

федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»

Б1.В.ДВ.01.02
(индекс дисциплины)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Интеллектуальный анализ на основе хранилищ данных
(наименование дисциплины)

по направлению подготовки (специальности)
09.04.03 Прикладная информатика

(код и наименование направления подготовки, специальности в соответствии с ФГОС ВПО/ ФГОС ВО)

Информационные системы и технологии корпоративного управления
(направленность (профиль))

Форма обучения: заочная

Год набора: 2019

Распределение часов дисциплины по семестрам и видам занятий (по учебному плану)

Количество ЗЕТ	7						
Часов по РУП	252						
Виды контроля на курсах	Экзамены	Зачеты	Курсовые проекты	Курсовые работы	Контрольные работы (для заочной формы обучения)		
		1					
	№№ курса						
	1	2	3	4	5	6	Итого
ЗЕТ по курсам	7						7
Лекции	4						4
Лабораторные	6						6
Практические							
Контактная работа	10,25						10,25
Сам. работа	238						238
Контроль	3,75						3,75
Итого	252						252

Тольятти, 2019

Рабочая программа составлена на основании ФГОС ВО и учебного плана направления подготовки (специальности) 09.03.03 Прикладная информатика
(код и наименование направления подготовки, специальности в соответствии с ФГОС ВПО/ ФГОС ВО)

Рецензирование рабочей программы дисциплины:



Отсутствует



Учебная (рабочая) программа одобрена на заседании кафедры «Прикладная математика и информатика» (протокол заседания № 6 от «13 » февраля 2019 г.).



Рецензент

(должность, ученое звание, степень)

«__»____20__г.

(подпись)

(И.О. Фамилия)

Срок действия рабочей программы дисциплины до « 25 » мая 2022 г.

Информация об актуализации рабочей программы дисциплины:

Протокол заседания кафедры № 1 от «09» сентября 2019 г.

Протокол заседания кафедры № 1 от « 28» августа 2020 г.

Протокол заседания кафедры № ____ от «__»_____ 20__ г.

Протокол заседания кафедры № ____ от «__»_____ 20__ г.

УТВЕРЖДАЮ

Заведующий кафедрой

Прикладная математика и информатика

(разработавшей РПД)

«__»____20__г.

(подпись)

А.В. Очеповский

(И.О. Фамилия)

АННОТАЦИЯ
дисциплины (учебного курса)
Б1.В.ДВ.01.02 Интеллектуальный анализ на основе хранилищ дан-
ных

(индекс и наименование дисциплины (учебного курса))

Данный курс предназначен для ознакомления с современными методами интеллектуального анализа на основе хранилищ данных. Технологии интеллектуального анализа данных изучает процесс нахождения новых, действительных и потенциально полезных знаний в хранилище данных. Эти технологии лежат на пересечении таких наук как системы баз данных, статистика и искусственный интеллект. Интеллектуальный анализ данных применяется в областях, имеющих дело с большими объемами данных. В науке к этим областям относятся астрономия, биологии, биоинформатика, медицина, физика, в бизнесе – это торговля, телекоммуникации, банковское дело, промышленное производство.

В курсе рассматриваются технологии и алгоритмы для хранения и организации данных, включая Хранилища Данных и OLAP, а также методы и алгоритмы реализации интеллектуального анализа данных (Data Mining).

1. Цель и задачи изучения дисциплины (учебного курса)

Цель – формирование представления о современных методах и технологиях интеллектуального анализа на основе хранилищ данных, которые применяются на практике в различных областях науки и техники .

Задачи:

1. Сформировать у студентов представление о концепциях и принципах организации хранилища данных.
2. Обучить студентов понятиям и архитектуре OLAP-систем.
3. Сформировать у студентов представление о теоретических основах интеллектуального анализа данных.
4. Сформировать у студентов практические навыки по решению задач классификации и регрессии при интеллектуальном анализе данных.
5. Сформировать у студентов практические приемы по поиску ассоциативных правил при интеллектуальном анализе данных.

2. Место дисциплины (учебного курса) в структуре ОПОП ВО

Данная дисциплина (учебный курс) относится к Блоку 1 «Дисциплины (модули)» (вариативная часть, дисциплины по выбору).

Дисциплины, учебные курсы, на освоении которых базируется данная дисциплина (учебный курс) – «Дискретная математика», «Базы данных», «Объектно-ориентированное программирование».

Дисциплины, учебные курсы, для которых необходимы знания, умения, навыки, приобретаемые в результате изучения данной дисциплины (учебного курса) – научно-исследовательская работа (подготовка к магистерской диссертации).

3. Планируемые результаты обучения по дисциплине (учебному курсу), соотнесенные с планируемыми результатами освоения образовательной программы

Формируемые и контролируемые компетенции	Планируемые результаты обучения
- Способен проводить обследование организаций, выявлять информационные потребности пользователей, формировать требования к программному обеспечению (ПК-1)	Знать: способы решения прикладных задач с помощью методов интеллектуального анализа данных
	Уметь: применять методы интеллектуального анализа данных при решении прикладных задач
	Владеть: Навыками программной реализации методов интеллектуального анализа данных для решения прикладных задач
- Способен проектировать информационные системы по видам обеспечения (ПК-4)	Знать: Способы применения методов интеллектуального анализа данных при проведении научных исследований
	Уметь: Планировать научные эксперименты с использованием методов интеллектуального анализа данных
	Владеть: Навыками оценки результатов, получаемых в результате интеллектуального анализа данных

Тематическое содержание дисциплины (учебного курса)

Раздел, модуль	Подраздел, тема
Модуль 1 - хранилища данных, OLAP-системы	Тема 1 - основные понятия принципы работы хранилищ данных. Тема 2 - основные понятия принципы работы OLAP-систем. Тема 3 - Кластеризация при интеллектуальном анализе данных.
Модуль 2 - алгоритмы интеллектуального анализа данных	Тема 4 - классификация при интеллектуальном анализе данных. Тема 5 - классификация на основе алгоритмов ID3, C4.5. Тема 6 - классификация на основе алгоритма CART.

	<p>Тема 7 - регрессионный анализ при интеллектуальном анализе данных.</p> <p>Тема 8 - регрессионный анализ на основе нейронных сетей.</p> <p>Тема 9 - регрессионный анализ на основе алгоритма CART.</p> <p>Тема 10 - аффинитивный анализ при интеллектуальном анализе данных.</p> <p>Тема 11 - аффинитивный анализ на основе алгоритма Apriori.</p>
--	--

Общая трудоемкость дисциплины (учебного курса) – 7 ЗЕТ.

4. Структура и содержание дисциплины (учебного курса) Интеллектуальный анализ на основе хранилищ данных
(наименование дисциплины (учебного курса))

Семестр изучения 2

Раздел, модуль	Подраздел, тема	Виды учебной работы						Необходимые материально- технические ресурсы	Формы текущего кон- троля	Реко- мендуе- мая ли- терату- ра (№)	
		Аудиторные занятия (в часах)					Самостоятельная работа				
		всего			в т.ч. в ин- терактивной	Формы проведения лекций, лабораторных, практических занятий, методы обучения, ре- ализующие применяе- мую образовательную технологию	в часах				формы организа- ции самостоятельной работы
		лекций	лабораторных	практических							
Хранилища данных, OLAP- системы	Основные понятия принципы работы хра- нилищ данных. Основ- ные понятия принципы работы OLAP-систем	2				Лекция	138	Работа с лек- ционным ма- териалом, подготовка задания по лабораторной работе к за- щите	Компьютер- ный класс общего дос- тупа	Собеседо- вание	1,2
	Кластеризация при ин- теллектуальном анали- зе данных		2			Лабораторная				Отчет по лаборатор- ной работе	
Алгоритмы ин- теллектуального анализа данных	Классификация при ин- теллектуальном анали- зе данных. Регрессион- ный анализ при интел- лектуальном анализе данных	2				Проблемная лек- ция	100	Работа с лек- ционным ма- териалом, подготовка задания по лабораторной работе к за- щите	Компьютер- ный класс общего дос- тупа	Собеседо- вание	1,2
	Классификация на ос- нове алгоритмов ID3, C4.5. Классификация на основе алгоритма CART		2			Лабораторная				Отчет по лаборатор- ной работе	
	Регрессионный анализ		2							Лабораторная	

	на основе нейронных сетей. Регрессионный анализ на основе алгоритма CART. Аффинитивный анализ на основе алгоритма Apriori.									ной работе	
Итого:		4	6				238				
		10									

5. Критерии и нормы текущего контроля и промежуточной аттестации

Формы текущего контроля	Условия допуска	Критерии и нормы оценки
Собеседование	Присутствие на занятии.	<p>Отметка «зачтено» ставится студенту, который не продемонстрировал знание материала, ориентируется в изученном материале.</p> <p>Отметка «не зачтено» ставится студенту, который не продемонстрировал знание материала.</p>
Отчет по лабораторной работе	Выполнение лабораторной работы на компьютере и представление преподавателю результатов выполненной работы	<p>Отметка «зачтено» ставится студенту, который продемонстрировал результаты выполнения лабораторной работы, соответствующие поставленным задачам, и предоставил отчет, оформленный должным образом и содержащий краткое описание полученных результатов</p> <p>Отметка «не зачтено» ставится студенту, который не продемонстрировал результаты выполнения лабораторной работы или не представил по ней отчет или представленный отчет не соответствует требованиям по оформлению.</p>

Форма проведения промежуточной аттестации	Условия допуска	Критерии и нормы оценки	
Зачет (устно)	Получение «зачтено» по всем формам текущего контроля	«Зачтено»	Студент должен <i>знать</i> основные определения, методы и технологии искусственного интеллекта, уметь <i>анализировать</i> и сравнивать различные алгоритмы искусственного интеллекта.
		«Не зачтено»	Студент не знает наиболее важные определения, не может ориентироваться в материале.

6. Критерии и нормы оценки курсовых работ (проектов)

Учебным планом данный подраздел не предусмотрен

7. Примерная тематика письменных работ (курсовых, рефератов, контрольных, расчетно-графических и др.)

Учебным планом данный подраздел не предусмотрен

8. Вопросы к зачету

№ п/п	Вопросы
1	Концепция и организация хранилища данных.
2	Многомерная модель данных.
3	Определение и архитектура OLAP-систем.
4	Задачи интеллектуального анализа данных.
5	Практическое применение интеллектуального анализа данных.
6	Модели интеллектуального анализа данных.
7	Методы Data Mining.
8	Метода построения деревьев принятия решений.
9	Алгоритмы поиска ассоциативных правил.
10	Неиерархические алгоритмы кластеризации.
11	Нечеткая логика в интеллектуальном анализе данных.
12	Генетические алгоритмы в интеллектуальном анализе данных.
13	Нейронные сети в интеллектуальном анализе данных.
14	Алгоритм k-means в интеллектуальном анализе данных.
15	Алгоритм C-means в интеллектуальном анализе данных.
16	Алгоритм Apriori в интеллектуальном анализе данных.
17	Неэффективность использования OLTP-систем анализа данных.
18	Процесс обнаружения новых знаний в данных.
19	Алгоритм ID3 в интеллектуальном анализе данных.
20	Алгоритм C4.5 в интеллектуальном анализе данных.
21	Алгоритм CART в интеллектуальном анализе данных.
22	Генетический алгоритмы в интеллектуальном анализе данных.
23	Нейронные сети в интеллектуальном анализе данных.
24	Основные понятия и назначения алгоритма k-means.
25	Математический аппарат алгоритма k-means.
26	Начальные данные необходимые для использования алгоритма k-means.
27	Метрики применяемые в алгоритме k-means.
28	Критерии остановки выполнения алгоритма k-means.
29	Вопрос о необходимости нормировки данных, ее назначение в алгоритме k-means.
30	Алгоритм ID3: критерий выбора условия для разбиения исходного множества.
31	Особенности алгоритма ID3 по сравнению с алгоритмами построения ДПР.
32	Алгоритм ID3 и проблема переобучения.

33	Математический аппарат алгоритма ID3.
34	Алгоритм C4.5: критерий выбора условия для разбиения исходного множества.
35	Особенности алгоритма C4.5 по сравнению с алгоритмами построения ДПР.
36	Алгоритм C4.5 и проблема переобучения.
37	Математический аппарат алгоритма C4.5.
38	Основные понятия искусственных нейронных сетей (ИНС)
39	Биологическая модель нейрона.
40	Преобразование сигналов искусственным нейроном
41	Цель обучения нейронной сети
42	Обучение нейронной сети методом дельта-правила
43	Обучение многослойной нейронной сети как задача многопараметрической оптимизации
44	Обучение нейронных сетей методом обратного распространения ошибки
45	Проблема выбора количества слоев и нейронов в ИНС
46	Виды активационных функций, их графики
47	Взаимосвязь между используемой в нейронах активационной функцией и ограничениями по диапазону изменения входных сигналов
48	Способы улучшения результатов обучения многослойных нейронных сетей
49	Персептрон и проблема «исключающего ИЛИ»
50	Задачи, решаемые с помощью различных типов нейронных сетей
51	Проблема переобучения искусственных нейронных сетей и способы их преодоления
52	Оценка точности работы нейронных сетей.
53	Требования предъявляемые к обучающей выборке.
54	Самоорганизующаяся карта Кохонена: назначение, особенности.
55	Назначение и особенности рекуррентных нейронных сетей.
56	Нейронные сети Хопфилда и Хемминга, их особенности и отличия.
57	Способы кодирования сигналов в биполярный вид для нейронных сетей Хопфилда и Хемминга.
58	Влияние скорости обучения нейронной сети на точность работы нейронной сети
59	Адаптивный выбор скорости обучения нейронных сетей
60	Коррекция весовых коэффициентов в процессе обучения нейронных сетей

9. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

9.1. Паспорт фонда оценочных средств

№ п/п	Контролируемые разделы (темы) дисциплины	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
1	Основные понятия принципы работы хранилищ данных	ОПК-5, ПК-3, ПК-4	Собеседование
2	Основные понятия принципы работы OLAP-систем	ОПК-5, ПК-3, ПК-4	Собеседование
3	Кластеризация при интеллектуальном анализе данных	ОПК-5, ПК-3, ПК-4	Отчет по лабораторной работе
4	Классификация при интеллектуальном анализе данных	ОПК-5, ПК-3, ПК-4	Собеседование
5	Классификация на основе алгоритмов ID3, C4.5	ОПК-5, ПК-3, ПК-4	Отчет по лабораторной работе
6	Классификация на основе алгоритма CART	ОПК-5, ПК-3, ПК-4	Отчет по лабораторной работе
7	Регрессионный анализ при интеллектуальном анализе данных	ОПК-5, ПК-3, ПК-4	Собеседование
8	Регрессионный анализ на основе нейронных сетей	ОПК-5, ПК-3, ПК-4	Отчет по лабораторной работе
9	Регрессионный анализ на основе алгоритма CART	ОПК-5, ПК-3, ПК-4	Отчет по лабораторной работе
10	Аффинитивный анализ при интеллектуальном анализе данных	ОПК-5, ПК-3, ПК-4	Собеседование
11	Аффинитивный анализ на основе алгоритма Apriori	ОПК-5, ПК-3, ПК-4	Отчет по лабораторной работе

9.2. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

9.2.1. Комплект вопросов для проведения собеседования

Вопросы для собеседования

1. Охарактеризуйте этапы выполнения алгоритма k-means.
2. Какими данными необходимо обладать для запуска алгоритма k-means?
3. Какие метрики определения расстояний применяются в алгоритме k-means?
4. В чем заключается рекурсивность алгоритма k-means.
5. По каким параметрам производится остановка выполнения алгоритма k-means?
6. Охарактеризуйте этапы выполнения алгоритма ID3.
7. Какими данными необходимо обладать для запуска алгоритма ID3?
8. По какому критерию определяется условие разбиения в текущем узле дерева решений построенного с помощью алгоритма ID3?
9. В чем заключается рекурсивность алгоритма ID3.
10. По каким параметрам производится остановка выполнения алгоритма ID3?
11. Охарактеризуйте этапы выполнения алгоритма C4.5.
12. Какими данными необходимо обладать для запуска алгоритма C4.5?
13. По какому критерию определяется условие разбиения в текущем узле дерева решений построенного с помощью алгоритма C4.5?
14. В чем заключается рекурсивность алгоритма C4.5.
15. По каким параметрам производится остановка выполнения алгоритма C4.5?
16. В чем отличия алгоритмов ID3 и C4.5?
17. Охарактеризуйте этапы выполнения алгоритма CART.
18. Какими данными необходимо обладать для запуска алгоритма CART?
19. По какому критерию определяется условие разбиения в текущем узле дерева решений построенного с помощью алгоритма CART?
20. В чем заключается рекурсивность алгоритма CART.
21. По каким параметрам производится остановка выполнения алгоритма CART?
22. В чем отличия алгоритмов CART и C4.5?
23. Охарактеризуйте понятие регрессионный анализ.
24. Опишите общий алгоритм работы с нейронными сетями.
25. Какие существуют типы нейронных сетей?
26. Назовите рекомендации по выбору структуры нейронных сетей в зависимости от решаемой задачи.
27. Охарактеризуйте понятия регрессионного анализа данных.
28. Назовите причины, по которым алгоритмы ID3 и C4.5 не могут применяться для регрессионного анализа данных?

29. По каким параметрам производится остановка выполнения алгоритма CART?
30. Охарактеризуйте понятие аффинитивного анализа
31. Охарактеризуйте алгоритм Apriori.
32. Какие параметры может задавать аналитик для управления процессом выполнения алгоритма Apriori.
33. Какие объективные и субъективные метрики применяются для оценки ассоциативных правил?
34. Что является результатом выполнения алгоритма Apriori?

Процедура оценивания

Основой для определения оценки на собеседовании служит уровень усвоения студентами материала, предусмотренного рабочей программой дисциплины.

Критерии оценки:

- отметка «зачтено» ставится студенту, который не продемонстрировал знание материала, ориентируется в изученном материале.
- отметка «не зачтено» ставится студенту, который не продемонстрировал знание материала.

9.2.2. Комплект отчетов по лабораторным работам

Лабораторная работа № 1. Кластеризация при интеллектуальном анализе данных

Цель: изучить основы применения алгоритма k-means для решения задач кластеризации при интеллектуальном анализе данных.

Содержание задания.

1. В соответствии с номером варианта сгенерировать данные, на которых будет тестироваться работа алгоритма k-means.

Варианты заданий:

- 1 Вариант: У записей должно быть 3 параметра. Первый изменяется в диапазоне [0,1; 5], второй параметр – в диапазоне [0,1; 3], третий параметр может принимать значения – 10%, 20%, 80%, 90%.
- 2 Вариант: У записей должно быть 3 параметра. Первый изменяется в диапазоне [0,01; 1], второй параметр – в диапазоне [1; 300], третий параметр может принимать значения – Самара, Тольятти, Чапаевск.
- 3 Вариант: Первый изменяется в диапазоне [0; 1], второй параметр – в диапазоне [-2; 2], третий параметр может принимать значения – да или нет.
- 4 Вариант: Первый изменяется в диапазоне [-10; 1], второй параметр – в диапазоне [1; 2], третий параметр может принимать значения – отрицательное значение, положительное значение.
3. На любом известном объектно-ориентированном языке программирования разработать программу, на практике реализующую выполнение кластериза-

ции сгенерированных данных по алгоритму k-means с применением не менее двух метрик (например, Евклидова расстояние и расстояние Манхэттена). В программе должна быть предусмотрена возможность фиксирования после каждой итерации алгоритма суммы квадратов ошибок и координат центроидов.

4. Ответить на контрольные вопросы

1. Охарактеризуйте этапы выполнения алгоритма k-means.
2. Какими данными необходимо обладать для запуска алгоритма k-means?
3. Какие метрики определения расстояний применяются в алгоритме k-means?
4. В чем заключается рекурсивность алгоритма k-means.
5. По каким параметрам производится остановка выполнения алгоритма k-means?

Лабораторная работа № 2. Классификация на основе алгоритмов ID3, C4.5

Цель: изучить основы применения алгоритма ID3, C4.5 для решения задач классификации при интеллектуальном анализе данных.

Содержание задания.

1. В соответствии с номером варианта сгенерировать данные, на которых будет тестироваться работа алгоритма ID3.

Варианты заданий:

- 1 Вариант: Первый атрибут может принимать значения: А, В, С, D, Е. Второй атрибут может принимать значения: да, нет. Третий атрибут может принимать значения из диапазона [0, 600]. Четвертый атрибут может принимать значения из диапазона [0, 0.9]. Пятый атрибут может принимать значения: 6, 7, 10. Выходной атрибут может принимать значения С1, С2, С3, С4, С5. Набор данных должен содержать не менее 40 записей.
- 2 Вариант: Первый атрибут может принимать значения: блондинка, брюнетка, шатенка, рыжая. Второй атрибут может принимать значения: да, нет. Третий атрибут может принимать значения из диапазона [160, 185]. Четвертый атрибут может принимать значения 0, 1, 2, 3. Пятый атрибут может принимать значения: 80, 90, 100. Выходной атрибут может принимать значения С1, С2, С3, С4. Набор данных должен содержать не менее 40 записей.
- 3 Вариант: Первый атрибут может принимать значения: стрелок1, стрелок2, стрелок3, стрелок4, стрелок5. Второй атрибут может принимать значения: нарезной, гладкий. Третий атрибут может принимать значения из диапазона [50, 95]. Четвертый атрибут может принимать значения 100, 150, 200. Пятый атрибут может принимать значения: [500, 800]. Выходной атрибут может принимать значения С1, С2, С3, С4. Набор данных должен содержать не менее 40 записей.

- 4 Вариант: Первый атрибут может принимать значения: среднее, средне-специальное, высшее не оконченное, высшее, два высших. Второй атрибут может принимать значения из диапазона [18, 50]. Третий атрибут может принимать значения да, нет. Четвертый атрибут может принимать значения 0, 1, 2, 3. Четвертый атрибут может принимать значения из диапазона [50, 100]. Выходной атрибут может принимать значения C1, C2, C3, C5. Набор данных должен содержать не менее 40 записей.
2. На любом известном объектно-ориентированном языке программирования разработать программу, на практике реализующую выполнение алгоритма ID3 с использованием сгенерированных данных.
 3. Ответить на вопросы для контроля
 1. Охарактеризуйте этапы выполнения алгоритма ID3.
 2. Какими данными необходимо обладать для запуска алгоритма ID3?
 3. По какому критерию определяется условие разбиения в текущем узле дерева решений построенного с помощью алгоритма ID3?
 4. В чем заключается рекурсивность алгоритма ID3.
 5. По каким параметрам производится остановка выполнения алгоритма ID3?
 4. На любом известном объектно-ориентированном языке программирования разработать программу на практике реализующую выполнение алгоритма C4.5 с использованием сгенерированных данных.
 5. Сравнить построенные деревья принятия решений по алгоритму ID3 и C4.5
 6. Ответить на вопросы для контроля.
 1. Охарактеризуйте этапы выполнения алгоритма C4.5.
 2. Какими данными необходимо обладать для запуска алгоритма C4.5?
 3. По какому критерию определяется условие разбиения в текущем узле дерева решений построенного с помощью алгоритма C4.5?
 4. В чем заключается рекурсивность алгоритма C4.5.
 5. По каким параметрам производится остановка выполнения алгоритма C4.5?
 6. В чем отличия алгоритмов ID3 и C4.5?

Лабораторная работа № 3. Классификация на основе алгоритма CART

Цель: изучить основы применения алгоритма CART для решения задач классификации при интеллектуальном анализе данных.

Содержание задания:

1. Данные, на которых будет проводиться тестирование алгоритма CART необходимо взять из предыдущей лабораторной работы.
2. На любом известном объектно-ориентированном языке программирования разработать программу на практике реализующую выполнение алгоритма CART с использованием сгенерированных данных.
3. Сравнить построенные деревья принятия решений по алгоритму ID3, C4.5, CART
4. Ответить на вопросы для контроля.
 1. Охарактеризуйте этапы выполнения алгоритма CART.

2. Какими данными необходимо обладать для запуска алгоритма CART?
3. По какому критерию определяется условие разбиения в текущем узле дерева решений построенного с помощью алгоритма CART?
4. В чем заключается рекурсивность алгоритма CART.
5. По каким параметрам производится остановка выполнения алгоритма CART?
6. В чем отличия алгоритмов CART и C4.5?

Лабораторная работа № 4. Регрессионный анализ на основе нейронных сетей

Цель: изучить основы применения нейронных сетей для решения задач регрессии при интеллектуальном анализе данных.

Содержание задания:

1. Данные, на которых будет проводиться тестирование нейронных сетей необходимо сгенерировать следующим образом. Выбрать функцию с двумя независимыми переменными $y = f(x_1, x_2)$. При разных значениях x_1, x_2 рассчитать значения y . Принять значения x_1, x_2 – входными сигналами нейронной сети, а y – выходным сигналом.
2. На любом известном объектно-ориентированном языке программирования разработать программу на практике реализующую выполнение простейшей нейронной сети с использованием сгенерированных данных.
3. Построить контурную диаграмму исходной функции $y = f(x_1, x_2)$ (например в среде Mathcad) . Построить аналогичную контурную диаграмму по данным генерируемых нейронной сетью. Сравнить обе контурных диаграммы.
4. Ответить на вопросы для контроля.
 1. Охарактеризуйте понятие регрессионный анализ.
 2. Опишите общий алгоритм работы с нейронными сетями.
 3. Какие существуют типы нейронных сетей?
 4. Назовите рекомендации по выбору структуру нейронных сетей в зависимости от решаемой задачи.

Лабораторная работа № 5. Регрессионный анализ на основе алгоритма CART

Цель: изучить основы применения алгоритма CART для решения задач регрессии при интеллектуальном анализе данных.

Содержание задания:

1. Данные, на которых будет проводиться тестирование алгоритма CART необходимо сгенерировать следующим образом. Выбрать функцию с двумя независимыми переменными $y = f(x_1, x_2)$. При разных значениях x_1, x_2 рассчитать значения y . Принять значения x_1 – значениями первого атрибута, значения x_2 – значениями второго атрибута, а y – метками класса.
2. На любом известном объектно-ориентированном языке программирования разработать программу на практике реализующую выполнение алгоритма CART с использованием сгенерированных данных.

3. Построить контурную диаграмму исходной функции $y = f(x_1, x_2)$ (например в среде Mathcad) . Построить аналогичную контурную диаграмму по правилам выделенным из построенного дерева решений. Сравнить обе контурных диаграммы.

4. Ответить на вопросы для контроля.

1. Охарактеризуйте понятия регрессионного анализа данных.
2. Охарактеризуйте этапы выполнения алгоритма CART.
3. Назовите причины, по которым алгоритмы ID3 и C4.5 не могут применяться для регрессионного анализа данных?
4. В чем заключается рекурсивность алгоритма CART.
5. По каким параметрам производится остановка выполнения алгоритма CART?

Лабораторная работа № 6. Аффинитивный анализ на основе алгоритма Apriori

Цель: изучить основы применения алгоритма Apriori для решения задач аффинитивного анализа при интеллектуальном анализе данных.

Содержание задания:

1. Выбрать предметную область, в которой возможно применение аффинитивного анализа и сгенерировать не менее 100 транзакций, содержащих не менее 25 различных предметов.
2. На любом известном объектно-ориентированном языке программирования разработать программу, на практике реализующую выполнение алгоритма Apriori с использованием сгенерированных данных.
3. Ответить на вопросы для контроля.
 1. Охарактеризуйте понятие аффинитивного анализа
 2. Охарактеризуйте алгоритм Apriori.
 3. Какие параметры может задавать аналитик для управления процессом выполнения алгоритма Apriori.
 4. Какие объективные и субъективные метрики применяются для оценки ассоциативных правил?
 5. Что является результатом выполнения алгоритма Apriori?

Требования к оформлению

Отчет должен содержать подробное описание (включая иллюстративный материал) последовательности действий проделанных студентом для выполнения заданий. Оформление отчета должно соответствовать методическому указанию рекомендациям, изложенным учебно-методическом пособии [Очеповский А.В. Общие требования по выполнению и оформлению контрольных, курсовых и выпускных квалификационных работ : Учебно-методическое пособие. – Тольятти : ТГУ, 2015. 78 с.].

Процедура оценивания

Оценка за лабораторные работы выставляется на основе письменного отчета студента. Преподаватель может потребовать студента исправить замечания по оформлению или содержанию отчета по лабораторной работе.

Критерии оценки:

- оценка «зачтено» ставится студенту, который продемонстрировал результаты выполнения лабораторной работы, соответствующие поставленным задачам, и предоставил отчет, оформленный должным образом и содержащий краткое описание полученных результатов
- оценка «не зачтено» ставится студенту, который не продемонстрировал результаты выполнения лабораторной работы или не представил по ней отчет или представленный отчет не соответствует требованиям по оформлению.

10. Образовательные технологии и методические указания по освоению дисциплины (учебного курса)

В рамках изучения дисциплины «Интеллектуальный анализ на основе хранилищ данных» предусмотрено использование следующих образовательных технологий:

- технология традиционного обучения: лекции и практические занятия, самостоятельная работа.

Рекомендации по подготовке к лекционным занятиям

Изучение дисциплины требует систематического и последовательного накопления знаний, следовательно, пропуски отдельных тем не позволяют глубоко освоить предмет.

В ходе лекционных занятий задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений, разрешения спорных ситуаций.

Студент может дополнить список использованной литературы современными источниками, не представленными в списке рекомендованной литературы, и в дальнейшем использовать собственные подготовленные учебные материалы при написании курсовых и выпускных квалификационных работ.

Рекомендации по подготовке к практическим занятиям

Студентам следует:

- при подготовке к практическим занятиям следует обязательно использовать не только лекции, учебную литературу, но и другие источники;
- в начале занятий задать преподавателю вопросы по материалу, вызвавшему затруднения в его понимании и освоении при решении задач, заданных для самостоятельного решения;
- на занятии доводить каждую задачу до окончательного решения, демонстрировать понимание проведенных расчетов (анализов, ситуаций), в случае затруднений обращаться к преподавателю.

Тема 1. Основные понятия принципы работы хранилищ данных

Учебные вопросы:

1. Концепция хранилища данных
2. Организация хранилища данных
3. Очистка данных
4. Анализ данных из хранилища данных
5. Формулирование выводов

Изучив данную тему, студент должен:

иметь представление о назначении хранилищ данных и способах представления в них информации;

знать:

- понятие хранилище данных;
- понятие витрины данных;
- способы организации хранилища данных;
- способы очистки данных;

уметь:

- на основе базы данных формировать хранилище данных;

владеть навыками:

- пользования учебной и монографической литературой;
- формирования SQL-запросов для формирования хранилища данных.

Методические рекомендации по изучению темы

При освоении темы необходимо:

- изучить учебный материал;
- акцентировать внимание на понятии хранилище данных;
- понять способы формирования хранилищ данных;
- понять причины, по которым анализ данных удобнее производить с использованием хранилища данных

Тема 2. Основные понятия принципы работы OLAP-систем

Учебные вопросы:

1. Многомерная модель данных
2. Определение OLAP-систем
3. Концептуальное многомерное представление
4. Архитектура OLAP-систем

Изучив данную тему, студент должен:

иметь представление о назначении, структуре и основных выполняемых задачах OLAP- систем;

знать:

- назначение OLAP систем
- структуру OLAP систем
- задачи, решаемые с помощью OLAP систем

уметь:

- выделять задачи, присущие OLAP системам

владеть навыками

- постановки требований к OLAP задачам

Методические рекомендации по изучению темы

При освоении темы необходимо:

- изучить учебный материал темы;

- акцентировать внимание на задачах и структуре OLAP систем;
- необходимо понять 12 правил Кодда и обратить внимание о наличии дополнительных правил Кодда;
- понять различия OLAP-систем: MOLAP, ROLAP и HOLAP

Тема 3. Кластеризация при интеллектуальном анализе

Учебные вопросы:

1. понятие кластеризации данных
2. общие сведения об алгоритме k-means
3. математический аппарат алгоритма k-means
4. метрики, применяемые в алгоритме k-means
5. вопрос сходимости алгоритма к локальному решению

Изучив данную тему, студент должен:

иметь представление о назначении кластеризации данных и математическом аппарате алгоритма k-means;

знать:

- понятие кластеризации данных
- математический аппарат алгоритма k-means
- проблему сходимости алгоритма к локальному решению

уметь:

• анализировать решаемую задачу для оценки возможности ее решения с использованием алгоритма k-means

владеть навыками

- применения алгоритма k-means для решения прикладных задач

Методические рекомендации по изучению темы

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на понятии кластеризации данных;
- акцентировать внимание на математическом аппарате алгоритма k-means;
- понять, что получаемая с помощью k-means кластерная структура не всегда является оптимальной

Тема 4. Классификация при интеллектуальном анализе данных

Учебные вопросы:

1. понятие классификации данных
2. задачи классификации
3. проблемы классификации данных
4. обзор интеллектуальных алгоритмов для классификации данных

Изучив данную тему, студент должен:

*иметь представление о назначении и задачах классификации данных;
знать:*

- назначение классификации данных
- задачи, решаемые с помощью классификации данных

уметь:

- выделять задачи, которые можно решить с помощью классификации данных

владеть навыками

- анализа задач по классификации данных

Методические рекомендации по изучению темы

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на задачах, решаемых с помощью классификации данных;
- акцентировать внимание на различия между классификацией и кластеризацией данных;

Тема 5. Классификация на основе алгоритмов ID3, C4.5

Учебные вопросы:

1. общие сведения об алгоритме ID3
2. математический аппарат алгоритма ID3
3. пример использования алгоритма ID3
4. общие сведения об алгоритме C4.5
5. математический аппарат алгоритма C4.5
6. пример использования алгоритма C4.5

Изучив данную тему, студент должен:

иметь представление о математическом аппарате алгоритмов ID3, C4.5 и их назначении;

знать:

- назначение алгоритмов ID3 и C4.5
- различия между алгоритмами ID3 и C4.5

уметь:

- применять алгоритмы ID3 и C4.5 для решения практических задач

владеть навыками

- математического моделирования деревьев принятия решений с помощью алгоритмов ID3, C4.5

Методические рекомендации по изучению темы

При освоении темы необходимо:

- изучить учебный материал темы;

- акцентировать внимание на различиях между алгоритмами ID3 и C4.5
- акцентировать внимание на проблеме построения оптимальных классификационных моделей с помощью алгоритмов ID3 и C4.5

Тема 6. Классификация на основе алгоритма CART

Учебные вопросы:

1. общие сведения об алгоритме CART
2. математический аппарат алгоритма CART
3. пример использования алгоритма CART

Изучив данную тему, студент должен:

иметь представление о математическом аппарате алгоритма CART и его назначении;

знать:

- назначение алгоритма CART.

уметь:

- применять алгоритм CART для решения практических задач

владеть навыками

- математического моделирования деревьев принятия решений с помощью алгоритма CART

Методические рекомендации по изучению темы

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на различиях между алгоритмами CART, ID3 и C4.5
- акцентировать внимание на критерий оптимальности разбиения множества решений в каждом из узлов классификационной модели.

Тема 7. Регрессионный анализ при интеллектуальном анализе данных

Учебные вопросы:

1. понятие регрессионного анализа данных
2. задачи регрессионного анализа
3. проблемы регрессионного анализа данных
4. обзор интеллектуальных алгоритмов для регрессионного анализа данных

ных

Изучив данную тему, студент должен:

иметь представление о назначении и задачах регрессионного анализа данных;

знать:

- назначение регрессионного анализа данных
 - задачи, решаемые с помощью регрессионного анализа данных
- уметь:*
- выделять задачи, которые можно решить с помощью регрессионного анализа данных

владеть навыками

- анализа задач по регрессионному анализу данных

Методические рекомендации по изучению темы

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на задачах, решаемых с помощью регрессионного анализа данных;
- акцентировать внимание на различия между регрессионным анализом, классификацией и кластеризацией данных;

Тема 8. Регрессионный анализ на основе нейронных сетей

Учебные вопросы:

1. общие сведения об нейронных сетях
2. математический аппарат нейронных сетей
3. пример использования нейронных сетей

Изучив данную тему, студент должен:

иметь представление о проведении регрессионного анализа данных с помощью нейронных сетей;

знать:

- назначение нейронных сетей

уметь:

- использовать нейронные сети для решения задач регрессионного анализа

владеть навыками

- математического моделирования искусственных нейронных сетей.

Методические рекомендации по изучению темы

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на способы обучения нейронных сетей;
- акцентировать внимание на требованиях к обучающей выборке данных, применяемой для настройки нейронной сети;

Тема 9. Регрессионный анализ на основе алгоритма CART

Учебные вопросы:

1. математический аппарат алгоритма CART
2. пример использования алгоритма CART для решения задач регрессионного анализа данных

Изучив данную тему, студент должен:

3. *иметь представление* о возможности применения алгоритма CART для решения задач регрессионного анализа данных

знать:

- назначение алгоритма CART.

уметь:

- применять алгоритм CART для решения практических задач

владеть навыками

- математического моделирования деревьев принятия решений с помощью алгоритма CART

Методические рекомендации по изучению темы

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на требованиях к обучающей выборке данных, применяемой для построения дерева принятия решения с использованием алгоритма CART.

Тема 10. Аффинитивный анализ при интеллектуальном анализе данных

Учебные вопросы:

1. понятие аффинитивного анализа данных
2. задачи аффинитивного анализа данных
3. проблемы аффинитивного анализа данных
4. обзор интеллектуальных алгоритмов для аффинитивного анализа данных

Изучив данную тему, студент должен:

иметь представление о назначении и задачах аффинитивного анализа данных;

знать:

- назначение аффинитивного анализа данных
- задачи, решаемые с помощью аффинитивного анализа данных
- *уметь:*

- выделять задачи, которые можно решить с помощью аффинитивного анализа данных

владеть навыками

- анализа задач по поиску причинно-следственных связей

Методические рекомендации по изучению темы

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на задачах, решаемых с помощью аффинитивного анализа данных;
- акцентировать внимание на различия между классификацией, кластеризацией и аффинитивным анализом данных;

Тема 11. Аффинитивный анализ на основе алгоритма Apriori

Учебные вопросы:

1. общие сведения об алгоритме Apriori
2. математический аппарат алгоритма Apriori
3. пример использования алгоритма Apriori

Изучив данную тему, студент должен:

иметь представление о математическом аппарате алгоритма Apriori и его назначении;

знать:

- назначение алгоритма Apriori.

уметь:

- применять алгоритм Apriori для решения практических задач
- владеть навыками*
- математического моделирования алгоритма Apriori

Методические рекомендации по изучению темы

При освоении темы необходимо:

- изучить учебный материал темы;
- акцентировать внимание на объективных и субъективных метриках, применяемых в алгоритме Apriori.

11. Учебно-методическое и информационное обеспечение дисциплины (учебного курса)

11.1. Обязательная литература

№ п/п	Библиографическое описание	Тип (учебник, учебное пособие, учебно-методическое пособие, практикум, др.)	Количество в библиотеке
1.	Сысоев Д. В. Введение в теорию искусственного интеллекта [Электронный ресурс] : учеб. пособие / Д. В. Сысоев, О. В. Курипта, Д. К. Проскурин. - Воронеж : ВГАСУ : ЭБС АСВ, 2014. - 170 с. - ISBN 978-5-89040-498-5.	учебное пособие	ЭБС "IPRbooks"

11.2. Дополнительная литература и учебные материалы (аудио-, видеопособия и др.)

№ п/п	Библиографическое описание	Тип (учебник, учебное пособие, учебно-методическое пособие, практикум, аудио-, видеопособия и др.)	Место хранения (методический кабинет кафедры, городские библиотеки и др.)
1.	Павлов С. Н. Системы искусственного интеллекта [Электронный ресурс] : учеб. пособие. В 2-х ч. Ч. 1 / С. Н. Павлов ; Томский государственный университет систем управления и радиоэлектроники. - Томск : Эль Контент, 2011. - 174 с. : ил. - ISBN 978-5-4332-0013-5.	учебное пособие	ЭБС "IPRbooks"
2.	Павлов С. Н. Системы искусственного интеллекта [Электронный ресурс] : учеб. пособие. В 2-х ч. Ч. 2 / С. Н. Павлов ; Томский государственный университет систем управления и радиоэлектроники. - Томск : Эль Контент, 2011. - 193 с. : ил. - ISBN 978-5-4332-0014-2.	учебное пособие	ЭБС "IPRbooks"

СОГЛАСОВАНО

Директор научной библиотеки

(подпись)

А.М. Асаева
(И.О. Фамилия)

«__» _____ 20__ г.

11.3. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"

Springer International Publishing, Part of Springer Science+Business Media [Электронный ресурс] – Springer International Publishing AG, 2016. — Режим доступа к журн.: <http://link.springer.com> . – Загл. с экрана

11.4. Перечень программного обеспечения

№ п/п	Наименование ПО	Количество лицензий	Реквизиты договора (дата, номер, срок действия)
1	Microsoft Access	1	Договор 652/2014 от 07.07.2014, бессрочный
2	Microsoft Visual Studio	1	Договор 652/2014 от 07.07.2014, бессрочный
3	MathCAD	15	Акт п/п от 21.07.09 (Гос. Контракт 487 от 28.05.09), бессрочный

11.5. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

№ п/п	Наименование оборудованных учебных кабинетов, лабораторий, мастерских и др. объектов для проведения практических и лабораторных занятий	Перечень основного оборудования	Фактический адрес учебных кабинетов, лабораторий, мастерских и др.	Площадь, м ²	Количество посадочных мест
1.	Компьютерный класс	Компьютер (монитор 19", системный блок Pentium (R) Dual-Core E5500 2,8 GHz / 4 Gb / 500 Gb) - 12 шт, стол ученический - 7 шт., стол компьютерный - 12шт., стол преподавательский -1 шт., стулья -35шт. Доска аудиторная(меловая) - 1 шт.	445667 Самарская область, г.Тольятти, Центральный р-н, ул. Белорусская, д.16В, УЛК-401	49,5	24

2.	Лекционная аудитория	80 посадочных мест. Стол ученический двухместный (моно-блок) – 41 шт., доска аудиторная 3-х секционная (меловая)-1 шт., стол преподавательский -1 шт., стул-2 шт., проектор Асер	445667 Самарская область, г.Тольятти, Центральный р-н, ул. Белорусская, д.16В, УЛК-418	90,6	80
3.	Помещение для самостоятельной работы студентов.	Стол ученический-26 шт., стул-26 шт., компьютер с выходом в сеть интернет-16 шт.	445020 Самарская область, г.Тольятти, Центральный р-н, ул. Белорусская, д.14, Г-401	16	84,2