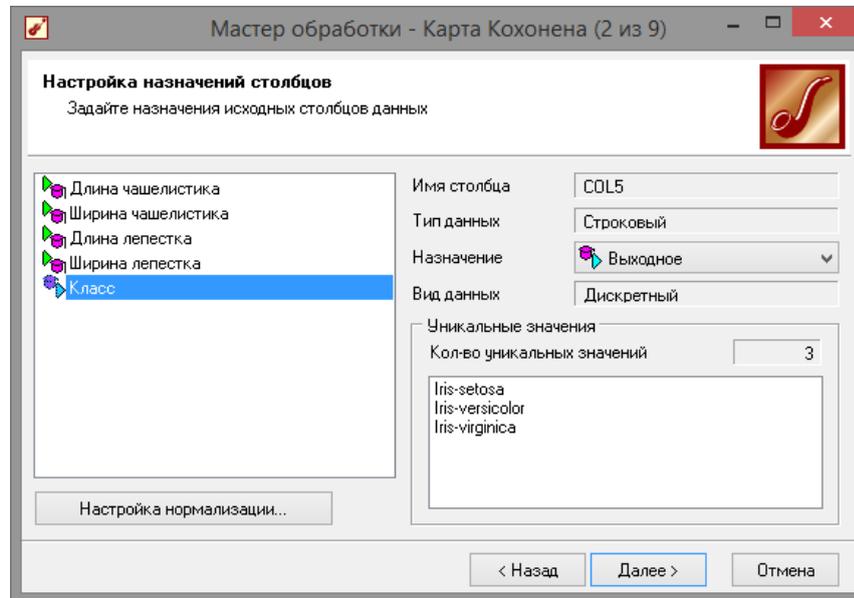


Рис. 4.3. Настройки мастера карт Кохонена

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).



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 70 из 81

Первый экземпляр _____

КОПИЯ № _____

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 | 93,33 | 140 | По возрастанию |
| <input checked="" type="checkbox"/> Тестовое | 6,67 | 10 | По возрастанию |
| ИТОГО: | 100,00 | 150 | |

Количество строк (всего): 150

< Назад Далее > Отмена

Рис.4.4 - Настройки тестового и обучающего множества

Настройка параметров карты Кохонена

Укажите значения параметров карты Кохонена

Параметры карты

Размер по оси X: 16

Размер по оси Y: 12

Кол-во ячеек: 192

Форма ячеек: Шестиугольные

Вид карты

< Назад Далее > Отмена

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

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 71 из 81

Первый экземпляр _____

КОПИЯ № _____

Мастер обработки - Карта Кохонена (5 из 9)

Настройка параметров остановки обучения

Укажите условия прекращения обучения. Обучение будет остановлено при выполнении одного из условий.

Считать пример распознанным, если ошибка меньше: 0,05

По достижению эпохи: 500

Обучающее множество

Средняя ошибка меньше

Максимальная ошибка меньше

Распознано примеров (%): 0

Тестовое множество

Средняя ошибка меньше

Максимальная ошибка меньше

Распознано примеров (%): 0

< Назад Далее > Отмена

Рис.4.6 - Настройки параметров остановки обучения

Мастер обработки - Карта Кохонена (6 из 9)

Настройка параметров обучения карты Кохонена

Укажите значения параметров обучения карты Кохонена

Способ начальной инициализации карты: Из собственных векторов

Количество эпох, через которое необходимо перемешивать строки: 20

Скорость обучения

В начале обучения: 0,3

В конце обучения: 0,005

Радиус обучения

В начале обучения: 4

В конце обучения: 0,1

Функция соседства: Ступенчатая

Классеризация

Автоматически определить количество кластеров

Уровень значимости, %: 0,1

Фиксированное кол-во кластеров: 3

< Назад Далее > Отмена

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

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 72 из 81

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.

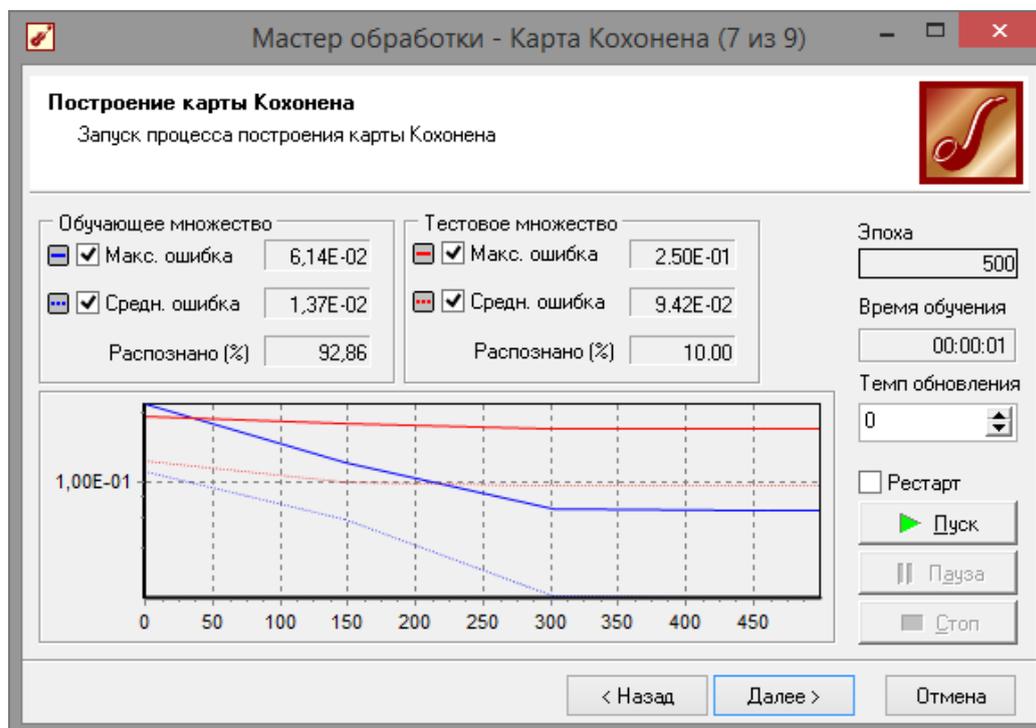


Рис. 4.8 - Процесс обучения

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

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 73 из 81

Первый экземпляр _____

КОПИЯ № _____

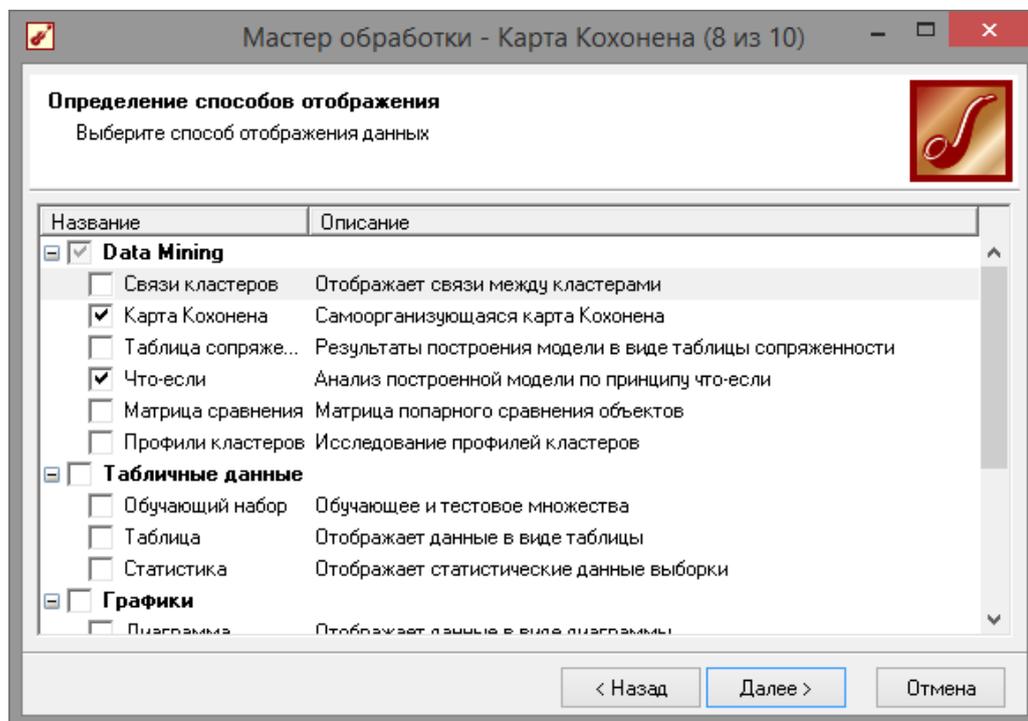


Рис. 4.9 - Способ отображения данных

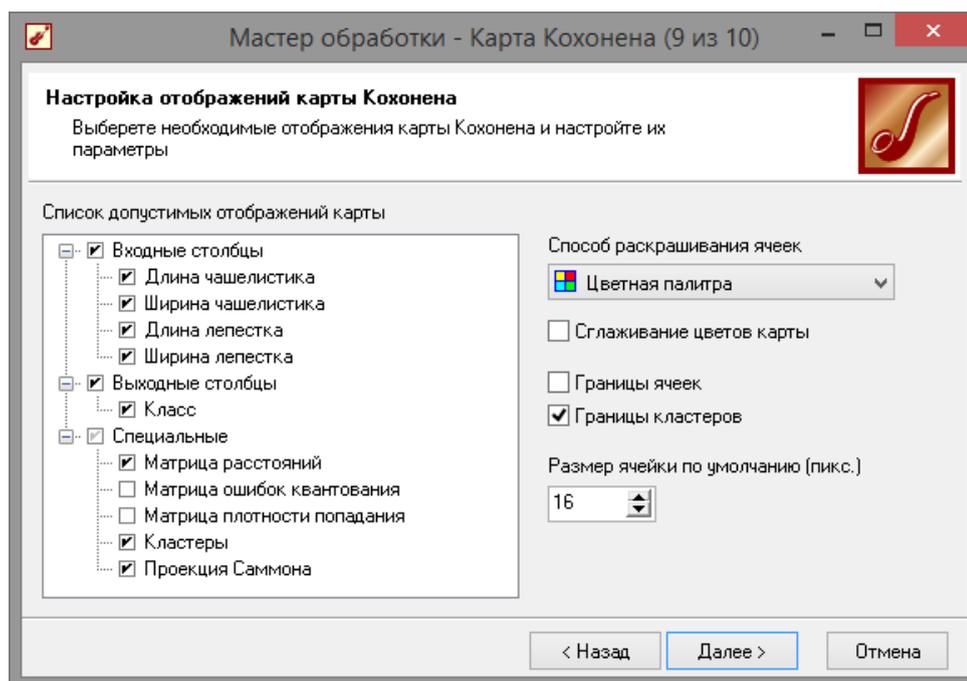




Рис. 4.10 - Настройка отображения кластеров

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.

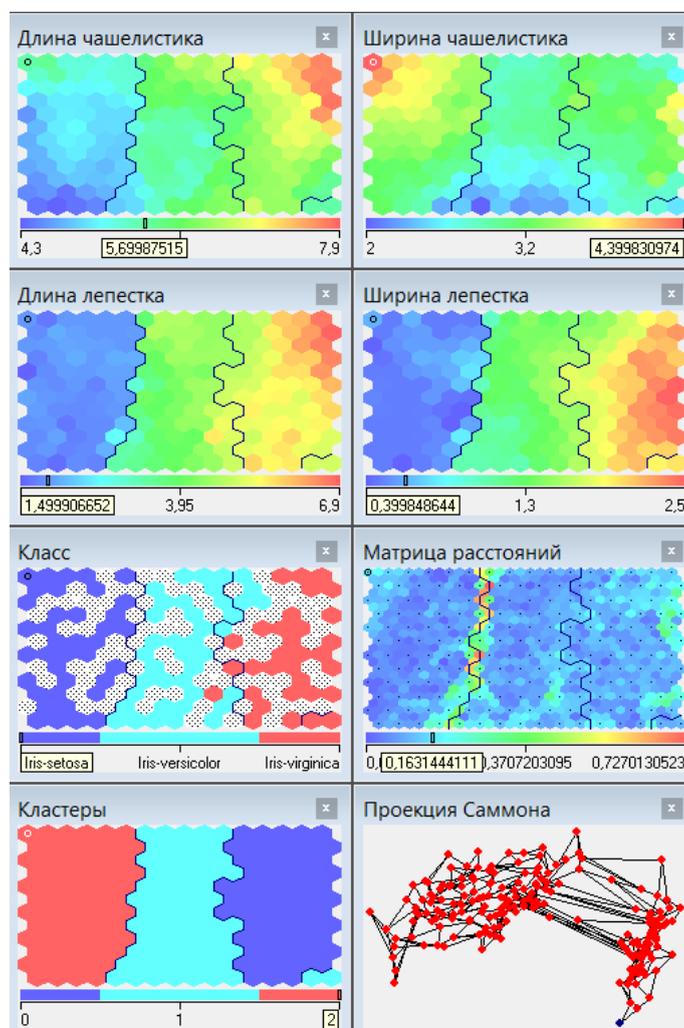


Рис. 4.11 - Карта Кохонена



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 75 из 81

Первый экземпляр _____

КОПИЯ № _____

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

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 76 из 81

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

Методические рекомендации для подготовки к занятиям лекционного типа, к выполнению практических работ и к самостоятельной работе студентов по дисциплине (модулю) «Интеллектуальный анализ данных» по основной профессиональной образовательной программе высшего образования – программе бакалавриата «Управление бизнесом» по направлению подготовки 38.03.02 Менеджмент

Версия документа - 1

стр. 77 из 81

Первый экземпляр _____

КОПИЯ № _____

| Поле | Значение |
|-----------------------------------|-------------------|
| Входные | |
| 9.0 Длина чашелистика | 4,7 |
| 9.0 Ширина чашелистика | 3,6746 |
| 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», «Versicolour», «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 скопления данных. Мастер предоставляет широкий набор настройки параметров обучения: настройка нормализации столбцов, настройка разбиения на тестовое и обучающее множество, настройка условий остановки 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

Методические рекомендации для подготовки к занятиям лекционного, семинарского типа и к самостоятельной работе студентов 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. 78 из 11

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.

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ СТУДЕНТАМ ПО ПОДГОТОВКЕ К САМОСТОЯТЕЛЬНОЙ РАБОТЕ

Общие методические рекомендации для выполнения СРС

Самостоятельная работа студентов является наиболее продуктивной формой образовательной и познавательной деятельности студента в период обучения. Текущая самостоятельная работа направлена на углубление и закрепление знаний студентов, развитие практических навыков и умений. Текущая самостоятельная работа включает в себя: работу с лекционным материалом, опережающую самостоятельную работу, подготовку к промежуточной аттестации. Самостоятельная работа студентов заключается в выполнении индивидуальных домашних заданий. При выполнении индивидуальных домашних заданий необходимо использовать теоретический материал, а так же знания и умения, полученные в ходе выполнения практических работ.

Самостоятельность в учебной работе способствует развитию заинтересованности студента в изучаемом материале, вырабатывает у него умение и потребность самостоятельно получать знания, что весьма важно для специалиста с высшим образованием.

Целью самостоятельной работы студентов является овладение фундаментальными знаниями, профессиональными умениями и навыками деятельности по профилю, опытом творческой, исследовательской деятельности.

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ВЫПОЛНЕНИЮ И ОФОРМЛЕНИЮ СРС ДЛЯ СТУДЕНТОВ ОЧНОЙ ФОРМЫ ОБУЧЕНИЯ

В одном семестре предусмотрено четыре СРС: №1, №2, №3, №4

| | | | |
|---|--|------------------------|---------------|
|  | МИНОБРНАУКИ РОССИИ Федеральное государственное бюджетное образовательное учреждение высшего образования «Челябинский государственный университет» (ФГБОУ ВО «ЧелГУ») Костанайский филиал Кафедра социально-гуманитарных и естественнонаучных дисциплин | | |
| | Методические рекомендации для подготовки к занятиям лекционного, семинарского типа и к самостоятельной работе студентов 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. 79 из 11 | Первый экземпляр _____ | КОПИЯ № _____ |

При выполнении заданий (самостоятельная работа студента) необходимо руководствоваться лекционным материалом, использовать для работы учебники.

СРС включает в себя наиболее типичные и распространённые практические задания по основным разделам учебной программы.

Каждый студент очной формы обучения (таблица) обязан выполнить все задания и предоставить их преподавателю для проверки в установленные сроки (не позднее установленных сроков) или прикрепить в СДО «Moodle», с случае внедрения дистанционного обучения в период карантина в группе (COVID-19).

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.

Индивидуальное задание для прогнозирования (изменение курса валют, график синусоиды, прогнозирование суммы или разности чисел и др.).

СРС № 3

Задание

Сети прямого распространения ошибки. Карты Кохонена

Индивидуальное задание 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.

СРС № 4

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.

Критерии оценивания конспектов

«отлично» (90-100%) выставляется студенту, если демонстрируются полнота использования учебного материала, логика изложения (наличие схем, количество смысловых связей между



МИНОБРНАУКИ РОССИИ

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

Костанайский филиал

Кафедра социально-гуманитарных и естественнонаучных дисциплин

Методические рекомендации для подготовки к занятиям лекционного, семинарского типа и к самостоятельной работе студентов 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. 80 из 11

Первый экземпляр _____

КОПИЯ № _____

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 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 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

Методические рекомендации для подготовки к занятиям лекционного, семинарского типа и к самостоятельной работе студентов 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. 81 из 11

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 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 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.