

Аннотация

Дисциплина «Машинное обучение» входит в вариативную часть образовательной программы подготовки обучающихся по направлению 45.03.02 «Лингвистика» направленность «Теоретическая и прикладная лингвистика». Дисциплина реализуется кафедрой №63.

Дисциплина нацелена на формирование у выпускника

профессиональных компетенций:

ПК-21 «владение основными математико-статистическими методами обработки лингвистической информации с учетом элементов программирования и автоматической обработки лингвистических корпусов».

Содержание дисциплины охватывает круг вопросов, связанных с изучением некоторых алгоритмов обучения с учителем и без с использованием соответствующих функций библиотеки компьютерного зрения OpenCV и IPP.

Теория обучения машин (machine learning, машинное обучение) находится на стыке прикладной статистики, численных методов оптимизации, дискретного анализа, и за последние 50 лет оформилась в самостоятельную математическую дисциплину. Методы машинного обучения составляют основу ещё более молодой дисциплины — *интеллектуального анализа данных* (data mining).

В курсе рассматриваются основные задачи обучения по прецедентам: классификация, кластеризация, регрессия, понижение размерности. Изучаются методы их решения, как классические, так и новые, созданные за последние 10–15 лет. Упор делается на глубокое понимание математических основ, взаимосвязей, достоинств и ограничений рассматриваемых методов. Отдельные теоремы приводятся с доказательствами.

Все методы излагаются по единой схеме:

исходные идеи и эвристики;

их формализация и математическая теория;

описание алгоритма в виде слабо формализованного псевдокода;

анализ достоинств, недостатков и границ применимости;

пути устранения недостатков;

сравнение с другими методами.

примеры прикладных задач.

Данный курс расширяет и углубляет набор тем, рекомендованный международным стандартом **ACM/IEEE Computing Curricula 2001** по дисциплине «Машинное обучение и нейронные сети» (machine learning and neural networks) в разделе «Интеллектуальные системы» (intelligent systems).

Данная цель предполагает решение следующих задач:

1. Изучить основные идеи, лежащие в основе следующих алгоритмов классификации:

- машина опорных векторов;
- дерево решений;
- случайный лес;

- градиентный бустинг деревьев решений.
- 2. Изучить идеи метода центров тяжести (k-means) для кластеризации.
- 3. Рассмотреть прототипы функций и интерфейсы классов, реализующих перечисленные алгоритмы в библиотеке OpenCV.
- 4. Рассмотреть простые примеры использования указанного набора функций.
- 5. Разработать приложения для решения задач классификации и кластеризации рассмотренными методами.
- 6. Применить разработанное приложение для решения модельных задач и проанализировать полученные результаты.

В результате освоения дисциплины студенты должны иметь представление об основных идеях машинного обучения, знать алгоритмы классификации методов машинного обучения, уметь решать задачи классификации, кластеризации и разрабатывать приложения для решения задач классификации и кластеризации.

Преподавание дисциплины предусматривает следующие формы организации учебного процесса: лекции, практические занятия, самостоятельная работа студента, курсовое проектирование.

Программой дисциплины предусмотрены следующие виды контроля: текущий контроль успеваемости, промежуточная аттестация в форме экзамена.

Общая трудоемкость освоения дисциплины составляет 5 зачетных единиц, 180 часов.

Язык обучения по дисциплине «русский».