

Finding efficient mappings of observations when reducing multiclass to binary learning

Massih-Reza Amini, Yury Maximov

25 февраля 2014 г.

Название. Finding efficient mappings of observations when reducing multiclass to binary learning

Источник задачи. Совместная постановка МФТИ, Universite Joseph Fourier и Inria Rhone Alpes.

Задача. Исследовать различные подходы к решению задач классификации с многими классами и сравнить их эффективность.

Подходы к решению. Пусть задана выборка $\{(x_i, y_i)\}$; классы y_i могут принимать значения из множества $\{1, 2, 3, \dots, k\}$. k предполагается достаточно большим.

- **One vs all approach.** В этом подходе строятся k бинарных классификаторов, по следующему принципу

$$f_i(x) = \begin{cases} 1, & y(x) = i \\ -1, & y(x) \neq i \end{cases}$$

Настройка алгоритма происходит по следующему принципу

$$\operatorname{argmin}_f \frac{1}{m} \sum_{t=1}^m \mathbf{1}[\max_j(x^t, j) > f(x^t, y(x^t))]$$

Решение принимается следующим образом:

$$\operatorname{argmax}_i f_i(x).$$

- **One vs one approach.** В этом случае строятся $\binom{k}{2}$ классификаторов, которые разделяют элементы пар классов, а именно

$$h_{ij}(x) = \begin{cases} 1, & y(x) = i \\ -1, & y(x) = j \end{cases}$$

Обозначим

$$h(i, x) = \sum_{j:j \neq i} h_{ij}(x).$$

Настройка алгоритма происходит по следующему принципу

$$\operatorname{argmin}_f \frac{1}{m} \sum_{t=1}^m \mathbf{1}[\max_j h(j, x^t) > h(y(x), x^t)]$$

Решение принимается следующим образом:

$$\operatorname{argmax}_i \sum_j h_{ij}(x)$$

- **All pairs approach.** Этот подход близок к первому подходу в том смысле, что производится обучение k бинарных классификаторов $f(\cdot, i)$, $1 \leq i \leq k$, но отличается функционалом ошибки:

$$\operatorname{argmin}_f \frac{1}{m} \sum_{t=1}^m \frac{1}{k-1} \sum_{j \neq y(x^t)} \mathbf{1}[f(x^t, j) > f(x^t, y(x^t))]$$

где первая сумма берется по числу объектов.

В этом случае мы избираем обучения на малых выборках, что благоприятно сказывается на качестве алгоритма.

Решение принимается стандартным образом:

$$\operatorname{argmax}_i f_i(x)$$

Последний подход оказывается достаточно эффективным (в смысле качества классификации) для задач с большим числом классов.

- **Error-Correcting Output Codes** еще один подход, с которым предстоит ознакомиться слушателю. Для первого представления и сведений о реализации мы рекомендуем ознакомиться с
 - [Dietterich, Bakiri. Solving Multiclass Learning Problems via Error-Correcting Output Codes. 1995.](#)
 - [Sergio Escalera, Oriol Pujol, Petia Radeva. Error-Correcting Output Codes Library](#)

Научная новизна и значимость. Подход 3 дает существенный выигрыш в качестве (подтверждено публикациями группы в ведущих международных конференциях: NIPS, ECCV, SIGIR, ECIR и др.), однако обладает не самой высокой скоростью, что является несомненным недостатком. Аналогичная ситуация для подхода 4, особенно в случае данных высокой размерности. Таким образом, повышение производительности классификаторов в рамках данного направления представляется актуальной задачей. В то же время (математические) методы снижения размерности для подходов 3 и 4 (успешно) не применялись.

Предполагаемый подход к решению. базируется на использовании Леммы Джонсона-Линдерштраусса и аналогичных утверждений, обеспечивающих квази-изометрические вложения точек (в смысле евклидова расстояния). Предполагается, что студент освоит эту, в общем-то несложную технику, и поставит необходимые эксперименты.

Данные. Данные с различным числом классов. Студент волен выбирать наиболее удобную для него коллекцию.

Текстовые данные

- Текстовые данные коллекции [Reuters](#)
- Данные [нашего конкурса](#) Kaggle от LIG

Данные изображений:

- [MNIST collection](#)

Для начальных экспериментов, подойдут препарированные данные [CSIE](#). Рекомендуется использовать данные sector или aloi.

Литература. Общее введение в тематику:

- [Rifkin's lecture on multiclass classification.](#)
- [Xia's lecture on the same topic](#)

Статьи по теме:

- Tax, Duin. Using two-class classifiers for multiclass classification. Pattern Recognition, 2002. Proceedings. 16th International Conference on (Volume:2).
- Dietterich, Bakiri. Solving Multiclass Learning Problems via Error-Correcting Output Codes. 1995.
- Allwein, Schapire, Singer. Reducing Multiclass to Binary: A Unifying Approach for Margin Classifiers. Journal of Machine Learning Research 1 (2000) 113-141.
- Nicolas Usunier, Massih-Reza Amini, Patrick Gallinari: Generalization error bounds for classifiers trained with interdependent data. NIPS 2005

Базовые алгоритмы. Начинать нужно с наиболее простого: линейные классификаторы. Вторым шагом следует попробовать применить SVM с различными ядрами, 3-им шагом Adaboost.

Контакты. yury.maximov@phystech.edu Написание отчетов предпочтительно(но не обязательно) на английском. Скайп: yuramaximov