EDGE SEARCH IN GRAPHS AND HYPERGRAPHS

Ingo Althöfer and Matthias Löwe

Fakultät für Mathematik
Universität Bielefeld
Postfach 100131
33501 Bielefeld
Germany

Abstract

Consider an undirected hypergraph $H = (X, E)$ with a probability distribution P on the set E of its hyperedges. We investigate the average case complexity $\overline{L}(H, P)$ of finding an unknown hyperedge $e^* \in E$, chosen according to P, if only tests are allowed that check, whether e^* is contained in the induced subhypergraphs $H[Y]$ for $Y \subset X$, or not.

In this note we prove that for hypergraphs with average size of the hyperedges r $H(P) \leq \overline{L}(H, P) < H(P) + 3r$. Here $H(\cdot)$ denotes the usual entropy function. The lower bound is the information theoretic bound, whereas the upper bound stems from a construction presented in this note.

More precisely, our construction gives search lengths which are bounded from above by $-\log_2 P(e^*) + 3 \cdot |e^*|$ for all possible choices of e^*.

1. Introduction

In [Aig 86] the following search problem on graphs was introduced which generalizes group testing problem with two defectives.

Given a finite graph $G = (X, E)$ without loops and multiple edges, we search for an edge $e^* \in E$. Admissible tests are the following: For any subset $Y \subset X$ we may test whether e^* is contained in the induced subgraph $G[Y]$ or not. For the worst case complexity $L(G)$ of this problem the information theoretic lower bound is $L(G) \geq \lceil \log_2 |E| \rceil$ and there are graphs with $L(G) = \lceil \log_2 |E| \rceil + 1$. Recently Damaschke [Dam 93] proved that indeed for all graphs G

$$L(G) \leq \lceil \log_2 |E| \rceil + 1.$$

In this note we generalize from the above problem in two ways. First we consider hypergraphs $H = (X, E)$ instead of graphs. Second we investigate the average search length in case of a probability distribution $P = \big(P(e_1) \ldots, P(e_m)\big)$ $(m = |E|)$ on the set E of hyperedges. For a fixed search scheme and every $e \in E$ we denote by $\mathrm{cost}(e)$ the number of questions required in case of $e^* = e$. For given H and P let $\overline{L}(H, P)$ be the minimum possible value of the average search cost $\sum_{e \in E} P(e)\mathrm{cost}(e)$. Recall, that the information theoretic lower bound is

$$\overline{L}(H, P) \geq H(P)$$

where $H(P) = -\sum_{i=1}^{m} P(i) \log_2 P(i)$ denotes the entropy of P. Naively we might hope to derive an upper bound for $\overline{L}(H, P)$ such that $\overline{L}(H, P) \leq H(P) + c$ with a constant c independent of H (especially with c independent of $|E|$). The example, where $|X| = n$, E consists of all hyperedges of size $n - 1$, and P is the uniform distribution on E, such that the average search cost are obviously $\frac{n}{2}$, shows that this hope is not very sensible without any further requirements.

In fact we are able to give such an upper bound on $\overline{L}(H, P)$ under the weak hypothesis that the average size of the hyperedges is known, i.e.

$$\sum_{e \in E} P(e)|e| = r \tag{1}$$

for some $r < \infty$ (here $|e|$ stands for the number of vertices in e). Indeed we prove

Theorem 1. *For every hypergraph $H = (X, E)$ with a probability distribution P on E there is a search scheme yielding*

$$\textit{cost}(e) \leq -\log_2 P(e) + 3|e| \quad \textit{for every} \quad e \in E.$$

An immediate consequence is: