

Белорусский государственный университет



УТВЕРЖДАЮ
Проректор по учебной работе и
образовательным инновациям
О. И. Чуприс
2018 г.

Регистрационный № УД- 6248 /уч.

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Учебная программа учреждения высшего образования по учебной дисциплине для специальностей первой степени высшего образования:

**1-31 03 06 Экономическая кибернетика
направления специальности**

1-31 03 06 - 01 Экономическая кибернетика (математические методы и компьютерное моделирование в экономике)

2018 г.

Учебная программа составлена на основе образовательного стандарта высшего образования ОСВО 1-31 03 06-2013 и учебного плана G31-166/уч. от 30.05.2013, G31и-191/уч. от 30.05.2013.

СОСТАВИТЕЛЬ:

С.Н. Сталевская, доцент кафедры математического моделирования и анализа данных факультета прикладной математики и информатики Белорусского государственного университета, кандидат физико-математических наук.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой математического моделирования и анализа данных Белорусского государственного университета (протокол № 13 от 29.03.2018);

Научно-методическим Советом Белорусского государственного университета (протокол № 7 от 13 июля 2018 г.).

Тимо / Бодени И.А., зав. кафедрой ММАД /

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи дисциплины

Целью дисциплины «Интеллектуальный анализ данных» является изучение методов и алгоритмов анализа данных большого объема, получаемых с различных практических сфер человеческой деятельности.

В рамках поставленной цели *задача* учебной дисциплины состоят в следующем:

- изучение методов и алгоритмов анализа данных большого объема
- формирование практических умений и навыков работы с большими объемами данных.
- приобретение практических навыков для решения прикладных задач статистического анализа больших объемов данных.

Принципы изложения материала

Интеллектуальный анализ данных (англ. Data Mining) — выявление скрытых закономерностей или взаимосвязей между переменными в больших массивах необработанных данных. Подразделяется на задачи классификации, моделирования и прогнозирования и другие. Термин «Data Mining» введен Григорием Пятецким-Шапиро в 1989 году.

Английский термин «Data Mining» не имеет однозначного перевода на русский язык (добыча данных, вскрытие данных, информационная проходка, извлечение данных/информации) поэтому в большинстве случаев используется в оригинале. Наиболее удачным непрямым переводом считается термин «интеллектуальный анализ данных» (ИАД).

ИАД включает методы и модели статистического анализа и машинного обучения, дистанцируясь от них в сторону автоматического анализа данных. Инструменты ИАД позволяют проводить анализ данных предметными специалистами (аналитиками), не владеющими соответствующими математическими знаниями.

Методы ИАД приобретают все большую популярность в качестве инструмента для анализа информации с различных сфер человеческой деятельности, особенно в тех случаях, когда предполагается, что из имеющихся данных можно будет извлечь знания для принятия решений в условиях неопределенности. Хотя в последнее время возрос интерес к разработке новых методов анализа данных, специально предназначенных для сфер экономики и биологии, в целом системы добычи данных по-прежнему основываются на классических принципах разведочного анализа данных (РАД) и построения моделей и используют те же подходы и методы.

Специфика современных требований к обработке данных следующая:

- данные имеют неограниченный объем,
- данные являются разнородными (количественными, качественными, текстовыми),
- результаты должны быть конкретны и понятны,

- инструменты для обработки сырых данных должны быть просты в использовании.

В основу современной технологии ИАД положена концепция шаблонов (паттернов), отражающих *фрагменты* многоаспектных взаимоотношений в данных. Эти шаблоны представляют собой закономерности, свойственные *подвыборкам данных*, которые могут быть компактно выражены в понятной человеку форме. Поиск шаблонов производится методами, не ограниченными рамками априорных предположений о структуре выборке и виде распределений значений анализируемых показателей.

Важное положение ИАД — нетривиальность разыскиваемых шаблонов. Это означает, что найденные шаблоны должны отражать неочевидные, неожиданные регулярности в данных, составляющие так называемые скрытые знания. К обществу пришло понимание, что сырые данные содержат глубинный пласт знаний, при грамотной раскопке которого могут быть обнаружены настоящие самородки.

Учебная дисциплина «Интеллектуальный анализ данных» относится к циклу дисциплин специализации компоненты учреждения высшего образования.

Учебная программа составлена с учетом межпредметных связей с учебными дисциплинами. Так, основой для изучения дисциплины «Интеллектуальный анализ данных» являются дисциплины «Теория вероятностей и математическая статистика», «Эконометрика». Знания, полученные в результате изучения дисциплины, будут использованы при изучении дисциплин «Компьютерное прогнозирование», «Статистический анализ многомерных данных», а также способствовать успешному прохождению преддипломной практики и подготовки дипломной работы.

В результате изучения дисциплины студент должен:

-*знать*:

- понятия добычи данных, интеллектуального анализа данных, разведочного анализа данных;
- этапы интеллектуального анализа данных;
- методы подготовки данных, их первичной обработки, организации и хранения данных;
- методы уменьшения объема данных и Отбора значимых признаков;
- методы и средства графического анализа данных;
- методы оперативной аналитической обработки данных;
- вычислительные методы разведочного анализа данных;
- методы машинного обучения;
- эвристические методы интеллектуального анализа данных,

- *уметь* использовать методы, алгоритмы, а также компьютерные средства для анализа данных большой размерности.

Освоение учебной дисциплины «Интеллектуальный анализ данных» должно обеспечить формирование следующих академических социально-личностных и профессиональных компетенций:

академические компетенции:

- АК-1. Уметь применять базовые научно-теоретические знания для решения
 - теоретических и практических задач.
- АК-2. Владеть системным и сравнительным анализом.
- АК-3. Владеть исследовательскими навыками.

профессиональные компетенции:

- ПК-1. Работать с научно-технической, нормативно-справочной и специальной
 - литературой.
- ПК-4. Профессионально ставить задачи, выработать и принимать решения.
- ПК-8. Эксплуатировать, сопровождать и разрабатывать соответствующие программные компьютерные системы.
- ПК-12. Взаимодействовать со специалистами смежных профилей.
- ПК-16. Пользоваться глобальными информационными ресурсами.
- ПК-17. Владеть современными информационными технологиями и средствами телекоммуникаций.
- ПК-20. Выявлять закономерности в функционировании экономических систем на основе статистического анализа экономических и финансовых данных.
- ПК-21. Владеть методами эконометрического анализа и прогнозирования экономических систем и процессов.
- ПК-24. Владеть методами инвестиционного и финансового анализа, а также анализировать и прогнозировать динамику основных показателей финансовых рынков.
- ПК-25. Анализировать и прогнозировать конъюнктуру товарных рынков.

Структура содержания учебной дисциплины включает такие дидактические единицы, как темы (разделы), в соответствии с которыми разрабатываются и реализуются соответствующие лекционные и лабораторные занятия. Примерная тематика занятий приведена в информационно-методической части.

Дисциплина изучается в 6 семестре. Всего на освоение учебной дисциплины «Компьютерное прогнозирование» отведено 54 часов, в том числе 34 аудиторных часов, из них: лабораторные занятия – 30 часов, управляемая самостоятельная работа – 4 часа.

Трудоемкость учебной дисциплины составляет 1.5 зачетные единицы.

Форма текущей аттестации – зачет.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Введение

Введение в ИАД. Понятия добычи данных, интеллектуального анализа данных, разведочного анализа данных. Задачи, решаемые ИАД и РАД. Этапы ИАД. Примеры применения для практических сфер человеческой деятельности.

Раздел 2. Первичный анализ данных

Первичная подготовка и обработка данных. Подготовка данных в стандартной форме. Организация и хранение данных. Преобразование данных. Обработка пропущенных данных. Обработка зависимых во времени данных. Обработка текста.

Раздел 3. Уменьшение объема данных

Уменьшение объема данных. Отбор значимых признаков. Метод главных компонент. Отбор признаков с использованием деревьев решений. Определение необходимой размерности исследуемых данных (количества признаков и размера выборок)

Раздел 4. Методы и средства графического анализа данных

Методы визуального (графического) ИАД. Средства визуального ИАД. Использование средств визуализации для выявления закономерностей и верификации моделей.

Раздел 5. OLAP

Оперативная аналитическая обработка данных (OLAP).

Раздел 6. Вычислительные методы разведочного анализа данных

В данном разделе изучаются статистические методы разведочного анализа данных и методы многомерного разведочного анализа.

Статистические методы:

- анализа распределений,
- просмотр корреляционных,
- анализ многомерных таблиц частот;

Методы многомерного разведочного анализа:

- кластерный анализ,
- факторный анализ,
- анализ дискриминантных функций,
- многомерное шкалирование,

- логлинейный анализ,
- канонические корреляции,
- пошаговая линейная и нелинейная регрессия,
- анализ соответствий,
- анализ временных рядов,
- деревья классификаций.

Раздел 7. Методы машинного обучения

Деревья решений, методы кросс-проверки, метод опорных векторов, ядерные методы.

Раздел 8. Эвристические методы интеллектуального анализа данных

Эволюционные алгоритмы, генетическое программирование, алгоритмы оптимизации «муравья», параллельные методы разведочного анализа данных.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов			Количество часов в УСР	Форма контроля знаний
		Лекции	Семинарские занятия	Лабораторные занятия		
1.	Введение в РАД.			2		
1.1.	Понятия добычи данных, интеллектуального анализа данных, разведочного анализа данных. Задачи, решаемые ИАД и РАД. Этапы ИАД. Примеры применения для практических сфер человеческой деятельности.			2		Защита лабораторной работы
2.	Первичная подготовка и обработка данных			6		
2.1.	Подготовка данных в стандартной форме. Организация и хранение данных.			2		Защита лабораторной работы
2.2.	Преобразование данных. Обработка пропущенных данных. Обработка зависимых во времени данных.			2		Защита лабораторной работы
2.3.	Обработка текста.			2		Защита лабораторной работы
3.	Уменьшение объема данных			2		
3.1.	Отбор значимых признаков. Метод главных компонент. Отбор признаков с использованием деревьев решений. Определение необходимой размерности исследуемых данных (количества признаков и размера выборок)			2		Защита лабораторной работы
4.	Методы визуального (графического) ИАД			2		
4.1	Средства визуального ИАД. Использование средств визуализации для выявления закономерностей и верификации моделей.			2		Защита лабораторной работы. Контрольная работа №1.
5.	Оперативная аналитическая обработка данных (OLAP)				2	
5.1.	Оперативная аналитическая обработка данных (OLAP)				2	Устный опрос
6.	Вычислительные методы			6	2	

	разведочного анализа данных					
6.1	Статистические методы (анализа распределений, просмотр корреляционных, анализ многовходовых таблиц частот).			2		Защита лабораторной работы
6.2	Методы многомерного разведочного анализа: кластерный анализ, факторный анализ, анализ лискриминантных функций, многомерное шкалирование.			2	2	Защита лабораторной работы. Защита подготовленного реферата.
6.3	Логлинейный анализ, канонические корреляции, пошаговая линейная и нелинейная регрессия, анализ соответствий, анализ временных рядов и деревья классификаций.			2		Защита лабораторной работы
7.	Методы машинного обучения.			6		
7.1.	Деревья решений. Методы кросс-проверки.			2		Защита лабораторной работы
7.2.	Метод опорных векторов.			2		Защита лабораторной работы
7.3.	Ядерные методы.			2		Защита лабораторной работы. Контрольная работа №2.
8.	Эвристические методы интеллектуального анализа данных			6		
8.1.	Эволюционные алгоритмы. Генетическое программирование.			2		Защита лабораторной работы
8.2.	Алгоритмы оптимизации «муравья».			2		Защита лабораторной работы
8.3	Параллельные методы разведочного анализа данных.			2		Защита лабораторной работы
ИТОГО				34		

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Айвазян С.А. и др. Прикладная статистика. Первичный анализ данных.
2. Чубукова И.А. *Data Mining*. Электронный учебник. Intuit.
3. Abbass H.A., Sarker R.A., Newton C.S. *Data Mining: A Heuristic Approach*. Idea Group Publishing, 2002, 310 pages.
4. Han J., Kamber M. *Data mining. Concepts and techniques*.
5. Hand D., Mannila H., Smyth P. *Principles of data mining*.
6. Parr Rud O. *Data Mining Cookbook. Modeling Data for Marketing, Risk, and Customer Relationship Management*. John Wiley & Sons, Inc., 2001, 439 pages.

Перечень дополнительной литературы

1. Guo Y., Grossman R. (eds.) *High Performance Data Mining: Scaling Algorithms, Applications and Systems*. Kluwer Academic Publisher, 1999, 111 pages.
2. Mitra S., Acharya T. *Data mining. Multimedia, soft computing, and bioinformatics*.
3. Soukup T. and Davidson I. *Visual Data Mining: Techniques and Tools for Data Visualization and Mining*. John Wiley & Sons, 2002, 382 pages.
4. Suykens J.A.K., Horvath G., Basu S. *Advances in learning theory: methods, models and applications*.
5. Wong M.L., Cheung K.S. *Data mining using grammar based genetic programming and applications*.

Рекомендуемая тематика контрольных работ

- 1) Контрольная работа №1. *Метод главных компонент и снижение размерности*.
- 2) Контрольная работа №2. *Подготовка данных к анализу*.

Методические рекомендации по организации самостоятельной работы обучающихся

Для организации самостоятельной работы студентов магистратуры по учебной дисциплине следует использовать современные информационные технологии: разместить в сетевом доступе комплекс учебных и учебно-методических материалов (учебно-программные материалы, ссылки на учебные издания для теоретического изучения дисциплины, методические указания к лабораторным занятиям, материалы текущего контроля и текущей аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательных стандартов высшего

образования и учебно-программной документации, в т.ч. вопросы для подготовки к зачету, задания, тесты, вопросы для самоконтроля, тематика рефератов и др., список рекомендуемой литературы, информационных ресурсов и др.). Эффективность самостоятельной работы студентов магистратуры проверяется в ходе текущего и итогового контроля знаний. Для общей оценки качества усвоения студентами магистратуры учебного материала рекомендуется использование рейтинговой системы.

Перечень рекомендуемых средств диагностики

Для текущего контроля качества усвоения знаний студентами магистратуры используется следующий диагностический инструментарий:

1. Устная форма: устные опросы; защиты отчетов по домашним заданиям, при выполнении студентами магистратуры лабораторных работ; проведение коллоквиума; защита подготовленного студентом магистратуры реферата (рефераты используются для обобщения и систематизации учебного материала; в процессе подготовки реферата студент магистратуры мобилизует и актуализирует имеющиеся умения, приобретает самостоятельно новые знания, необходимые для раскрытия темы, сопоставляя разные позиции и точки зрения).

2. Письменная форма: письменные контрольные работы по отдельным темам учебной дисциплины.

Методика формирования итоговой оценки

Формой текущей аттестации по учебной дисциплине «Интеллектуальный анализ данных» учебным планом предусмотрен зачет.

При оценивании реферата внимание обращается на:

- содержание, корректность и последовательность изложения – 35%;
- релевантность и полноту раскрытия темы – 20 %;
- самостоятельность суждений – 35%;
- оформление – 10%.

Рекомендуется использовать рейтинговую оценку знаний студента магистратуры, дающую возможность проследить и оценить динамику процесса достижения целей обучения. Рейтинговая оценка предусматривает использование весовых коэффициентов для текущего контроля знаний и текущей аттестации студентов по дисциплине. Примерные весовые коэффициенты, определяющие вклад текущего контроля знаний в рейтинговую оценку:

- подготовка реферата – 15 %;
- работа на лабораторных занятиях – 35 %;
- контрольные работы – 25 %;
- коллоквиум – 25 %.

Итоговая оценка формируется на основе:

- 1) Правил проведения аттестации студентов (Постановление Министерства образования Республики Беларусь № 53 от 29 мая 2012г.);
- 2) Положение о рейтинговой системе оценки знаний по дисциплине в БГУ (Приказ ректора БГУ от 18.08.2015 № 382-ОД);
- 3) Критериев оценки знаний студентов (письмо Министерства образования от 22.12.2003).

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УВО

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Статистический анализ многомерных данных	Математического моделирования и анализа данных	нет	Оставить содержание учебной дисциплины без изменения, протокол № 13 от 29.03.2018 г.
Компьютерное прогнозирование	Математического моделирования и анализа данных	нет	Оставить содержание учебной дисциплины без изменения, протокол № 13 от 29.03.2018 г.

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ ПО
ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ**
на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
_____ (протокол № ____ от _____ 201_ г.)

Заведующий кафедрой

УТВЕРЖДАЮ
Декан факультета
