Об эффективности эмпирических функционалов качества решающей функции

Недельмо В. М.
nedelko@math.nsc.ru
Новосибирск, Институт математики СО РАН

В работе предложен способ сравнения эффективности различных функционалов, оценивающих риск решающей функции по той же выборке, по которой построена функция, что позволяет находить в некотором смысле оптимальные функционалы.

Задача построения решающей функции

Пусть X — пространство значений переменных, используемых для прогноза, а Y — пространство значений прогнозируемых переменных, и пусть C — множество всех вероятностных мер на заданной σ-алгебре подмножеств множества $D = X \times Y$. При каждом $c \in C$ имеем вероятностное пространство: $\langle D, B, P_c \rangle$, где B — σ-алгебра, $P_c[D]$ — вероятностная мера. Параметр будем называть *стратегией природы.*

Решающей функцией называется соответствие $f\colon X \to Y$.

Качество принятого решения оценивается заданной функцией потерь $L : Y^2 \to [0, \infty)$. Под риском будем понимать средние потери:

$$R(c, f) = \int_D L(y, f(x))\, dP_c[D].$$

Пусть $\nu = \big\{(x^i, y^i) \in D \bigm| i = 1, \ldots, N\big\}$ — случайная независимая выборка из распределения $P_c[D]$. Эмпирический риск определим как средние потери на выборке:

$$\widetilde{R}(\nu, f) = \frac{1}{N} \sum_{i=1}^{N} L(y^i, f(x^i)).$$

Заметим, что значение риска зависит от стратегии природы — распределения, которое неизвестно. Функционал скользящего экзамена определяется как

$$\breve{R}(\nu, Q) = \frac{1}{N} \sum_{i=1}^{N} L(y^i, f_{Q,\nu'_i}(x^i)),$$

где $\nu'_i = \nu \backslash \{(x^i, y^i)\}$ — выборка, получаемая из ν удалением i-го наблюдения, $Q\colon \{\nu\} \to \Phi$ — алгоритм построения решающих функций, $f_{Q,\nu}$ — функция, построенная по выборке ν алгоритмом Q, Φ — заданный класс решающих функций.

Доверительный интервал для риска

Доверительный интервал для R будем задавать в виде $[0, \hat{R}(\nu)]$. Здесь мы ограничиваемся односторонними оценками, поскольку на практике для риска важны именно оценки сверху. Таким образом, в данном случае построение доверительного интервала эквивалентно выбору функции $\hat{R}(\nu)$, которую будем называть оценочной функцией или просто оценкой (риска).

При этом должно выполняться условие:

$$\forall c \quad P_c\bigl(R \leqslant \hat{R}(\nu)\bigr) \leqslant \eta,$$

где η — заданная доверительная вероятность.

При построении оценок риска первая проблема, которую нужно решить, это сравнение качества различных оценок.

Можно положить, что задан функционал качества $K\bigl(F_{c,\hat{R}}(\cdot)\bigr)$, где $F_{c,\hat{R}}(\cdot)$ — функция распределения оценки $\hat{R}(\nu)$. Выбор данного функционала, так же как и выбор функции потерь, определяется практическими соображениями. Простейшим вариантом такого функционала является математическое ожидание.

При фиксированной стратегия природы c функционал K позволяет сравнивать качество оценок риска и находить оптимальную оценку.

Однако на практике распределение c неизвестно, а оценки, оптимальной при всех распределениях, может не существовать. В этом случае естественным является поиск множества Парето недоминируемых оценок.

Эмпирические функционалы качества

Известные на данный момент оценки риска (напр. [1]) строятся не как функция непосредственно выборки, а через композицию $\hat{R}(\nu) = \hat{R}_e\bigl(\dot{R}(\nu)\bigr)$, как функция значений некоторого эмпирического функционала $\dot{R}$, в качестве которого обычно выступает эмпирический риск или скользящий экзамен.

Эмпирический функционал здесь выступает в роли точечной оценки риска, на основе которой строится интервальная оценка.

В докладе представлены исследования эффективности функционалов эмпирического риска и скользящего экзамена для частных задач. При этом под эффективностью понимается, насколько хорошая интервальная оценка риска может быть построена на основе данного функционала.

Определение 1. *Оценочную функцию $\hat{R}(\nu)$ назовем согласованной с эмпирическим функционалом $\dot{R}$, если для выборок ν_1 и ν_2 одинакового объема*

$$\dot{R}(\nu_1) > \dot{R}(\nu_1) \Rightarrow \hat{R}(\nu_1) \geqslant \hat{R}(\nu_2).$$

Достаточно естественным представляется ограничиться рассмотрением только таких оценочных функций, которые согласованы с функционалами эмпирического риска и скользящего экзамена. Это означает, что оценка вероятности ошибки не должна убывать при увеличении значения эмпирического функционала.

Данное условие позволяет резко сузить пространство поиска при нахождении Парето-оптимальных оценочных функций.

Работа выполнена при поддержке РФФИ, проект № 07-01-00331-а, и СО РАН, проект № 7.

Литература

[1] *Вапник В. Н., Червоненкис А. Я.* Теория распознавания образов. — Москва: Наука, 1974. — 415 с.

[2] *Лбов Г. С., Старцева Н. Г.* Логические решающие функции и вопросы статистической устойчивости решений. — Новосибирск: Изд-во ИМ СО РАН, 1999. — 211 с.

[3] *Nedelko V. M.* Estimating a Quality of Decision Function by Empirical Risk // LNAI 2734. Machine Learning and Data Mining in Pattern Recognition. Third International Conference, MLDM 2003, Leipzig. Proceedings. Berlin: Springer-Verlag, 2003. — P. 182–187.

[4] *Недельько В. М.* Оценивание смещения эмпирического риска для линейных классификаторов // Таврический вестник информатики и математики. Изд-во НАН Украины. — 2004. — № 1. — С. 47–53.

[5] *Недельько В. М.* Оценка смещения эмпирической оценки риска решающей функции // Докл. всеросс. конф. «Математические методы распознавания образов», ММРО-11 — Москва: ВЦ РАН, 2003. — С. 148–150.