

АЛГОРИТМЫ НАСТРОЙКИ ГИПЕРПАРАМЕТРОВ НА ОСНОВЕ ОБЪЕДИНЕНИЯ АПРИОРНЫХ И АПОСТЕРИОРНЫХ ЗНАНИЙ О ЗАДАЧЕ

Смирнова В.С. (Национальный исследовательский Университет ИТМО)

Научный руководитель – доцент, к.т.н. Фильченков А.А.

(Национальный исследовательский Университет ИТМО)

В данной научной работе исследованы известные и сформулирован оптимальный алгоритм настройки гиперпараметров на основе объединения априорных и апостериорных знаний о конкретной задаче.

Введение. Одни и те же виды моделей машинного обучения могут требовать различные предположения, веса или скорости обучения для различных видов данных. Эти параметры называются гиперпараметрами и недавний интерес к сложным и вычислительно дорогостоящим моделям машинного обучения со многими гиперпараметрами, таким как системы автоматического машинного обучения и глубокие нейронные сети, привел к возобновлению исследований по оптимизации гиперпараметров. Как до, так и после обучения, по конкретной задаче могут быть известны некоторые факты, которые можно использовать для настройки гиперпараметров, что может дать существенное преимущество перед случайной расстановкой тех же параметров. В зарубежных исследованиях используются отдельно априорные (известные до обучения) и апостериорные (известные после обучения) знания о задаче. Исходя из этих знаний определяется функция сбора данных, которая использует прогнозирующее распределение вероятностной модели и определяет то, насколько полезны различные точки-кандидаты, что позволяет оптимально настраивать гиперпараметры. По сравнению с оценкой более сложной функции черного ящика, функция сбора данных является не такой трудоёмкой для вычисления и поэтому может быть полностью оптимизирована. Байесовская оптимизация достигла впечатляющих результатов, однако она использует исключительно априорные знания.

Основная часть. Суть предлагаемого решения: объединить априорные и апостериорные сведения о задаче, чтобы добиться лучших результатов в обучении. Используя методы байесовской оптимизации и оптимизации на основе модели неизвестной функции (чёрный ящик), а также средства мета-обучения, задействовать ранее извлечённые из обучения на аналогичных задачах знания для настройки гиперпараметров в конкретной задаче. Проанализировать наиболее влияющие факторы и представить общий алгоритм.

Выводы. Поскольку высокая вычислительная потребность многих современных задач машинного обучения делает чистую оптимизацию черного ящика чрезвычайно дорогостоящей, мы сосредоточились на современных многоцелевых методах, которые используют (намного) более дешёвые варианты функции черного ящика для приблизительной оценки качества настроек гиперпараметров. Более того, объединение априорных и апостериорных знаний позволило сделать обучение эффективнее, нежели обучение на основе исключительно априорных знаний. Разработанный подход может быть применим к большому ряду актуальных на сегодняшний день задач, как AutoML, глубокие нейронные сети, задачи классификации, кластеризации и так далее.