

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Локтионова Оксана Геннадьевна
Должность: проректор по учебной работе
Дата подписания: 12.10.2023 13:28:42
Уникальный программный ключ:
0b817ca911e6668abb13a5d426d39e5f1c11eabbf73e943df4a4851fda56d089

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра вычислительной техники

УТВЕРЖДАЮ

Проректор по учебной работе

О.Г. Локтионова

« 5 » 10

10.10.2023 г.



ВВЕДЕНИЕ В БОЛЬШИЕ ДАННЫЕ

Методические указания по выполнению практической работы
для студентов направления подготовки 09.04.01 по дисциплине
«Системы искусственного интеллекта»

Курск 2023

2

УДК 004.89

Составитель: Д.В. Титов

Рецензент

Доктор технических наук, профессор *Чернецкая И.Е.*

Введение в большие данные: методические указания по выполнению практической работы по дисциплине «Системы искусственного интеллекта» / Юго-Зап. гос. ун-т; сост.: Д.В. Титов. - Курск, 2023. – 8 с.

Методические указания по выполнению практической работы являются дополнением к конспекту лекций «Системы искусственного интеллекта» и содержат сведения, необходимые для выполнения работы.

Методические указания соответствуют рабочей программе дисциплины «Системы искусственного интеллекта» направления подготовки 09.04.01 Информатика и вычислительная техника.

Предназначены для студентов направления очной формы обучения.

Текст печатается в авторской редакции

Подписано в печать
Усл.печ.л. . Уч.-изд.л. . Тираж 50 экз. Заказ *1133* Бесплатно
Юго-Западный государственный университет.
305040, г.Курск, ул 50 лет Октября, 94

СОДЕРЖАНИЕ

Теоретическая часть	4
Задание к практической работе	7
Контрольные вопросы	8

Теоретическая часть

Машинное обучение (англ. machine learning, ML) — класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение за счёт применения решений множества сходных задач. Для построения таких методов используются средства математической статистики, численных методов, математического анализа, методов оптимизации, теории вероятностей, теории графов, различные техники работы с данными в цифровой форме.

Различают два типа обучения: индуктивное обучение и дедуктивное обучение.

Обучение по прецедентам, или индуктивное обучение, основано на выявлении эмпирических закономерностей в данных. Многие методы индуктивного обучения разрабатывались как альтернатива классическим статистическим подходам. Другие методы тесно связаны с извлечением информации (англ. information extraction, information retrieval), интеллектуальным анализом данных (data mining).

Дедуктивное обучение предполагает формализацию знаний экспертов и их перенос в компьютер в виде базы знаний. Дедуктивное обучение принято относить к области экспертных систем, поэтому термины машинное обучение и обучение по прецедентам можно считать синонимами.

Имеется множество объектов (ситуаций) и множество возможных ответов (откликов, реакций). Существует некоторая зависимость между ответами и объектами, но она неизвестна. Известна только конечная совокупность прецедентов — пар «объект, ответ», называемая обучающей выборкой. На основе этих данных требуется восстановить неявную зависимость, то есть построить алгоритм, способный для любого возможного входного объекта выдать достаточно точный классифицирующий ответ. Эта зависимость не обязательно выражается аналитически, и здесь нейросети реализуют принцип эмпирически формируемого решения. Важной особенностью при этом является способность обучаемой системы к обобщению, то есть к адекватному отклику на данные, выходящие за пределы имеющейся обучающей выборки. Для измерения точности ответов вводится оценочный функционал качества. Данная постановка является обобщением классических задач аппроксимации функций. В классических задачах аппроксимации объектами являются действительные числа или векторы. В реальных прикладных задачах входные данные об объектах могут быть неполными, неточными, нечисловыми, разнородными. Эти особенности приводят к большому разнообразию методов машинного обучения.

Раздел машинного обучения, с одной стороны, образовался в результате разделения науки о нейросетях на методы обучения сетей и виды топологий их архитектуры, с другой стороны — вобрал в себя методы математической статистики. Указанные ниже способы машинного обучения основаны на применении нейросетей, хотя существуют и другие методы, основанные на обучающей выборке — например, дискриминантный анализ, оперирующий обобщённой дисперсией и ковариацией наблюдаемой статистики, или байесовские классификаторы. Базовые виды нейросетей, такие как перцептрон и многослойный перцептрон (а также их модификации), могут обучаться как с учителем, так и без учителя, с подкреплением и самоорганизацией. Но некоторые нейросети и большинство статистических методов можно отнести только к одному из способов обучения. Поэтому, если нужно классифицировать методы машинного обучения в зависимости от способа обучения, то будет некорректным относить нейросети к определенному виду, правильнее было бы типизировать алгоритмы обучения нейронных сетей.

Способы машинного обучения

Обучение с учителем — для каждого прецедента задаётся пара «ситуация, требуемое решение»:

- искусственная нейронная сеть;
- глубокое обучение;
- метод коррекции ошибки;
- метод обратного распространения ошибки;
- метод опорных векторов.

Обучение без учителя — для каждого прецедента задаётся только «ситуация», требуется сгруппировать объекты в кластеры, используя данные о попарном сходстве объектов, и/или понизить размерность данных:

- альфа-система подкрепления;
- гамма-система подкрепления;
- метод ближайших соседей.

Обучение с подкреплением — для каждого прецедента имеется пара «ситуация, принятое решение»:

- генетический алгоритм.

Активное обучение — отличается тем, что обучаемый алгоритм имеет возможность самостоятельно назначать следующую исследуемую ситуацию, на которой станет известен верный ответ.

Обучение с частичным привлечением учителя (англ. semi-supervised learning) — для части прецедентов задается пара «ситуация, требуемое решение», а для части — только «ситуация».

Трансдуктивное обучение — обучение с частичным привлечением учителя, когда прогноз предполагается делать только для прецедентов из тестовой выборки.

Многозадачное обучение (англ. multi-task learning) — одновременное обучение группе взаимосвязанных задач, для каждой из которых задаются свои пары «ситуация, требуемое решение».

Многовариантное обучение (англ. multiple-instance learning) — обучение, когда прецеденты могут быть объединены в группы, в каждой из которых для всех прецедентов имеется «ситуация», но только для одного из них (причем, неизвестно какого) имеется пара «ситуация, требуемое решение».

Бустинг (англ. boosting — улучшение) — это процедура последовательного построения композиции алгоритмов машинного обучения, когда каждый следующий алгоритм стремится компенсировать недостатки композиции всех предыдущих алгоритмов.

ЗАДАНИЕ

1. «Жизненный цикл аналитики по методологии CRISP-DM»

Необходимо пошагово описать методологию CRISP-DM. Изобразите модель жизненного цикла исследования данных в методологии.

2. «Очистка и предобработка данных»

С использованием персонального компьютера требуется составить опорный конспект по предобработке и очистке данных.

КОНТРОЛЬНЫЕ ВОПРОСЫ

1. Что такое машинное обучение?
2. Перечислите способы машинного обучения.
3. Приведите примеры использования машинного обучения
4. Как можно использовать машинное обучение в искусственном интеллекте?
5. Искусственный интеллект против машинного обучения. Плюсы и минусы.