(cc) BY

Автор:

Дмитриев Егор Андреевич

студент

ФГАОУ ВО «Самарский национальный исследовательский

университет им. академика С.П. Королева»

г. Самара, Самарская область

ЛОГИЧЕСКИЕ КЛАССИФИКАТОРЫ

Аннотация*: в данной работе автором рассматриваются один из методов машинного обучения – использование метрических классификаторов.*

Ключевые слова*: машинное обучение, метрические классификаторы.*

Основные определения

Определение 1. Логическая закономерность – легко интерпретируемое правило, по которому достаточно легко выделить объекты обучающей выборки какого-то одного класса и не выделить объекты другого класса.

Определение 2. Предикат – функция, у которой значения лежат на множестве $\{0,1\}$.

Определение 3. Алгоритм классификации – отображение $a: X \to Y$, где X – метрическое пространство признаков классифицируемого объекта, а Y – множество классов. Классифицировать объект – однозначно определить, к какому классу относится объект.

Определение 4. Машинное обучение – нахождение отображения, в частности алгоритма классификации, который строится по множеству, называемому обучаемой выборкой, а качество обучения проверяется по множеству, называемому тестовой выборкой.

Введение

В последнее время к задачам машинного обучения, в частности к классификации объектов, привлекают все больше и больше внимания. В большинстве компаний, приходиться иметь дело с многочисленными данными, такие как: пер-

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

сональные данные сотрудников и клиентов, информация на крупных веб ресурсах, научные источники, наименования болезней, растений, животных и т. д. Для того, чтобы автоматизировать различные процессы, требуется использовать обучение системы для выполнения определенных задач.

Постановка проблемы

Необходимо ознакомиться с принципами логических классификаторов объектов.

Описание логических классификаторов

Суть задачи классификации заключается в построении алгоритма классификации по обучаемой выборке. Идея логических классификаторов заключается в поиске логических закономерностей, которые присущи объектам определенных классов.

Логической закономерностью является предикат, который обладает определенными свойствами:

1. *Интерпретируемость*, то есть предикат или логическое выражение представляет собой несложное выражение, которое включает в себя небольшое количество признаков и является понятным для людей, работающих в прикладной области.

2. *Информативность.*

Зафиксируем y принадлежащий множеству классу Y. Объекты данного класса будем называть положительными, а другого класса – отрицательными.

Рассмотрим 4 выражения:

1. $P_y = \sum_{i=1}^{m}[y_i = y]$ – число положительных объектов в выборке.

2. $N_y = \sum_{i=1}^{m}[y_i \neq y]$ – число отрицательных объектов в выборке.

3. $P_y(\phi) = \sum_{i=1}^{m}[\phi(x_i) = y][y_i = y]$ – число положительных объектов в выборке, который выделил предикат.

4. $N_y(\phi) = \sum_{i=1}^{m}[\phi(x_i) = y][y_i \neq y]$ – число отрицательных объектов, который выделил предикат.

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

Интуитивно понятно, что информативность предиката зависит от того, насколько хорошо он выделяет положительные объекты и не выделяет отрицательные, то есть $P_y(\text{ф}) \to max$ и $N_y(\text{ф}) \to min$. На практике свернуть эти предикаты – достаточно проблематичная задача. Однако существую довольно хорошие критерии, такие как:

– точный статистический тест Фишера:

$$IStat(\text{p, n}) = -log_2 C_P^p C_N^n / C_{P+N}^{p+n} \to max$$

– алгоритм бустинга:

$$\sqrt{p}\text{-}\sqrt{n} \to max$$

Бинарное решающее дерево

Бинарное решающее дерево – алгоритм классификации a(x), который состоит из композиции предикатов.

Каждое внутренний узел дерева – предикат, а каждый лист дерева – это класс объекта. Таким образом, начиная с корня дерева, мы постепенно продвигаемся к листьям. Путь по дереву зависит от значений предикатов. Дерево – это удобная модель. В общем случае, принадлежность к какому-либо классу можно заменить одним предикатом. На практике чаще всего, где у объектов логические признаки, предикат включает себя знаки сравнения ≥ и ≤. Для бинарных и номинальных признаков используется только равенство.

Заключение

В данной работе был рассмотрен принцип работы логических классификаторов.

Список литературы

1. Журавлев Ю.И. «Распознавание». Математические методы. Программная система. Практические применения / Ю.И. Журавлев, В.В. Рязанов, О.В. Сенько. – М.: Фазис, 2006.

2. Загоруйко Н.Г. Прикладные методы анализа данных и знаний. – Новосибирск: ИМ СО РАН, 1999.

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)