Colloques Internationaux du C.N.R.S.
N° 276. – THEORIE DE L'INFORMATION

SEARCH USING SETS WITH SMALL INTERSECTION

Gyula O.H. KATONA

We consider the following type of questionnaires. Let X be a set of cardinality n, and x an unknown element of X. Let $\mathcal{A} = \{A_1, \ldots, A_m\}$ a family of subsets of X. The informations relative to the belonging of x to A_i ($1 \leqslant 1 \leqslant m$) are given. These informations must be sufficient to identify an unknown element of X. The aim of the work is to determine the minimum of m if we impose different conditions on $\mathcal{A}$. The case $|A_i| \leqslant k$ is solved in [1]. In this paper, we consider the case $|A_i \cap A_j| \leqslant k$, for every $A_i, A_j \in \mathcal{A}$.

Recherches utilisant des codes de petite intersection

On considère le type suivant de questionnaire. Soit X un ensemble de cardinalité n et x un élément inconnu de X. Soit $\mathcal{A} = \{A_1, \ldots, A_m\}$ une famille de parties de X. Les informations concernant l'appartenance de x à A_i ($1 \leqslant i \leqslant m$) sont à notre disposition. La famille $\mathcal{A}$ doit posséder la propriété que ces informations suffisent à identifier un élément quelconque de X. L'objet de l'examen est de déterminer le minimum de m si on pose sur $\mathcal{A}$ différentes conditions. Le cas $|A_i| \leqslant k$ est résolu dans [1]. Dans le présent article on considère le cas $|A_i \cap A_j| \leqslant k$ (pour chaque $A_i, A_j \in \mathcal{A}$).

INTRODUCTION

Let X be a finite set of n elements and x be an unknown element of X. We can have the following type of information on x. For some subsets $A \subset X$ we may ask if $x \in A$ or not. Using this type of information we have to determine uniquely the unknown element x. There are many practical situations where this model is more or less correct. We say that the *search* is *sequential* if the choice of the next question (subset) can depend on the answers of the previous questions.

Otherwise the search is called *unsequential*. In this paper we shall consider the latter type, only. In this case the search is simply a family $\mathcal{A} = \{A_1, \ldots, A_m\}$ of subsets with the property that the information "$x \in A_i$ or $x \notin A_i$ ($1 \leqslant i \leqslant m$)" determines x uniquely. The mathematical problem is to minimize m under some constrains on $\mathcal{A}$. In [1] this problem has been solved when the constrain is $|A_i| \leqslant k$ ($1 \leqslant i \leqslant m$; $k \leqslant n/2$ fixed; $|\ |$ denotes the number of elements. In the present paper the following constrain is considered:

$$|A_i \cap A_j| \leqslant k \quad (1 \leqslant i, j \leqslant n, i \neq j, k \leqslant n/4).$$

The case $k = 1$ is completely solved. If $k \leqslant c\sqrt[3]{n}$ we have fairly good lower and upper estimation. The other cases are practically unsolved.

It is worth-while to mention that under the constrain $A_i \leqslant k$ the best sequential search (minimizing the maximum number of necessary questions) is determined in [2]. However it is an open question under the present constrain ($|A_i \cap A_j| \leqslant k$ can be assumed either for all the possible questions or only for the questions occuring at a given x).

LOWER ESTIMATIONS

Let X be a finite set and A be a subset of X. We say that A *separates* two elements x and y of X if A contains exactly one of them. The family $\mathcal{A} = \{A_1, \ldots, A_m\}$ of subsets of X is called a *separating system* if any pair $x, y \in X$ $(x \neq y)$ is separated by some A_i $(1 \leqslant i \leqslant m)$. It is easy to see tha the information "$x \in A_i$ or $x \notin A_i$ $(1 \leqslant i \leqslant m)$" determines x uniquely iff $\mathcal{A}$ is a separating system. Consequently, our aim is to minimize m under the conditions that

a) $\mathcal{A}$ is a separating system on an n-element set X,

b) $|A_i \cap A_j| \leqslant k$ $(1 \leqslant i, j \leqslant m, i \neq j, k \leqslant n/4)$.

LEMMA 1 – Let $\mathcal{A} = \{A_1, \ldots, A_m\}$ *be a family of subsets of the set*

$$X = \{x_1, \ldots, x_n\}$$

$(n \geqslant 1)$, *satisfying (a) and (b). Then m is not smaller than the minimal m satisfying*

$$\sum_{i=0}^{r} \binom{m}{i}\binom{i}{2} + \left(n - \sum_{i=0}^{r}\binom{m}{i}\right)\binom{r+1}{2} \leqslant k\binom{m}{2} \tag{1}$$

and

$$\sum_{i=0}^{r}\binom{m}{i} \leqslant n < \sum_{i=0}^{r+1}\binom{m}{i} \tag{2}$$

for some r $(1 \leqslant r \leqslant m$; *for* $r = m$ (2) *should be understood as* $\sum_{i=0}^{m}\binom{m}{i} = n$.

Proof. 1) Let $B = (b_{ij})$ be the incidence matrix of $\mathcal{A}$, that is, $b_{ij} = 1$ if $x_j \in A_i$ and 0 otherwise. Denote by s_i the number of columns of B having exactly i 1's $(0 \leqslant i \leqslant m)$. Then the next inequalities can be verified:

$$0 \leqslant s_i \leqslant \binom{m}{i} \tag{3}$$

$$\sum_{i=0}^{m} s_i = n \tag{4}$$

$$\sum_{i=0}^{m} s_i \binom{i}{2} \leqslant k \binom{m}{2}. \tag{5}$$

(3) follows from (a): Since any two elements of X are separated by an A_i, to any two columns there is a row (the i^{th}) what contains 1 in exactly one of these columns. It means that the columns are different. The maximum number of different columns with i 1's is $\binom{m}{i}$.

(4) is obvious. (5) follows from (b): The number of pairs $b_{ij} = b_{i'j} = 1$ can not exceed k for any fixed (i, i'). Consequently, their total number (left hand side of (5)) can not exceed $k\binom{m}{2}$.

2) Let $m, s_0, s_1, \ldots, s_m$ satisfy (3)-(5) and suppose $s_i < \binom{m}{i}$, $s_{i+1} > 0$ for some $0 \leqslant i < m$. We shall prove that $m, s_0, s_1, \ldots, s_i + 1, s_{i+1} - 1, \ldots, s_m$ also satisfy (3)-(5). (3) and (4) are satisfied trivially. The left hand side of (5) is increased by

$$\binom{i}{2} - \binom{i+1}{2} \leqslant 0$$

that is, it remains true.

3) By repeated application of the previous result we arrive at the situation when

$$s_i = \binom{m}{i} \qquad 1 \leqslant i \leqslant r$$

$$s_{r+1} = n - \sum_{i=0}^{r} \binom{m}{i}$$

and $s_i = 0$, $r + 1 < i \leqslant m$ for some r. For these numbers (3)-(5) is equivalent to (1)-(2).

LEMMA 2 – The minimal m of the solution of (1) *and* (2) *is attained for the maximal r for which there is a solution in m at all.*

Proof. At a fixed r the set of solutions of (1)-(2) is contained in the set of m's satisfying (2). However, this is an interval of integers and moreover these intervals are disjoint for different r's. Finally if $r > r'$ then the corresponding intervals are ordered in the opposite way. The statement of the lemma immediately follows.

LEMMA 3 – If $1 < k < \frac{2}{\sqrt[3]{36}} \sqrt[3]{n}$ *then the minimal m satisfying* (1) *and* (2) *is the minimal m satisfying*

$$n \leqslant 1 + m + \binom{m}{2} + (k-1)\frac{\binom{m}{2}}{3}. \tag{6}$$

Proof. 1) Let us verify that

$$n \leqslant k \binom{m}{2} \tag{7}$$

follows form (1) and (2) if $r \geqslant 3$ and $m \geqslant 4$. Indeed, the left hand side of (2) implies

$$\sum_{i=0}^{r} \binom{m}{i} \frac{\binom{r+1}{2} - \binom{i}{2}}{\binom{r+1}{2} - 1} \leqslant n$$

(supposing $r \geqslant 3, m \geqslant 4$). Hence we obtain

$$n \leqslant \sum_{i=0}^{r} \binom{m}{i}\binom{i}{2} + \left(n - \sum_{i=0}^{r} \binom{m}{i}\right)\binom{r+1}{2},$$

and (7) follows from (1). (The case $r = 3, m = 3$ can be checked by easy computations).

2) From (7) we have $\sqrt{2n/k} \leqslant m$, that is, by the supposition on k,

$$m > \sqrt[3]{6n}. \tag{8}$$

If $r \geqslant 3$, then $n \geqslant 1 + m + \binom{m}{2} + \binom{m}{3} \geqslant \frac{m^3}{6}$ follows from (2) and this contradicts (8). We have proved that under the conditions of the lemma $r \leqslant 2$ must hold for any solutions of (1)-(2).

3) For $r = 2$ (1) and (2) are of the from

$$n \leqslant 1 + m + \frac{k+2}{3}\binom{m}{2} \tag{9}$$

and

$$1 + m + \binom{m}{2} \leqslant n < 1 + m + \binom{m}{2} + \binom{m}{3}. \tag{10}$$

Let us verify that

$$1 + m + \frac{k+2}{3}\binom{m}{3} < 1 + m + \binom{m}{2} + \binom{m}{3} \tag{11}$$

for any solution of (9)-(10), supposing $k < \frac{2}{\sqrt[3]{36}} \sqrt[3]{n}$. (11) holds iff $k - 1 < m - 2$. Assume, in the contrary that $k + 1 \geqslant m$. Hence and from (10) we have

$$n < 1 + m + \binom{m}{2} + \binom{m}{3} = \frac{m^3 + 5m + 6}{6} \leqslant \frac{(k+1)^3 + 5(k+1) + 6}{6}.$$

The supposition of the lemma is $\frac{9}{2} k^3 < n$. Comparing it with the above inequality we obtain

$$\frac{9}{2}k^3 < \frac{k^3 + 3k^2 + 8k + 12}{6}$$

and this is a contradiction for any $k > 0$.

Thus we have proved that we have to consider

$$1 + m + \binom{m}{2} \leqslant n \leqslant 1 + m + \frac{k+2}{3}\binom{m}{2}$$

rather than (9)-(10)

4) We prove now that (12) always have a (n integer) solution in m (supposing $1 < k < \frac{2}{\sqrt[3]{36}} \sqrt[3]{n}$).

For a given m, $1 + m + \binom{m}{2}$ and $1 + m + \frac{k+2}{3}\binom{m}{2}$ determine an interval. We have to see that these intervals cover all the natural numbers. Indeed, it is easy to see that

$$1 + (m+1) + \binom{m+1}{2} \leqslant 1 + m + \frac{k+2}{3}\binom{m}{2}$$

holds if $k \geqslant 2$ and $m \geqslant 8$. However, the beginning of the interval corresponding to $m = 8$ is 37. The intervals cover all the integers $n \geqslant 37$. If $n \leqslant 36$ there is no k satisfying $1 < k < \frac{2}{\sqrt[3]{36}} \sqrt[3]{n}$.

We have proved that (1)-(2) has no solution for $r \geqslant 3$, and it has a solution for $r = 2$. On the other hand (1)-(2) is equivalent to (12). By lemma 2 this means that we have to consider the minimal m, satisfying (12). Our lemma is proved.

Case $k = 1$

The left hand side of (1) $> \binom{m}{2}\binom{2}{2}$ if $r \geqslant 3$. For $k = 1$ this contradicts (1). If $r = 2$, the contradiction is avoided only in the case $n = 1 + m + \binom{m}{2}$. It means that if n can be written in the form $1 + m + \binom{m}{2}$ for some integer m, then this m is the minimal solution of (1)-(2) by lemma 2. Otherwise there is no solution for $r \geqslant 2$, and for $r = 1$ we obtain $n \leqslant 1 + m + \binom{m}{2}$, $1 + m \leqslant n < 1 + m + \binom{m}{2}$ from (1) and (2). The minimal m in this case is the minimal m satisfying $n < 1 + m + \binom{m}{2}$.

Summarizing the two cases: the minimal m satisfying (1) and (2) is the minimal m satisfying $n \leqslant 1 + m + \binom{m}{2}$. By lemma 1 this is a lower estimation. This lower estimation is exact. It is easy to construct $\mathcal{A} = \{A_1, \dots, A_m\}$ by constructing the corresponding incidence matrix B. Let B consist of the column containing m $0's$, all the different columns containing one 1 and $m - 1$ $0's$ and $n - 1 - m$ different columns with two $1's$. Here $0 \leqslant n - 1 - m$ follows from the minimality of m (in $n \leqslant 1 + m + \binom{m}{2}$). The matrix obtained in this way obviously satisfies the conditions that it has different columns and any two rows have at least one place with $1 - 1$. The corresponding system $\mathcal{A}$ satisfies (a) and (b). We have proved the following.

THEOREM 1 – If $\mathcal{A} = \{A_1, \dots, A_m\}$ *is a separating system on an n-element set and* $|A_i \cap A_j| \leqslant 1$ $(1 \leqslant i, j \leqslant m, i \neq j)$, *then*

$$\left\{ \frac{\sqrt{8n - 7} - 1}{2} \right\} \leqslant m,$$

(where $\{a\}$ *denotes the least integer* $\geqslant a$*) and this is the best possible estimation.*

Case $k = 2$

Using the theory of Steiner-triple systems we obtain an almost complete solution here.

THEOREM 2 – If $= \{A_1, \dots, A_m\}$ *is a separating system on an n-element set and* $|A_i \cap A_j| \leqslant 2$ $(1 \leqslant i, j \leqslant m, i \neq j)$ *then the minimum of* m *is*

$$\left\{ \frac{\sqrt{24n - 23} - 1}{4} \right\} \tag{13}$$

or

$$\left\{ \frac{\sqrt{24n - 23} - 1}{4} \right\} + 1$$

Proof. 1) It follows from lemma 3 that (13) gives a lower estimation for m, if $n \geqslant 37$. For $n \leqslant 36$ it can be shown by more or less clever computation.

2) Prove now that if $s \leqslant \binom{m}{2} \Big/ 3$, then there is a system of s different triples on m elements, such that any pair is contained in at most 1 triple.

It is known (e.g. [3]) that for one of the numbers $m_0 = m - 1, m, m + 1$ and $m + 2$ there is a Steiner triple system, that is, a system of $\binom{m_0}{2} \Big/ 3$ such that any pair is contained in exactly 1 triples.

If there is a Steiner triple system for $m-1$, let us arbitrarily delete $\binom{m-1}{2}\Big/3-s$ triples and add an isolated point. This construction proves the statement in this case.

If there is a Steiner triple system for m, delete $\binom{m}{2}\Big/3-s$ arbitrarily chosen triples.

If there is a Steiner triple system for $m+1$, delete a point with the $m/2$ triples containing it. Thus we have $\binom{m+1}{2}\Big/3-m/2$ triples on m points with the desired property. However, as this number $\geqslant \binom{m-1}{2}\Big/3$ we can delete some more triples until we have only s.

Finally, if there is a Steiner triple system for $m+2$ we have to delete 2 points with the triples containing them. In this way we obtain $\binom{m+2}{2}\Big/3-m$ triples on m point. As $\binom{m+2}{2}\Big/3-m=\binom{m-1}{2}\Big/3$ this case is also settled.

3) Let $m=\left\{\dfrac{\sqrt{24n-23}-1}{4}\right\}+1$. We shall construct the incidence matrix B by its columns. It will have one column full with $0's$, m columns with one and $\binom{m}{2}$ columns with two $1's$. It is easy to see, that the number of remaining columns

$$s=n-\left(1+m+\binom{m}{3}\right)\leqslant\binom{m-1}{2}\Big/3$$

Choose now s different columns with three $1's$ in each, according to the previous result (section 2 of this proof). It is easy to see that B satisfies the conditions. The theorem is proved.

Case $1<k<c\sqrt[3]{n}$

The same proof works for the following more general case.

THEOREM – $\sqrt[3]{n}$. *Let* $\mathfrak{A}=\{A_1,\ldots,A_m\}$ *be a separating system on an n-element set and*

$$|A_i\cap A_j|\leqslant k\qquad(1\leqslant i,\ j\leqslant m,\ i\neq j)$$

where

$$1<k<\frac{2}{\sqrt[3]{36}}\sqrt[3]{n}$$

Then the minimum of m *is between*

$$M = \left\{ \frac{k - 4 + \sqrt{(k-4)^2 + 24(k+2)(n-1)}}{2(k+2)} \right\}$$

and

$$M + \left\{ \frac{-3M + u + 1 + \sqrt{(3M - u - 1)^2 + 8Mu}}{4} \right\}$$

where $M + u$ *is the first integer* $\geqslant M$ *such that there is a system of triples on an* $M + u$*-element set containing each pair exactly* $k - 1$*-times.*

Open problems

The problem of the paper can be formulated in the following way :

Problem 1 – Let Y be a set of m elements. What is the maximum number of different subsets of Y, such that any pair of elements is contained in at most k subsets?

Another, more hopeful problem needed to the solution of the case $\sqrt[3]{n}$:

Problem 2 – The number n of triples in an m-element set is given. Minimize the maximal valency of the pairs. (The valency of a pair is the number of triples containing it.)

Problem 2a – Is it true that for any m and n, the n triples can be chosen so, taht the valencies differ at most by 2?

REFERENCES

[1] G. KATONA – On separating systems of finite sets, *J. Combinatorial Th. 1* (1966), 174-194.

[2] G. O.H. KATONA – Combinatorial Search Problems, in *A survey of combinatorial theory*, ed. by J.N. Srivastava, North-Holland/American Elsevier Amsterdam – New York, 1973, pp. 285-308.

[3] T.P. KIRKMAN – On a problem in combinations, *Cambridge and Dublin. Math. J. 2* (1847), 191-204.