Секция «Теория вероятностей и математическая статистика»

Задача поиска скрытого гиперграфа

Полянский Никита Андреевич

Аспирант

Московский государственный университет имени М.В.Ломоносова,
Механико-математический факультет, Кафедра теории вероятностей, Москва, Россия

E-mail: nikitapolyansky@gmail.com

Задача поиска скрытого гиперграфа [5] является естественным обобщением классической задачи группового тестирования. Рассмотрим семейство гиперграфов $\mathcal{H}(t, s, \ell)$, которое будет состоять из гиперграфов $H = (V, E)$ таких, что множество вершин $V = \{1, 2 \ldots t\}$, множество ребер $E = \{e_1, e_2, \ldots, e_{s'}\}$, $s' \le s$, и размер каждого ребра $1 \le |e_i| \le \ell$, причем никакое ребро не содержится ни в каком другом. Предположим, что задан *скрытый* гиперграф $H_{un} = (V, E)$, ребра которого нам неизвестны, но мы знаем, что $H_{un} \in \mathcal{H}(t, s, \ell)$. Наша цель – обнаружить ребра E этого гиперграфа, спросив N *вопросов* $Q(S)$, где множество S - это некоторое подмножество V, а ответ на вопрос *положительный*, т.е. $Q(S) = 1$, в случае, если множество S содержит полностью хотя бы одно ребро из E. В остальных случаях, ответ на вопрос *отрицательный*, т.е. $Q(S) = 0$. Будем называть поиск *неадаптивным*, если все вопросы заранее спланированы и задаются одновременно. Если же вопросы задаются последовательно, и последующие вопросы зависят от ответов на предыдущие, то поиск называют *адаптивным*.

Определим *асимптотическую скорость* оптимального поиска скрытого гиперграфа как

$$R(s, l) = \overline{\lim_{t\to\infty}} \frac{\log_2 t}{N(t, s, l)},$$

где число $N(t, s, l)$ равно минимальному числу вопросов, необходимых для нахождения произвольного гиперграфа $H_{un} \in \mathcal{H}(t, s, \ell)$. Будем приписывать верхний индекс a и na в зависимости от адаптивности поиска.

Теорема 1. *Скорость $R^{na}(s, l)$ удовлетворяет асимптотическому неравенству*

$$\left(\frac{\ell+1}{e}\right)^{\ell+1} \frac{\log_2 s}{s^{\ell+1}}(1 + o(1)) \le R^{na}(s, \ell) \le \frac{(\ell+1)^{\ell+1}}{2\, e^{\ell-1}} \frac{\log_2 s}{s^{\ell+1}}(1 + o(1)), \quad \ell \ge 2, \quad s \to \infty.$$

Доказательство теоремы 1 можно разбить на две части: во-первых, можно показать [3], что задача неадаптивного поиска скрытого гиперграфа и свободные от перекрытия коды сильно связаны между собой, а, во-вторых, уже для скорости свободных от перекрытий кодов верна соответствующая нижняя граница [2], а также верхняя граница [3].

Теорема 2. *Скорость $R^a(s, \ell)$ достигает теоретико-информационную границу, т.е.*

$$R^a(s, \ell) = \frac{1}{s\ell}.$$

Теорема 2 представляет собой уточнение основного результата работы [4], в которой доказано, что $R^a(s, \ell) \ge 1/(2s\ell)$.

Источники и литература

1) Лебедев В.С., “Асимптотическая верхняя граница скорости кодов, свободных от (w,r)-перекрытий ”, Пробл. передачи информ. (2003) 39:4, 3-9.
2) А.Г. Дьячков, И.В. Воробьев, Н.А. Полянский, В.Ю. Щукин, “Границы скорости дизъюнктивных кодов”. Пробл. передачи информ, (2014) 50:1, 31-63.

3) Dyachkov A., Vilenkin P., Macula A., Torney D., "Families of Finite Sets in Which No Intersection of l Sets Is Covered by the Union of s Others", J. Combin. Theory. Ser. A. (2002) v. 99, pp. 195-218.

4) Abasi, H., Bshouty, N.H., and Mazzawi, H., "On Exact Learning Monotone DBF from Membership Queries", Lecture Notes in Artificial Intelligence, (2014) pp. 111-124.

5) Angluin, D., and Chen, J. "Learning a hidden hypergraph. Journal of Machine Learning Research 7, (2006) pp. 2215-2236.