

УЧЕБНАЯ ПРОГРАММА.

Интеллектуальный анализ данных.

СОСТАВИТЕЛЬ.

Яцков Николай Николаевич, доцент кафедры системного анализа и компьютерного моделирования Белорусского государственного университета, кандидат физико-математических наук.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Введение в интеллектуальный анализ данных

Тема 1.1. Основные понятия дисциплины Интеллектуальный анализ данных (ИАД). Методы и задачи ИАД. Данные. Понятие данных. Типы шкал наборов данных. Типы наборов данных. Примеры практического применения ИАД. Медицина. Биоинформатика. Геология. Банковское дело. Торговля.

Раздел 2. Методы снижения размерности данных

Тема 2.1. Методы преобразования пространства признаков. Метод главных компонент. Метод главных координат. Факторный анализ.

Тема 2.2. Методы автоматического отбора информативных признаков.

Раздел 3. Кластерный анализ

Тема 3.1. Иерархические методы кластерного анализа. Основные элементы и этапы кластерного анализа. Расстояния между объектами данных. Математические характеристики кластера. Дендрограмма. Формула Ланса-Уильямса. Меры сходства кластеров данных. Агломеративный, дивизимный и гибридный кластерный анализ. Оценка значимости кластеров.

Тема 3.2. Неиерархические методы кластерного анализа. Алгоритмы на основе k -средних – Fuzzy k -средних, k -медоидов, PAM k , CLARA. Алгоритмы DBSCAN и спектральный.

Раздел 4. Методы классификации

Тема 4.1. k -ближайших соседей и байесовских сетей. Алгоритмы k -ближайших соседей. Методы V -кратного перекрестного контроля и bootstrap. Байесовская классификация. Метод опорных векторов.

Тема 4.2. Деревья решений. Методика «разделяй и властвуй». Критерии и функции качества разбиения узлов дерева. Индекс Джини. Энтропия. Ошибка классификации. Остановка построения дерева. Сокращение дерева. Извлечение правил из деревьев. Алгоритмы построения деревьев решений: ID3, покрытия, Conditional Inference Tree, CART, Random Forests.

Раздел 5. Нейронные сети

Тема 5.1. Нейронные сети. Нейрон. Нейронная сеть. Обучение нейронной сети, алгоритмы Хебба, адаптивный и обратного распространения ошибки. Нейронные сети Кохонена. Нейронные сети глубокого обучения.

Раздел 6. Стохастические методы поиска

Тема 6.1. Методы случайного поиска. Простой стохастический поиск. Случайный поиск. Метод Метрополиса. Метод имитации отжига. Эволюционные стратегии случайного поиска. Селекция. Скрещивание. Мутация. Генетические алгоритмы.

Раздел 7. Ассоциативные правила

Тема 7.1. Методы поиска ассоциативных правил. Ассоциативные правила. Алгоритмы поиска ассоциативных правил. Алгоритм Apriori. Алгоритм FPG.

Раздел 8. Методы визуализации данных и процесс ИАД

Тема 8.1. Методы визуализации многомерных данных.

Тема 8.2. Основные этапы процесса ИАД.

ТЕМАТИКА ЛЕКЦИЙ И ЛАБОРАТОРНЫХ ЗАНЯТИЙ

Тематика лекций (24 часа)

- 1) Введение в интеллектуальный анализ данных (2 часа).
- 2) Методы снижения размерности данных (4 часа).
- 3) Кластерный анализ (4 часа).
- 4) Методы классификации (4 часа).
- 5) Нейронные сети (4 часа).
- 6) Стохастические методы поиска (2 часа).
- 7) Ассоциативные правила (2 часа).
- 8) Методы визуализации данных и процесс ИАД (2 часа).

Тематика лабораторных занятий (40 часов)

- 1) Предварительный анализ больших наборов данных (4 часа).
- 2) Метод главных компонент (4 часа).
- 3) Метод главных координат (4 часа).
- 4) Иерархические методы кластерного анализа (4 часа).
- 5) Метод k-средних (4 часа).
- 6) Нейронные сети. Слой Кохонена (4 часа).
- 7) Спектральный метод кластерного анализа (4 часа).
- 8) Метод k-ближайших соседей (4 часа)
- 9) Метод опорных векторов (4 часа)
- 10) Стохастические методы поиска (4 часа).

ИНФОРМАЦИОННАЯ ЧАСТЬ

Перечень основной литературы

- 1) Яцков, Н. Н. Интеллектуальный анализ данных : пособие / Н. Н. Яцков. – Минск : БГУ, 2014. – 151 с.
- 2) Анализ больших данных : методические указания к лабораторным работам / Н. Н. Яцков, Е. В. Лисица. – Минск: БГУ, 2019. – 50 с.
- 3) Лагутин, М. Б. Наглядная математическая статистика [Электронный ресурс]: учеб. пособие / М. Б. Лагутин. – 5-е изд. (эл.). – Электрон. текстовые дан. (1 файл pdf : 475 с.). — М. : БИНОМ. Лаборатория знаний, 2015.
- 4) Анализ данных и процессов: учеб. пособие / А. А. Барсегян [и др.]. – 3-е изд., перераб. и доп. – СПб., 2009. – 512 с.
- 5) Николенко, С. Глубокое обучение. / С. Николенко, А. Кадурын, Е. Архангельская – СПб.: Питер, 2018. – 480 с.

Перечень дополнительной литературы

- 1) Лесковец, Ю. Анализ больших наборов данных / Ю. Лесковец, А. Раджараман, Д. Д. Ульман – Москва: ДМК Пресс, 2016. – 498 с.
- 2) Machine Learning Models and Algorithms for Big Data Classification / S. Suthaharan – Integrated Series in Information Systems 36, Springer Science+Business Media New York, 2016. – 359 pp.
- 3) Bishop, C. M., Pattern recognition and machine learning / C. M. Bishop. –New York: Springer, 2016. – 738 p.
- 4) Bramer, M. Principles of Data Mining / M. Bramer. – Third edition. London : Springer-Verlag London Ltd. 2016. – 526 p.
- 5) Aggarwal, C.C. Data Mining: The Textbook / C.C. Aggarwal. – Springer International Publishing Switzerland. 2015. 734 P.