Использование нелинейных смесей экспертов в задачах классификации

Трофимов Михаил Игоревич

Студент

ФУПМ МФТИ, Москва, Россия

E-mail: mikhail.trofimov@phystech.edu

Построение смеси экспертов [1, 2] — важная и актуальная задача, которая возникает в приложениях, где точность прогноза важнее интерпретируемости модели, например, в системах автоматической детекции запрещенного контента. Другой важной сферой применения являются задачи, в которых описание объектов задано разнородными признаками, например — вещественными и категориальными с большим числом категорий.

В данной работе рассматривается метод построения смеси экспертов, новизна заключается в использовании нелинейных отображений. Исследование можно условно разделить на 3 части: обучение базовых экспертов, выбор алгоритма для обучения смеси и тестирование метода на реальных данных.

Базовые эксперты могут являться представителями как разных семейств параметрических моделей, так и представителями одного семейства. Более того, сам исходный объект можно представить различными признаковыми описаниями, и каждое из них потребует обучение своего эксперта. Заметим, что множество экспертов естественным образом возникает в случаем, когда исходный объект представлен большим количеством как вещественных, так и категориальных признаков с большим количеством категорий — категориальные признаки представляются в виде разреженной бинарной матрицы и обучается линейная модель, в том время как на вещественных признаках может быть обучен более сложный эксперт.

Для предотвращения эффекта переобучения, исходная совокупность объектов разбивается на три множества: для обучения базовых алгоритмов, для обучения композиции и для валидации. Стоит отметить, что соотношение размеров множеств для обучения экспертов и смеси нуждается в подборе. Смесь обучается на ответах базовых алгоритма и некоторых исходных признаках. В данной работе в качестве модели смеси используются нелинейные отображения. Выбор алгоритма зависит от конкретного целевого функционала, автор использовал случайный лес [3] и градиентный бустинг над решающими деревьями [4]. Стоит отметить, что очень важным аспектом является

борьба с переобучением — в упомянутых алгоритмах использовалось большое (1000–5000) число деревьев малой (3–5) глубины.

Итогом проделанной работы является алгоритм, который превосходит каждый отдельный компонент смеси и их линейную комбинацию ценою увеличения вычислительной сложности. Эффективность данного метода была подтверждена на реальных данных конкурсов The Hunt for Prohibited Content и Tradeshift Text Classification, проводимых международной платформой Kaggle.com.

Литература

1. Журавлёв Ю. И. «Об алгебраическом подходе к решению задач распознавания или классификации», Проблемы кибернетики: Вып.33. — 1978. — С. 5–68.
2. Jacobs, Jordan, Nowlan, Hinton. «Adaptive mixtures of local experts», Neural Computation, 3, 79-87
3. L. Breiman, «Random forests», Machine learning 45 (1), 5-32
4. J. Friedman, «Greedy Function Approximation: A Gradient Boosting Machine», The Annals of Statistics, Vol. 29, No. 5, 2001.