

Калибровка метода многоклассовой классификации один-против-всех для бустинга

Вежневцев А. П., Соболев А. А., Вежневцев В. П.

avezhnevets@graphics.cs.msu.ru, neusobol@yandex.ru,
dmoroz@graphics.cs.msu.ru

Москва, МГУ им. Ломоносова, лаборатория машинной графики и
мультимедиа

В данной статье рассматривается задача многоклассовой классификации; показывается, что классический метод один против всех может быть существенно улучшен с помощью метода шкалирования Платта выходов бинарных классификаторов [1, 3].

Введение

Для сведения задачи классификации со многими классами к бинарной существует множество способов. Самый простой из них — один-против-всех [2]. Для бинарных классификаторов, основанных на методе опорных векторов, было показано, что [2] такой простой метод, при регуляризации бинарных классификаторов, не уступает многим более сложным и вычислительно трудным методам, основанных на самокорректирующихся кодах [4]. В данной статье показывается, что метод один-против-всех также очень эффективен и в случае использования в качестве бинарных классификаторов комитетов деревьев решений, построенных бустингом и откалиброванных алгоритмом Платта.

Описание подхода

Основная идея заключается в использовании независимого шкалирования выходов бинарных классификаторов методом Платта [1] для их лучшей согласованности, что повышает качество работы алгоритма. Пусть $f_c(x): X \rightarrow R$ — бинарный классификатор, настроенный на распознавание класса $c \in Y = [1, \dots, C]$. Предполагается что бинарный классификатор возвращает *уверенность* в том, что прецедент принадлежит классу c (например, возвращаемое значение может быть отступом от разделяющей поверхности). Тогда многоклассовый классификатор по классическому методу один-против-всех строится как

$$F(x) = \arg \max_{c \in Y} f_c(x).$$

В классическом методе не предполагается никакой калибровки выходов бинарных классификаторов. Оценим апостериорные вероятности следующим образом:

$$P(c|x) \approx \tilde{P}(c|x) = \frac{1}{1 + \exp(Af_c(x) + B)},$$

где параметры A и B оцениваются алгоритмом Платта. Будем строить финальный классификатор следующим образом:

$$F(x) = \arg \max_{c \in Y} \tilde{P}(c|x).$$

В результате получатся более согласованные бинарные классификаторы, что ведет к уменьшению ошибки.

Эксперименты

Для сравнения были взяты три метода: один-против-всех, самокорректирующиеся коды (ЕСС) и один-против-всех со шкалированием Платта. В качестве бинарных классификаторов использовались деревья классификации глубины 3, усиленные бустингом. Ниже представлены графики зависимостей ошибки на контрольных данных (использовался скользящий контроль) от итераций алгоритмов (Рис. 1): для ЕСС — каждая последующая точка соответствует ошибке на длине кодовых слов, большей на единицу, а для один против всех — количество итераций бинарных классификаторов. Для самокорректирующихся кодов в качестве бинарных классификаторов использовались комитеты из 40 деревьев, построенных бустингом (количество комитетов определяется длиной кодового слова).

Заключение

Эксперименты показывают, что применение шкалирования Платта к алгоритму один-против-всех дает существенное уменьшение количества ошибок этого метода — на трех из пяти представленных выборках метод один против всех, откалиброванный методом Платта, работает качественнее ЕСС.

Литература

- [1] *J. Platt* Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. // *Advances in Large Margin Classifiers*, 1999. — pp. 61–74.
- [2] *R. Rifkin, A. Klautau*. In *Defense of One-Vs-All Classification*. // *The Journal of Machine Learning Research*, 2004. — pp. 101–141.
- [3] *A. Niculescu-Mizil and R. Caruana* Predicting good probabilities with supervised learning. // *Proceedings of the 22nd international conference on Machine learning*, 2005. — pp. 625–632
- [4] *Chun-Nan Hsu and Yu-Shi Lin* Boosting Multiclass Learning with Repeating Codes // *Journal of Artificial Intelligence Research*, 2006. — pp. 263–286.

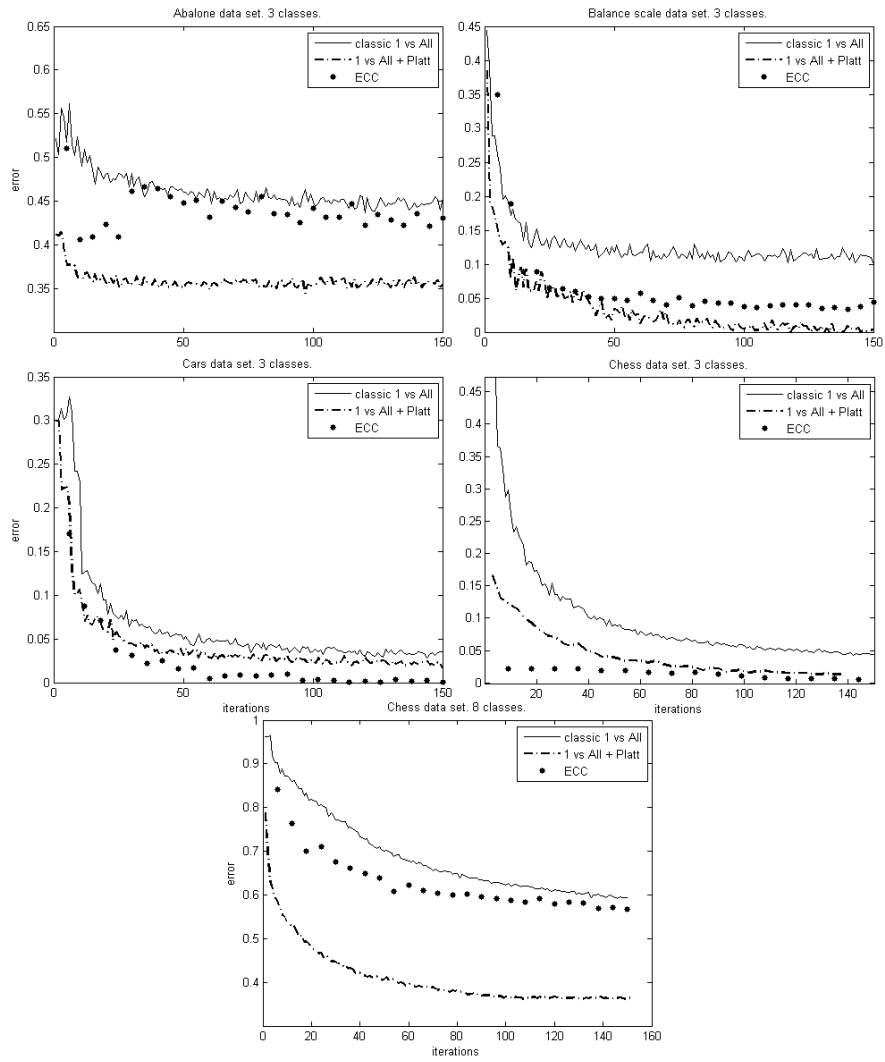


Рис. 1. Результаты экспериментов.