AdaBoost для задач многоклассовой классификации

Впервые AdaBoost (от Adaptive Boosting) был предложен в http://www.face-rec.org/algorithms/Boosting-Ensemble/decision-theoretic_generalization.pdf (1997) и стал настоящим прорывом в области построения точных алгоритмов на базе «слабых» моделей. Основная идея AdaBoost заключается в перевзвешивании выборки на каждой итерации таким образом, чтобы наиболее сложные объекты получали больший вес. Это позволяет обучить новый алгоритм, который сможет исправить ошибки предыдущих, уделяя усиленное внимание объектам с большим весом. Как и для всех видов бустинга, решающее правило представляет собой взвешенную сумму.

В изначальном виде AdaBoost был способен решать задачи классификации только на два класса. Обобщение алгоритма AdaBoost на многоклассовый случай было предложено в http://ww.web.stanford.edu/~hastie/Papers/SII-2-3-A8-Zhu.pdf. В данной работе авторы предлагают два алгоритма SAMME и SAMME.R, которые представляют собой многоклассовую модификацию AdaBoost для дискретного (когда базовые алгоритмы возвращают только метки классов) и непрерывного (алгоритмы вычисляют вероятности) случаев.

Алгоритм SAMME (Stagewise Additive Modeling)

Дано

1. K - количество классов, D - количество признаков
2. Обучающая выборка $S = \{(x_i, y_i)\}_{i=1}^N = (X, y)$, где $x_i \in R^D$ – вектор признаков, а y_i принимает значения из конечного множества $\{1,2,...,K\}$

Выборку S удобно записать в виде матрицы $X \in R^{N\times D}$, строками которой являются $x_i^\top$, и вектора $y = (y_1, \dots, y_N)^\top$: $S = (X, y)$

3. Количество базовых моделей: L

Найти

На основе данной выборки построить решающее правило $\hat{y} : R^D \to \{1,2,...,K\}$, которое для произвольного x предсказывает $\hat{y}(x)$ – принадлежность x к одному из K классов.

Алгоритм

1. Проинициализируем веса для каждого объекта $w_i^1 = \dfrac{1}{N}, i = 1,...,N$
2. Для всех $l = 1,...,L$:

    1. Обучим базовый алгоритм $\hat{y}_l$ при помощи весов w_n^l
    2. Вычислим взвешенную ошибку ε_l алгоритма $\hat{y}_l$:

$$\varepsilon_l = \sum_{i=1}^{n} w_i^l [\hat{y}_l(X_i) \neq y_i]$$

3. Вычислим вес нового классификатора:

$$(1)\alpha_l = ln\frac{1-\varepsilon_l}{\varepsilon_l} + ln(K-1)$$

4. Пересчитаем и нормализуем веса:

$$\bar{w}_i^{l+1} = w_i^l \cdot exp(\alpha_l \cdot [\hat{y}_l(X_i) \neq y_i]), i = 1,..,N,$$

$$w_i^{l+1} = \frac{\bar{w}_i^{l+1}}{\sum_{l=1}^{L} \bar{w}_i^{l+1}}, i = 1,..,N$$

3. Финальный прогон классификатора: $\hat{y}(X) = arg\max_k \sum_{l=1}^{L} \alpha_l [\hat{y}_l(X) = y]$

Теоретическое обоснование

Применительно к задачам классификации 2х классов, AdaBoost оказался чрезвычайно успешным. Брейман (1996) назвал AdaBoost «best off-the-shelf classifier in the world». Однако это не относится к многоклассовым проблемам, хотя AdaBoost также предполагалось использовать и для многоклассовой классификации. Заметим, что для повышения неправильно классифицированных обучающих данных требуется, чтобы ошибка каждого слабого классификатора была меньше $1/2$ (относительно распределения, в котором они были обучены), в противном случае вес классификатора будет отрицательным, и веса обучающих выборок будут обновлены в неправильном направлении. Для задач классификации с двумя классами это требование примерно такое же, как и при случайном угадывании, но когда $K > 2$, точность $1/2$ становится намного сложнее достижима, чем показатель точности случайного угадывания $1/K$. Следовательно, AdaBoost может потерпеть неудачу.

Алгоритм SAMME имеет одну и ту же структуру что и AdaBoost, с отличием в вычислении веса классификатора (1). Теперь для того, чтобы α_l была положительной, необходимо только $(1 - \varepsilon_l) > 1/K$, то есть чтобы точность каждого слабого классификатора была лучше, чем раннее угадывание, а не 1/2. Samme придает вес ошибочно классифицированным точкам больший, чем AdaBoost.

Обобщение функции потерь на многоклассовый случай следует естественным образом:

$$Loss(y, f) = exp(-\frac{1}{K}(y_1 f_1 + \ldots + y_K f_K)) = exp(-\frac{1}{K} y^T f)$$

где $f_1 + \ldots + f_K = 0$ (симметричное ограничение);

Очевидно, что при $K = 2$ SAMME сводится к AdaBoost.

Кулакова Виолетта, 895а