Вероятностный выход для методов многоклассовой классификации на основе самокорректирующихся кодов

Соболев А. А., Вежневец А. П., Вежневец В. П.
Москва, МГУ им. М. В. Ломоносова,
Лаборатория машинной графики и мультимедиа
neusobol@yandex.ru, {avezhnevets,dmoroz}@graphics.cs.msu.ru

Одним из способов сведения задачи классификации с множеством классов к задаче бинарной классификации (с двумя классами) является семейство методов основанных на самокорректирующихся кодах [?]. В этой статье рассматривается получение вероятностного выхода для данного семейства методов.

Введение

Пусть дана тренировочная выборка $D = \{(x_n, y_n)\}_{n=1}^{N} \in X \times Y$, где X — пространство образов, а $Y = \{c_1, \dots, c_n\}$ — множество меток классов. Пусть $M \in \{\pm 1\}^{C \times T}$ — кодовая матрица, где T — длина кодового слова. Финальный классификатор представляет собой комитет $f(x) = [f_1(x), \dots, f_T(x)]^T$, где $f_t(x) : x \to R$. В итоге настроенный классификатор работает по принципу минимального расстояния, то есть классом нового объекта x считается тот класс y^*, расстояние до кодового слова $M(y^*)$ которого минимально

$$y^* = \arg\min_y (\triangle(M(y), f(x))). \tag{1}$$

Классический подход к получению вероятностей

Такой подход дает существенный прирост в качестве классификации. Однако, для многих прикладных задач, например задач машинного зрения, требуется получение не просто наиболее вероятного класса для прецедента, но и вероятности принадлежности прецедента к тому или иному классу. Ранее для решения данной задачи предлагалось использовать подход, основанный на сведении задачи к решению системы линейных уравнений [?]. Пусть для каждого классификатора из комитета $f_t(x)$ можно вычислить вероятностный выход $P(f_t(x))$. Пусть $p = \langle P(f_1(x)), \dots, P(f_T(x)) \rangle$. Например, если столбец t кодовой матрицы M равен $\langle 1, 0, 1 \rangle$, то $P(f_t(x)) = P(c_1|x) + P(c_3|x)$. Обозначим $z = \langle P(c_1|x), \dots, P(c_k|x) \rangle$ — вектор искомых вероятностей. Тогда можно записать матричное уравнение $M^T z = p$. Решая эту систему, мы получим искомые вероятности. Система, вообще говоря, может быть не совместной, поэтому предлагается использовать метод наименьших квадратов. Данный метод имеет ряд недостатков:

— для каждого классифицируемого прецедента приходится решать систему заново, из-за чего метод становится вычислительно сложным;
— метод вычислительно неустойчив — зависит от обусловленности матрицы M.

Предлагаемый метод

Предлагается использовать подход, основанный на нормировке отступа, аналогично [?, ?]. Для этого требуется определить отступ для каждого конкретного класса и выбрать метод шкалирования значения отступа для аппроксимации условной вероятности класса. Отступ для класса c_i определим как

$$\rho(c_i, f(x)) = \min_{c \in Y, c \neq c_i} \triangle M(c, f(x)) - \triangle M(c_i, f(x)). \qquad (2)$$

Для того чтобы отступ максимально точно аппроксимировал апостериорную вероятность, применим алгоритм шкалирования, предложенный в [?] [?], в котором предлагается преобразовать отступ сигмоидальной функцией с параметрами, минимизирующими невязку предсказанной вероятности и реальной. Заметим, что при вычислении ответа на образ x вычисление отступов и вероятностей почти не требует дополнительных вычислений — расстояния до кодовых слов будут рассчитаны в любом случае во время классификации:

$$P(c_i|x) \approx \frac{1}{1 + \exp(A\rho(c_i, f(x)) + B)}. \qquad (3)$$

Оценка параметров сигмиоды A и B производится на отдельной верификационной выборке, не являющейся ни частью обучающей, ни частью контрольной. В качестве альтернативы можно использовать скользящий контроль с глубиной 3, как предложено в [?].

Эксперименты

Мы сравнили работу своего метода и классического на нескольких выборках из репозитория задач UCI [?]. Мы использовали скользящий контроль глубины 3 для настройки параметров сигмоиды (для каждого класса отдельно). В качестве бинарных классификаторов мы использовали комитет деревьев глубины один (stumps), построенный методом AdaBoost.

Ниже приводятся диаграммы калибровки [?] для различных классов из набора abalone для классического и предложенного метода (из-за ограниченного места мы, к сожалению, не можем привести больше графиков). Диаграммы строятся следующим образом. Весь диапазон

Рис. 1: Результаты экспериментов. Три верхних графика получены методом [**?**], нижние с помощью предложенного метода.

предсказанных вероятностей делится на ячейки. Для каждой ячейки считается реальная доля прецедентов исследуемого класса. Чем ближе точки лежат к диагональной прямой, тем лучше откалиброван метод.

Как видно из графиков, предложенный метод дает более адекватные результаты. Для второго класса классический метод всегда давал одно и то же решение. Еще раз отметим, что предложенный метод вычислительно намного проще.

Литература

[1] *T. Dietterich and G. Bakiri* Solving Multiclass Learning Problems via Error-Correcting Output Codes. // Journal of Artificial Intelligence Research. — 1995. — Pp. 263–286.

[2] *E. Kong and T. Diettrich* Probability estimation via error-correcting output coding. // In Int. Conf. of Articial Inteligence and soft computing. — 1997.

[3] *J. Platt* Probabilistic outputs for support vector machines and comparison to regularized likelihood methods. // Advances in Large Margin Classifiers. — 1999. — Pp. 61–74.

[4] *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.

[5] *D.J. Newman, S. Hettich, C.L. Blake and C.J. Merz* UCI Repository of machine learning databases. // University of California, Irvine, Dept. of Information and Computer Sciences.