

## МЕТА-ОБУЧЕНИЕ КРИТЕРИЕВ ИНФОРМАТИВНОСТИ ЛОГИЧЕСКИХ ЗАКОНОМЕРНОСТЕЙ

Воронцов К. В., Ивахненко А. А.  
Вычислительный центр РАН, Россия  
[vokov@forecsys.ru](mailto:vokov@forecsys.ru)

Логическим правилом или закономерностью называется предикат  $\varphi : X \rightarrow \{0, 1\}$ , заданный на множестве объектов  $X$ , зависящий от небольшого числа признаков, покрывающий достаточно много объектов обучающей выборки  $X^\ell = \{x_i, y_i\}_{i=1}^\ell$  заданного класса  $y \in Y$  и достаточно мало объектов остальных классов:

$$p_y(\varphi, X^\ell) = \#\{x_i : \varphi(x_i) = 1 \text{ и } y_i = y\} \rightarrow \max ;$$

$$n_y(\varphi, X^\ell) = \#\{x_i : \varphi(x_i) = 1 \text{ и } y_i \neq y\} \rightarrow \min .$$

Многие алгоритмы классификации строятся как композиции правил: решающие списки и деревья, голосование правил, и т. д. Для синтеза правил в них применяются различные переборные процедуры: поиск в глубину, поиск в ширину, адаптивный случайный поиск, эволюционный поиск, и другие. В этих алгоритмах используются два критерия информативности правил: на стадии поиска оценивается *перспективность* дальнейшей модификации правила; на стадии финального отбора оценивается *полезность* правила в композиции с другими правилами.

Для оценивания правил изобретены десятки эвристик, в основном это различные свёртки пары величин  $p_y, n_y$ . Только в обзоре [1] их представлено около 20. Возникает вопрос: какие критерии информативности правил являются оптимальными?

Пусть  $X^k = \{x'_i, y'_i\}_{i=1}^k$  — независимая контрольная выборка. Обобщающей способностью правила  $\varphi$  будем называть его *точность* (долю ошибочно покрываемых объектов) на контрольной выборке:

$$v_y(\varphi, X^k) = n_y(\varphi, X^k) / (p_y(\varphi, X^k) + n_y(\varphi, X^k)) .$$

Оптимальным критерием *полезности* правила  $\varphi$  является функция, предсказывающая его обобщающую способность  $v_y(\varphi, X^k)$  по значениям  $p_y(\varphi, X^\ell)$  и  $n_y(\varphi, X^\ell)$ . Оптимальным критерием *перспективности* правила  $\varphi$  является функция, предсказывающая обобщающую способность  $v_y(\psi, X^k)$  наилучшего правила  $\psi$ , полученного путём модификаций из  $\varphi$ , по значениям  $p_y(\varphi, X^\ell)$  и  $n_y(\varphi, X^\ell)$ .

Для построения предсказывающих функций можно использовать процедуру *мета-обучения* [1,2]. Фиксируется набор реальных задач классификации и метод поиска правил. Для каждой задачи выборка разбивается несколькими способами на обучение и контроль, по обучающей части строится достаточно большое количество правил. Полученная мета-выборка, объектами которой являются тройки «задача, разбиение, правило», используется для построения регрессионной зависимости точности  $v_y$  на контроле от значений  $p_y, n_y$  на обучении.

В данной работе предлагается усовершенствованная процедура мета-обучения, имеющая следующие отличия от процедуры Фюрнкранца.

1. Кроме характеристик правила  $p_y, n_y$  в число регрессоров включаются оценки сложности правила: число правил, просмотренных в процессе перебора перед тем, как найдено данное правило; из них число правил с достаточно высокой точностью на обучающей выборке; число термов в правиле; коэффициент разнообразия (функция роста) множества правил с данным числом термов.

2. Мета-обучение делается для каждой задачи и каждого класса в отдельности, чтобы учесть индивидуальные особенности задачи.

3. Вместо простейших жадных стратегий нисходящего поиска и покрытия выборки применяется «полужадный» поиск в ширину.

4. Вид регрессионной модели диктуется сложностными оценками обобщающей способности. Поскольку эти оценки сильно завышены, сложностный множитель  $\Delta$  заменяется на  $a\Delta^b$ , где параметры  $a$  и  $b$  определяются в результате решения регрессионной задачи. Это позволяет заодно оценить степень завышенности классических оценок.

5. Наряду с оценкой средней обобщающей способности (точности правил на контроле) строится её верхняя доверительная граница.

Эксперимент на 7 реальных задачах из репозитория UCI показал, что применение мета-обучения повышает обобщающую способность как отдельных правил, так и их композиции. Использование информации о сложности правил, накопленной в процессе поиска, позволяет ещё точнее предсказывать обобщающую способность правил и отбирать лучшие правила.

Работа выполнена при поддержке РФФИ, проект №08-07-00422.

#### **Литература**

1. Fürnkranz J., Flach P. A. Roc 'n' rule learning — towards a better understanding of covering algorithms // *Machine Learning*. — 2005. — Vol. 58, no. 1. — Pp. 39–77.
2. Janssen F., Fürnkranz J. Meta-learning rule learning heuristics // *LWA — Martin-Luther-University Halle-Wittenberg*, 2007. — Pp. 167–174.