Rounds in Combinatorial Search

Extended abstract

Gábor Wiener[1]

Budapest University of Technology and Economics,
Department of Computer Science and Information Theory
H-1521, Budapest, Hungary
wiener@cs.bme.hu

Abstract. The search complexity of a separating system $\mathcal{H} \subseteq 2^{[m]}$ is the minimum number of questions of type "$x \in H$?" (where $H \in \mathcal{H}$) needed in the worst case to determine a hidden element $x \in [m]$. If we are allowed to ask the questions in at most k batches then we speak of the k-*round* (or k-*stage*) complexity of $\mathcal{H}$, denoted by $\mathrm{c}_k(\mathcal{H})$. While 1-round and m-round complexities (called non-adaptive and adaptive complexities, respectively) are widely studied (see for example Aigner [1]), much less is known about other possible values of k, though the cases with small values of k (tipically $k = 2$) attracted significant attention recently, due to their applications in DNA library screening. It is clear that $|\mathcal{H}| \geq \mathrm{c}_1(\mathcal{H}) \geq \mathrm{c}_2(\mathcal{H}) \geq \ldots \geq \mathrm{c}_m(\mathcal{H})$. A group of problems raised by G. O. H. Katona [6] is to characterize those separating systems for which some of these inequalities are tight. In this paper we are discussing set systems $\mathcal{H}$ with the property $|\mathcal{H}| = \mathrm{c}_k(\mathcal{H})$ for any $k \geq 3$. We give a necessary condition for this property by proving a theorem about traces of hypergraphs which also has its own interest.

Keywords. Search, group testing, adaptiveness, hypergraph, trace

1 Preliminaries

We denote the set of the first m positive integers by $[m]$. A set system $\mathcal{A} \subseteq 2^{[m]}$ is said to be a *separating system* if for any pair of distinct elements $x, y \in [m]$ there exists a set in $\mathcal{A}$ that contains exactly one of them. A separating system $\mathcal{A}$ is *minimal* if no $\mathcal{B} \subset \mathcal{A}$ is separating.

A *hypergraph* is a pair $(V, \mathcal{E})$, where V is a finite set, called the *vertices* of the hypergraph and $\mathcal{E}$ is a collection of subsets of V, called the *(hyper)edges* of the hypergraph. Notice that $\mathcal{E}$ is not necessarily a set, that is, hyperedges may have multiplicity greater than 1. If every hyperedge has multiplicity 1 then the hypergraph is called *simple*. It is obvious that edge sets of simple hypergraphs and set systems are the same. If the restriction of a simple hypergraph to any proper subset of the vertices is not simple then we speak of a *minimal simple* hypergraph. The set of all minimal simple hypergraphs on the vertex set $[n]$ having m hyperedges is denoted by $MSH(n, m)$. The *multiplicity* of a set of

Dagstuhl Seminar Proceedings 09281
Search Methodologies
http://drops.dagstuhl.de/opus/volltexte/2009/2239

vertices X in a hypergraph $\mathcal{H}$ is the number of occurences of X as an edge and is denoted by $m_{\mathcal{H}}(X)$. Sometimes, if it does not cause any misunderstanding we identify hypergraphs by their edge set.

Let $\mathcal{H} = (V, \mathcal{E})$ be a hypergraph and consider any linear order of V and $\mathcal{E}$. The *incidence matrix* of $\mathcal{H}$ is a 0-1 matrix $M_{\mathcal{H}} = (m_{ij})_{|\mathcal{E}|,|V|}$, where m_{ij} is 1 if and only if the i^{th} edge contains the j^{th} vertex. The incidence matrix of a set system $\mathcal{A} \subseteq 2^S$ is defined as the incidence matrix of the simple hypergraph having vertex set S and edge set $\mathcal{A}$ and is denoted by $M_{\mathcal{A}}$. It is obvious that any row and column permutation of an incidence matrix of a hypergraph (set system) is also an incidence matrix of the same hypergraph (set system) and that any 0-1 matrix is an incidence matrix of some hypergraph. The *dual* of a hypergraph $\mathcal{H}$ is the hypergraph $\mathcal{H}^*$ whose incidence matrix is $M_{\mathcal{H}}^T$. The dual of a set system $\mathcal{A}$ is the collection of edges of the hypergraph whose incidence matrix is $M_{\mathcal{A}}^T$. Note that $\mathcal{A}^*$ is not necessarily a set system.

It is obvious that a hypergraph $\mathcal{H}$ is simple if and only if $M_{\mathcal{H}}$ has no identical rows and that a set system $\mathcal{A}$ is separating if and only if $M_{\mathcal{A}}$ contains no identical columns.

A set system $\mathcal{A}$ of cardinality $k+1$ is called a *k-star* if it contains a set A such that for any $B \in \mathcal{A}$, $B \neq A : A \subseteq B$ and $|B \setminus A| = 1$.

A set system $\mathcal{A} \subseteq 2^{[m]}$ is said to be *hereditary* if $A \in \mathcal{A}$ and $B \subseteq A$ implies $B \in \mathcal{A}$.

A set system $\mathcal{A} \subseteq 2^{[m]}$ is said to be a *representation* of a set system $\mathcal{B} \subseteq 2^{[m]}$ if there exist a linear order of the sets of $\mathcal{A}$ $(A_1, A_2, \dots, A_r)$ and $\mathcal{B}$ $(B_1, B_2, \dots, B_r)$ and a permutation π of the elements of $[m]$, such that for any $i \leq r = |\mathcal{B}|$ we have either $A_i = \{\pi(j) \ : \ j \in B_i\}$ or $A_i = \{\pi(j) \ : \ j \notin B_i\}$. In other words, $\mathcal{A}$ is a representation of $\mathcal{B}$ if they have the same cardinality and their incidence matrices can be transformed to each other by row and column permutations and by complementing some rows (but not columns), where complementing a row means that we change the 1 entries of the row to 0 entries and vice versa.

2 Introduction

Let $\mathcal{H} \subseteq 2^{[m]}$ be an arbitrary separating system (called the *question sets*) and $x \in [m]$ an unknown element. Our aim is to find x by asking questions of type "$x \in H$?", where $H \in \mathcal{H}$. A sequence of questions is called a *search algorithm* (or shortly an algorithm) if given the answers we can determine x uniquely.

An algorithm is said to be *adaptive* (or *dynamic*) if the choice of a question set may depend on the values obtained until then. If the questions are all fixed beforehand then we speak of a *non-adaptive* (or *static*) algorithm. More generally, if we are allowed to ask the questions in at most k batches (that is, we ask some questions, receive the answers, ask again some questions, receive the answers, and so on, at most k times) then we speak of a *k-round* (or *k-stage*) algorithm.

The length of an algorithm $\mathbf{A}$ for the element x, denoted by $l_x(\mathbf{A})$ is l if the sequence contains l questions and the first $l-1$ answers does not determine x

uniquely. The (worst case) cost of an algorithm $\mathbf{A}$ is $g(\mathbf{A}) = \max_{x\in[m]} l_x(\mathbf{A})$. The *adaptive (search) complexity* of the set system $\mathcal{H}$ is $\mathrm{c}(\mathcal{H}) = \min g(\mathbf{A})$ considering all adaptive algorithms $\mathbf{A}$. The *non-adaptive*, and *k-round complexities* are defined similarly and are denoted by $\mathrm{c}_{na}(\mathcal{H})$ and $\mathrm{c}_k(\mathcal{H})$, respectively. Notice that since $\mathcal{H}$ is separating, these definitions are correct. For a detailed treatment of adaptive and non-adaptive search the reader is referred to the book by Aigner [1]. Much less is known about k-round search for arbitrary values of k, though the cases with small values of k (tipically $k = 2$) attracted significant attention recently, due to their applications in DNA library screening.

It is obvious that

$$|\mathcal{H}| \geq \mathrm{c}_{na}(\mathcal{H}) = \mathrm{c}_1(\mathcal{H}) \geq \mathrm{c}_2(\mathcal{H}) \geq \ldots \geq \mathrm{c}_m(\mathcal{H}) = \mathrm{c}(\mathcal{H}). \tag{1}$$

A problem raised by G. O. H. Katona [6] is to characterize those separating systems $\mathcal{H} \subseteq 2^{[m]}$ for which certain inequalities of (1) are tight. In the present paper our aim is to examine separating systems $\mathcal{H} \subseteq 2^{[m]}$ with the property $|\mathcal{H}| = \mathrm{c}_k(\mathcal{H})$ for any $k \geq 3$.

More precisely, we will show that if $\mathrm{c}_k(\mathcal{H}) = |\mathcal{H}|$ for some $k \geq 3$ then the dual of $\mathcal{H}$ contains a $\lceil \frac{n^2}{2m-n-2} \rceil$-star, where $n = |\mathcal{H}|$.

3 Results

We would like to examine separating systems $\mathcal{H} \subseteq 2^{[m]}$ for which $\mathrm{c}_k(\mathcal{H}) = |\mathcal{H}|$ for some $k \geq 3$. This condition implies $\mathrm{c}(\mathcal{H}) = |\mathcal{H}|$, from which $|\mathcal{H}| \leq m - 1$ follows easily. It is more interesting that even $\mathrm{c}_1(\mathcal{H}) = |\mathcal{H}|$ implies $|\mathcal{H}| \leq m - 1$, in other words, a minimal separating system $\mathcal{H} \subseteq 2^{[m]}$ contains at most $m - 1$ sets, as it was first observed by Bondy [4]. Notice that both results are sharp, just consider $\mathcal{H} = \{\{1\}, \{2\}, \ldots, \{m - 1\}\}$.

Using Bondy's result it is not difficult to characterize those systems whose k-round complexity is $m - 1$ for any $k \geq 2$.

Lemma 1. *Let $k \geq 2$. For a separating system $\mathcal{H} \subseteq 2^{[m]}$, $c_k(\mathcal{H}) = m - 1$ if and only if $\mathcal{M} = \{\{1\}, \{2\}, \ldots, \{m - 1\}\}$ is a representation of $\mathcal{H}$.*

The main theorem of this paper is the following.

Theorem 1. *Let $\mathcal{H} \subseteq 2^{[m]}$ be a separating system for which $c_k(\mathcal{H}) = |\mathcal{H}|$ for some $k \geq 3$ and let $n = |\mathcal{H}|$. Then $\mathcal{H}^*$ contains a $\lceil \frac{n^2}{2m-n-2} \rceil$-star.*

Proof. If for some $k \geq 3$ we have $\mathrm{c}_k(\mathcal{H}) = |\mathcal{H}|$ then $\mathrm{c}_3(\mathcal{H}) = |\mathcal{H}|$. We show that this implies that $\mathcal{H}^*$ contains a $\lceil \frac{n^2}{2m-n-2} \rceil$-star.

The proof is based on the following theorem about hypergraphs.

Theorem 2. *Let $\mathcal{A} \in MSH(n, m)$. Then there exists a subset $X \subseteq [n]$ of cardinality $\lceil \frac{n^2}{2m-n-2} \rceil$, such that deleting X we obtain a hypergraph where every hyperedge has multiplicity at most $\lceil \frac{n^2}{2m-n-2} \rceil + 1$.*

The sketch of the proof of Theorem 2 can be found in Section 4.

Let us denote the number $\lceil \frac{n^2}{2m-n-2} \rceil$ by r. Consider now the set system $\mathcal{H}^*$. Since $\mathcal{H}$ is separating, $\mathcal{H}^*$ is also a set system (that is, it contains distinct sets), in other words it is the hyperedge set of a simple hypergraph $\mathcal{G}$ on the vertices corresponding to the sets of $\mathcal{H}$. Observe now that $\mathrm{c}_{na}(\mathcal{H}) = |\mathcal{H}|$ (because $\mathrm{c}_3(\mathcal{H}) = |\mathcal{H}|$), so $\mathcal{H}$ is a minimal separating system, thus $\mathcal{G}$ is a minimal simple hypergraph having n vertices and m hyperedges. Now applying Theorem 2 for $\mathcal{G}$ we see that there exists a subset of the vertices X, $|X| = r$, such that deleting X we obtain a hypergraph where every hyperedge has multiplicity at most $r+1$. This subset X of the vertices of $\mathcal{H}^*$ correspond to a subset $\mathcal{X}$ of the original set system $\mathcal{H}$. Considering the incidence matrix of $\mathcal{H}$ one can see that deleting the rows corresponding to $\mathcal{X}$ we obtain a matrix where every column appears at most $r+1$ times.

Suppose now that we ask the sets of $\mathcal{H} \setminus \mathcal{X}$ in the first round of a 3-round search algorithm. Given the answers we know that the unknown element is one from a set $Y \subseteq [m]$, where $|Y| \leq r+1$, because no column appears more than $r+1$ times in the incidence matrix of $\mathcal{H} \setminus \mathcal{X}$.

Since $\mathrm{c}_3(\mathcal{H}) = |\mathcal{H}|$, we have to ask all the remaining sets of $\mathcal{H}$ in two more rounds to determine the hidden element. That is, we have to ask $|\mathcal{X}| = r$ sets in two rounds to find an element in Y, which has at most $r+1$ elements. By Lemma 1 this is possible if and only if $Y = r+1$ and the restriction of $\mathcal{X}$ to Y contains only one-element sets. In other words, the incidence matrix of the restriction of $\mathcal{X}$ to Y is an $r \times r$ identity matrix plus an all-zero column. Since for the elements of Y we received the same answers in the first round, these elements form an r-star in $\mathcal{H}^*$.

4 Sketch of proof of Theorem 2

Let $\mathcal{H}$ be a hypergraph on the vertex set $[n]$. Let us denote the hypergraph obtained from $\mathcal{H}$ by deleting a subset X of the vertices (that is, taking the restriction of $\mathcal{H}$ to $\overline{X} = [n] \setminus X$) by $\mathcal{H}|_{\overline{X}}$. Recall that $m_{\mathcal{H}}(E)$ denotes the multiplicity of the hyperedge E in the hypergraph $\mathcal{H}$.

The following lemma can be proved using the down-compression technique of Alon [2] and Frankl [5].

Lemma 2. *The following two statements are equivalent.*

1. *For every $\mathcal{A} \in MSH(n,m)$ there exists a set $X \subseteq [n]$ of cardinality r, such that for any set $S \subseteq \overline{X}$ we have $m_{\mathcal{A}|_{\overline{X}}}(S) \leq s$.*
2. *For every hereditary $\mathcal{A} \in MSH(n,m)$ there exists a set $X \subseteq [n]$ of cardinality r, such that for any set $S \subseteq \overline{X}$ we have $m_{\mathcal{A}|_{\overline{X}}}(S) \leq s$.*

By Lemma 2 we only have to prove that for a hereditary minimal simple hypergraph having n vertices and m hyperedges there exists a subset of the vertices X of cardinality $\lceil \frac{n^2}{2m-n-2} \rceil$, such that deleting X we obtain a hypergraph where every hyperedge has multiplicity at most $\lceil \frac{n^2}{2m-n-2} \rceil + 1$.

Let $\mathcal{A}$ be such a hypergraph. Observe that every vertex v is contained in some hyperedge, otherwise $\mathcal{A}$ would not be minimal, thus by the hereditary property all 1-element sets are hyperedges of $\mathcal{A}$.

This means that the number of hyperedges of $\mathcal{A}$ containing at least two elements is $m-n-1$ (since $\mathcal{A}$ contains n 1-element hyperedges and also the empty set). Consider now the graph G on the vertex set $[n]$ whose edges are the 2-element sets of $\mathcal{A}$. G has n vertices and at most $m-n-1$ edges, thus by a corollary of Turán's theorem [7], [3, p. 282.] it contains a stable set X of size $\lceil \frac{n^2}{2(m-n-1)+n} \rceil = \lceil \frac{n^2}{2m-n-2} \rceil$. We show that $m_{\mathcal{A}|_{\overline{X}}}(S) \leq \lceil \frac{n^2}{2m-n-2} \rceil + 1$ for any $S \subseteq [n]$. Actually, it suffices to show that $m_{\mathcal{A}|_{\overline{X}}}(\emptyset) \leq \lceil \frac{n^2}{2m-n-2} \rceil + 1$, since by the hereditary property of $\mathcal{A}$ we have $m_{\mathcal{A}|_{\overline{X}}}(\emptyset) \geq m_{\mathcal{A}|_{\overline{X}}}(S)$ for any $S \subseteq [n]$.

By definition, $m_{\mathcal{A}|_{\overline{X}}}(\emptyset) = |\{A \in \mathcal{A} : A \subseteq X\}| = |\mathcal{A}|_X|$.

If $i, j \in X$ $(i \neq j)$, then $\{i, j\} \notin \mathcal{A}$, since X is stable in G. Furthermore, there is no hyperedge in $\mathcal{A}$ that contains both i and j, because $\mathcal{A}$ is hereditary. Thus $\mathcal{A}|_X$ does not contain sets of size greater than 1, so the number of distinct sets in $\mathcal{A}|_X$ is at most $|X| + 1 = \lceil \frac{n^2}{2m-n-2} \rceil + 1$.

References

1. M. Aigner. Combinatorial Search, John Wiley, Chichester and Teubner, Stuttgart, 1988
2. N. Alon. On the density of sets of vectors, *Discrete Mathematics* **46** (1983), 199–202.
3. C. Berge. Graphs and Hypergraphs, North Holland,
4. J. A. Bondy. Induced subsets, *Journal of Combinatorial Theory, Series B* **12** (1972), 201–202.
5. P. Frankl. On the trace of finite sets, *Journal of Combinatorial Theory, Series A* **34**, (1983), 41–45.
6. G. O. H. Katona. Personal communication, 2004
7. P. Turán. On an extremal problem in graph theory, (in Hungarian) *Mat. Fiz. Lapok* (1941) **4**8, p. 436-452.