

УДК 004.8

ЗАДАЧИ, РЕШАЕМЫЕ С ПОМОЩЬЮ МАШИННОГО ОБУЧЕНИЯ

М.А. ШЕВЦОВ

(Представлено: канд. физ.-мат. наук, доц. Ю.Ф. ПАСТУХОВ)

Задачи, связанные с машинным обучением можно разделить на то какой вид прогнозирования или предсказания они будут совершать. В этой статье мы рассмотрим основные виды задач, решаемые с помощью машинного обучения.

1) Задача классификации (англ. **classification problem**) — имеется множество объектов (ситуаций), разделённых некоторым образом на классы. Задано конечное множество объектов, для которых известно, к каким классам они относятся. Это множество называется обучающей выборкой. Классовая принадлежность остальных объектов не известна. Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

В математической статистике задачи классификации называются также задачами дискриминантного анализа. В машинном обучении задача классификации решается, в частности, с помощью методов искусственных нейронных сетей при постановке эксперимента в виде обучения с учителем.

Типы классов:

- Двух классовая классификация. Наиболее простой в техническом отношении случай, который служит основой для решения более сложных задач.
- Много классовая классификация. Когда число классов достигает многих тысяч, задача классификации становится существенно более трудной.
- Непересекающиеся классы.
- Пересекающиеся классы. Объект может относиться одновременно к нескольким классам.
- Нечёткие классы. Требуется определять степень принадлежности объекта каждому из классов, обычно это действительное число от 0 до 1.

Основные задачи:

- Определение принадлежности объекта к определенному классу.

2) Кластерный анализ (англ. **cluster analysis**) — процедура, выполняющая сбор данных, содержащих информацию о выборке объектов, и затем упорядочивающая объекты в однородные группы. Задача кластеризации относится к статистической обработке, а также к широкому классу задач обучения без учителя.

Независимо от предмета изучения применение кластерного анализа предполагает следующие этапы:

- Отбор выборки для кластеризации. Подразумевается, что имеет смысл кластеризовать только количественные данные.
- Определение множества переменных, по которым будут оцениваться объекты в выборке, то есть признакового пространства.
- Вычисление значений той или иной меры сходства (или различия) между объектами.
- Применение метода кластерного анализа для создания групп сходных объектов.
- Проверка достоверности результатов кластерного решения.

Основные задачи:

- Разработка типологии или классификации.
- Исследование полезных концептуальных схем группирования объектов.
- Порождение гипотез на основе исследования данных.
- Проверка гипотез или исследования для определения, действительно ли типы (группы), выделенные тем или иным способом, присутствуют в имеющихся данных.

3) Регрессионный анализ (англ. **regression analysis**) — метод моделирования измеряемых данных и исследования их свойств. Данные состоят из пар значений зависимой переменной (переменной отклика) и независимой переменной (объясняющей переменной).

Основные задачи:

- Определение степени детерминированности вариации критериальной (зависимой) переменной предикторами (независимыми переменными).
- Предсказание значения зависимой переменной с помощью независимой(-ых).
- Определение вклада отдельных независимых переменных в вариацию зависимой.

Регрессионный анализ нельзя использовать для определения наличия связи между переменными, поскольку наличие такой связи и есть предпосылка для применения анализа.

Заключение. Каждый тип задач классифицируется по типу решаемых задач, сферы применения и алгоритмам, которые поддерживают решение данной проблемы, но важно помнить, что многие задачи разделены весьма условной линией и могут быть взаимозаменяемы в определённых условиях.

ЛИТЕРАТУРА

1. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. Множественная регрессия = Applied Regression Analysis. — 3-е изд. — М.: «Диалектика», 2007. — С. 912.
2. Шлезингер М., Главач В. Десять лекций по статистическому и структурному распознаванию. — Киев: Наукова думка, 2004.
3. Mitchell T. Machine Learning. — McGraw-Hill Science/Engineering/Math, 1997
4. Ryszard S. Michalski, Jaime G. Carbonell, Tom M. Mitchell (1983), Machine Learning: An Artificial Intelligence Approach, Tioga Publishing Company.