

Устойчивый метод усиления слабых классификаторов

Вежневек Александр Петрович

студент

Московский государственный университет им. М.В. Ломоносова, Москва, Россия

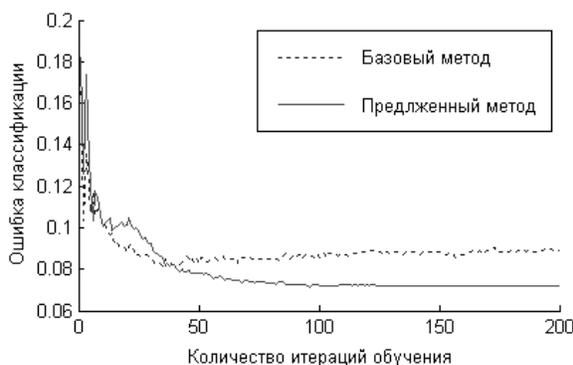
E-mail: vezhnick@gmail.com

Усиление слабых классификаторов это подход к решению задачи классификации, путём комбинирования примитивных классификаторов в один, более сильный комитет. Основная идея метода заключается в итеративной минимизации выпуклого функционала ошибки классификации, путем добавления в комитет очередного слабого классификатора.

Простота реализации, высокая скорость работы финального классификатора, быстрая сходимость, высокая обобщающая способность и универсальность придают методу большое прикладное значение. Тем не менее, последние исследования показали склонность метода к переобучению, особенно в случае наличия шума или выбросов в обучающей выборке. Подходы решающие данную проблему либо требуют априорных знаний о природе шума, которые в большинстве случаев недоступны, либо сильно меняют алгоритм, лишая его простоты реализации и уменьшая скорость работы результирующего комитета.

Предлагается новый алгоритм усиления простых классификаторов, устойчивый к шуму и выбросам. Предложенный метод не меняет общей структуры базового алгоритма и не требует априорных знаний о шуме. Предлагается лишь изменить формулу вычисления веса очередного слабого классификатора в комитете, таким образом, чтобы отступ прецедентов от разделяющей плоскости был более равномерным. Таким образом, сохраняется простота реализации и скорость работы.

Эффективность метода проверялась экспериментально на реальных данных из различных областей, в основном медицины, взятых из базы UCI Machine Learning Repository с помощью скользящего контроля с числом вложенности 5. Для оценки устойчивости алгоритма в обучающую выборку добавлялся искусственный шум.



Левый график показывает работу базового и предложенного метода на не зашумленных данных. На правом графике 20% прецедентов из обучающей выборки были зашумлены (инвертирован маркер принадлежности к классу).

Литература

- [1] Jerome Friedman, Trevor Hastie, and Robert Tibshirani. "Additive logistic regression: A statistical view of boosting." *The Annals of Statistics*, 38(2):337–374, April 2000.
- [2] Rosset, Zhu and Hastie. "Boosting as a Regularized Path to a Maximum Margin Classifier" *Journal of Machine Learning Research* 5, pp. 941–973, 2004.
- [3] Philip M. Long, Rocco A. Servedio. "Martingale Boosting." *Eighteenth Annual Conference on Computational Learning Theory (COLT)*, 2005