
Логическая подход в задаче восстановления регрессии

A Preprint

Листопадов Иван Сергеевич
ВМК МГУ
kramp87@mail.ru

Дюкова Елена Всеволодовна
ВМК МГУ
edjukova@mail.ru

2023

Abstract

Рассматривается одна из центральных задач машинного обучения — задача восстановления регрессии. Предлагается регрессионная модель с применением бустинга над элементарными классификаторами (эл.кл.). Ранее была реализована аналогичная композиция, которая базируется на генетических корректорах в качестве распознающих процедур. Предложенная модель строит голосование над представительными наборами эл.кл. с оптимизацией потерь для задачи восстановления регрессии.

Keywords Корректоры · Представительные эл.кл. · Регрессия

1 Введение

Задача восстановления регрессии является одной из основных задач обучения по прецедентам. Эта задача имеет следующую постановку.

Существует множество различных методов решения поставленной задачи. Наиболее распространенными алгоритмами являются линейная регрессия, метод ближайших соседей, градиентный бустинг, случайный лес. Одной из процедур решения сформулированной задачи может быть ее сведение к задаче классификации по прецедентам.

В качестве подхода к решению задачи классификации по прецедентам рассматривается дискретный или логический подход Л.В. and Ю.И. [1981]. Основное его достоинство заключается в возможности получения результата при отсутствии дополнительных предположений вероятностного характера и при небольшом числе прецедентов. Одним из направлений данного подхода является Correct Voting Procedures (CVP).

Основа работы Е.В. et al. [1996], Е.В. and Н.В. [2005] процедур CVP заключается в поиске среди обучающей информации корректных элементарных классификаторов (эл.кл.) – наборов из подмножеств признаков описаний, дающих возможность различать объекты из разных классов. Для этого используются методы построения покрытий матрицы и преобразования нормальных форм булевой функции.

Целью данной работы является разработка и реализация метода восстановления регрессии, основанного на процедурах корректного голосования (CVP). Алгоритм A1-Rg предполагает применение кластеризации, синтез классификаторов и восстановление значения целевой переменной. На каждом из этих этапов используются специфические алгоритмы, такие как DBSCAN и RUNC-M. Важной частью работы является экспериментальное сравнение предложенного метода с классическими алгоритмами восстановления регрессии на реальных данных.

2 Постановка задачи

Исследуется множество объектов M , каждый из которых описывается числовыми признаками $\{x_1, \dots, x_n\}$. Каждому объекту из M соответствует некоторое значение целевой переменной («ответа») y из числового множества Y , которое, быть может, неизвестно. Имеется набор объектов $X = \{X_1, \dots, X_m\}$ из множества M , каждому из которых соответствует определенное значение «ответа» $y_i, i = \{1, \dots, m\}$. Требуется по предъявленному набору значений признаков, описывающему некоторый объект из M , определить его значение целевой переменной ?.

Список литературы

- Баскакова Л.В. and Журавлёв Ю.И. Модель распознающих алгоритмов с представительными наборами и системами опорных множеств. Ж.вычисл.матем.и.матем.физ., 21(5):1264–1275, 1981.
- Дюкова Е.В., Журавлёв Ю.И., and Рудаков К.В. Об алгебраическом синтезе корректирующих процедур распознавания на базе элементарных алгоритмов. Ж. вычисл. матем. и матем. физ, 36(8):215–223, 1996.
- Дюкова Е.В. and Песков Н.В. Построение распознающих процедур на базе элементарных классификаторов. Математические вопросы кибернетики. Вып. 14 – М.: Физматлит, pages 57–92, 2005.