

Использование порождающих состязательных сетей для синтеза наборов данных в матричном представлении

Качальский И. В., Университет ИТМО, г. Санкт-Петербург
Забашта А. С., Университет ИТМО, г. Санкт-Петербург
Научный руководитель – Фильченков А. А., к.ф.-м.н., доц. ФИТиП Университета
ИТМО

Введение

В настоящее время продолжает набирать популярность задача мета-обучения, целью которой является автоматический выбор наилучшего алгоритма из множества для обработки набора данных, без непосредственного применения каждого из алгоритмов. Это позволяет сэкономить время необходимое на обработку набора данных алгоритмами и их настройку. В основе мета-обучения лежат мета-характеристики – численный вектор, описывающий свойства набора данных. Данные признаки позволяют понять зависимости между мета-признаками набора данных и качество работы алгоритмов на нём. Наборы данных задач классификации и кластеризации можно представить в виде матрицы произвольного размера. Поскольку в данной области имеется мало наборов данных, пространство мета-характеристик имеет точечный вид, из-за этого сложно найти зависимости работы алгоритмов от набора данных. Для решения этой проблемы требуется генерация набора данных, но если это делать случайным образом, то многие области пространства мета-характеристик будут всё равно не размеченными. Поэтому разработка алгоритма, способного генерировать наборы данных для заданных мета-характеристик является важной в области мета-обучения.

Цель работы

Целью данной работы является разработка алгоритма синтеза генерации наборов данных по заданным мета-характеристикам с применением условных генеративно-состязательных сетей, что позволит не только заполнить пробелы в пространстве мета-характеристик, но и обучить модель, решающую задачу мета-обучения.

Базовые положения исследования

В ходе данной работы предполагается разработать единый способ представления набора данных в матричном виде, который не будет терять информацию. Так же требуется разработать алгоритм свёртки и обратной свёртки, которые будут учитывать особенности наборов данных в матричном виде. Наконец разработать алгоритм обучения генеративно-состязательной сети, который будет учитывать оценки качества работы алгоритмов, на синтезированных наборах данных и будет стараться заполнить пространство мета-характеристик.

Предварительные результаты

На данный момент разработан алгоритм перевода набора данных в матричный вид. Предложен набор мета-характеристик и набор оценивающих алгоритмов, для обучения. Разработан и обучен прототип генеративно-состязательной сети, который протестирован на наборах данных, сформированных по полиномам. Результаты прототипа показали, что нужно продолжать исследования в данной области и продолжать разработку алгоритма обучения данной сети.

Список литературы

1. Забашта А.С., Фильченков А.А. Построения наборов данных для задачи классификации по их характеристическому описанию // Научно-технический вестник информационных технологий, механики и оптики -2017. - Т. 17. - № 3(109). - С. 498–505.