MATHEMATICAL PROBLEMS IN COMPUTATION THEORY
BANACH CENTER PUBLICATIONS, VOLUME 21
PWN – POLISH SCIENTIFIC PUBLISHERS
WARSAW 1988

THE COMPLEXITY OF CLASSIFICATION PROBLEMS

L. BUDACH

Laboratoire Informatique Théorique et Programmation, Université Paris, Institut de Programmation, Paris
on leave of absence from
Humboldt-Universität, Sektion Mathematik, Berlin, GDR

1. Information systems and classification problems

1.1. An *information system* (see [8]) is a quadrupel $S = (X, A, V, r)$, where X, A, V are finite sets and r is a mapping of $X \times A$ into V. The set X is interpreted as the set of all *objects* under consideration, A is the set of all *attributes*, and V is the set of *descriptors*. The mapping r is the so-called *information function*.

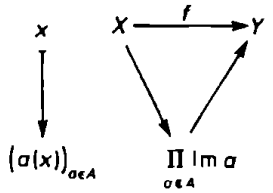
Let $a \in A$ be an attribute; a defines a function

$$X \to V, \quad x \to r(x, a),$$

of X into V. We assume that different attributes define different functions. This enables us to identify the attribute a with the corresponding function and to write $a(x)$ instead of $r(a, x)$. Let $\operatorname{Im} a = \{a(x) | x \in X\}$ be the image of the function a, then a can be considered as a function of X onto $\operatorname{Im} a$. By abuse of language we consider A to be a set of functions a: $X \to \operatorname{Im} a$ and write $S = (X, A)$ instead of $S = (X, A, V, r)$. Let f: $X \to Y$ be a mapping. We call f to be *dependent on* S if the following condition is satisfied:

If $a(x_1) = a(x_2)$ for all $a \in A$, then $f(x_1) = f(x_2)$, $x_1, x_2 \in X$.

The mapping f is dependent on S iff there is a function $\prod_{a \in A} \operatorname{Im} a \to Y$, such that the following diagram is commutative:



A tripel $C = (X, A, f)$ such that (X, A) is an information system, and f: $X \to Y$ is a function dependent on (X, A) is called a *classification problem.* The function f is called the *classifying function* or the *classification.*

By some technical reasons we will assume throughout the following paper, that all information systems satisfy the following condition:

S is fully faithfull, i.e., the function $X \to \prod_{a \in A} \operatorname{Im} a$ is bijective. This implies that all functions f: $X \to Y$ are dependent on S.

1.2. Examples

1.2.1. Every boolean function f: $\{0, 1\}^n \to \{0, 1\}$ defines a classification problem. The underlying information system consists of $X = \{0, 1\}^n$ as set of objects, $A = \{1, 2, \ldots, n\}$ is the set of attributes, $V = \{0, 1\}$ is the set of descriptors, $r = X \times A \to V$ is the selection function defined by $r((x_1, \ldots, x_n), i) = x_i$, and f is the classifying function.

1.2.2. Let Σ be a finite aphabet and L be a language over Σ. L is a subset of Σ^* and defines a subset $L^n := L \cap \Sigma^n$ for any natural number n. Let f_n: $\Sigma^n \to \{0, 1\}$ be the characteristic function of L^n, i.e.,

$$f_n(w) = \textit{if } w \in L^n \textit{ then } 1 \textit{ else } 0.$$

Therefore L defines for every n a classification problem with information function f_n and the underlying information system $(\Sigma^n, \{1, 2, \ldots, n\}, \Sigma, r)$ with $r((\sigma_1, \ldots, \sigma_n), i) = \sigma_i$.

1.2.3. Let V be a finite set, called the set of vertices and let v_0, v_e be two distinguished vertices of V. Take $V' = V - \{v_e\}$. A *V-Maze* is a function d: $V' \times \{0, 1\} \to V$ (see [1], [2], [11]). A V-maze can be considered as a directed graph $\Gamma(d)$, V being the set of vertices and $E = \{(v, d(v, i)) \mid v \in V', i \in \{0, 1\}\}$ the set of edges. This graph has two distinguished nodes v_0 and v_e. The outdegree of all nodes different from v_e is two and v_e is a sink of this graph, its outdegree being zero. The V-maze d is called to be threadable, if there is a path in $\Gamma(d)$ connecting v_0 with v_e.

Let $X = \operatorname{Map}(V' \times \{0, 1\}, V)$ be the set of V-mazes. These are the objects of the following information system:

$A = V' \times \{0, 1\}$ is the set of attributes,

V is the set of descriptors and

r: $X \times A \to V$ is the function defined by $r(d, (v, i)) := d(v, i)$.

$\operatorname{MAZES}(V) := (X, A, V, r)$ is an information system which is the underlying information system of the classification problem $\operatorname{GAP}(V, v_0, v_e)$ the classifying function t of which is defined by

$$t(d) := \textit{if } d \textit{ is threadable then } 1 \textit{ else } 0.$$

Since all $\operatorname{GAP}(V, v_0, v_e)$ with $\#(V) = n$ are isomorphic we write $\operatorname{GAP}(n)$

instead of GAP(V, v_0, v_e). Without restriction of generality we can assume $V = \{1, 2, \ldots, n\}$ and $v_0 = 1$, $v_e = n$.

2. Questionnaires or classifying graphs

2.1. A procedure to classify via a given classifying function are the questionnaires introduced by C. Picard [9] (see also [2] and [4]). A questionnaire or a classifying graph over a given information system $S = (X, A)$ is a quintuple $F = (Q, Y, \alpha, \delta, q_0)$, where

Q is a finite set, the set of nodes;

Y is a finite set with $Y \cap A = \emptyset$ and α: $Q \to Y \cup A$ is a mapping; the nodes of act $F := \alpha^{-1}(A)$ are called questions and the nodes of term $F := \alpha^{-1}(Y)$ are the results;

$\delta = \{\delta_q | \ q \in \text{act } F\}$, δ_q: $\text{Im}\,\alpha(q) \to Q$ describes the strategy of posing questions;

q_0 is the initial node, i.e., that node, in which all enquiries get started.

X operates partially on Q by the following function:

$$\text{act } F \times X \to Q, \quad (q, x) \mapsto qx := \delta_q((\alpha(q))(x)).$$

This action can be interpreted as follows: let q be a question; then $\alpha(q)$ is an attribute, i.e., a mapping $\alpha(q)$: $X \to V$. To pose question q to the object $x \in X$ means to apply $\alpha(q)$ on x. The answer of x to the question q is $\alpha(q)(x)$. This answer implies a new question or a result, namely $\delta_q(\alpha(q)(x))$ which we called qx. The partial action of X on Q can be extended to X^*, the free monoid generated by X. For $x \in X$ take $qx^{n+1} := (qx^n)x$ if $qx^n \in \text{act } F$. Let ξ_F: $X \to Y$ be the following function

$$\xi_F(x) := \textit{if } q_0 x^n \in \text{term } F \textit{ then } \alpha(q_0 x^n) \textit{ else not defined}.$$

The sequence $q_0, q_0 x, q_0 x^2, \ldots, q_0 x^n \in \text{term } F$ describes the strategy of F in asking questions: after having asked for the attribute $\alpha(q_0 x^m)$, $m < n$, F gets the answer $(\alpha(q_0 x^m))(x)$ which makes F move to the node $q_0 x^{m+1}$ if $m+1 < n$ or the result of the enquiry $\xi_F(x) = \alpha(q_0 x^n)$.

If F is free of cycles (more precisely if the directed graph

$$\Gamma(F) = \left(Q, \{(g, \delta_q(i)) | \ q \in \text{act } Q, \ i \in \text{Im}\,\alpha(q)\}\right)$$

is free of cycles), then $qx^n \neq q$ for all $q \in \text{act } F$, $x \in X$. In this case ξ_F is fully defined on X. So if we assume that F is free of cycles, then ξ_F is a fully defined function, which can be proved to be always (i.e., also in case when we do not assume that S is fully faithful) dependent on S. We say that F is a solution of a classification problem $C = (X, A, f)$ if $\xi_F = f$.

2.2. Easy to verify that every classification problem admits a solution F which is, moreover, a tree (for the easy proof of this fact we refer the reader

to [2]). In [2] and [4] we introduced different measures for the complexity of classifying graphs. One of these measure was the size of F:

$$\text{size}(F) = \#(Q).$$

Let $C = (X, A, f)$ be any classification problem. We introduce two numbers:

$$\text{size}(C) = \min \{\text{size}(F) |\ \xi_F = f\}, \quad \text{and}$$

$$\text{Size}(C) = \min \{\text{size}(F) |\ \xi_F = f \text{ and } \Gamma(F) \text{ is a tree}\}.$$

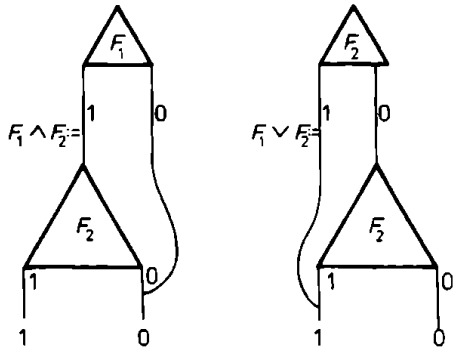
One of the most interesting and outstanding problems in theoretical computer science is the determination of the "small size" size(C) of certain classification problems C. Though also the determination of Size(C) is not easy, it can be done in certain cases. The following section presents some results concerning these questions. The detailed proofs of these results can be read in [2].

3. Classifying graphs for GAP(n) and related problems

3.1. Assume $Y = \{0, 1\}$ and let F_1, F_2 be classifying graphs with

$$F_i = (Q_i, Y, \alpha_i, \delta_i, q_{i0}), \quad i = 1, 2.$$

We define $F_1 \wedge F_2$ and $F_1 \vee F_2$ in the following straightforward manner:



Obviously

$$\text{size}(F_1 \wedge F_2) = \text{size}(F_1 \vee F_2) = \text{size}(F_1) + \text{size}(F_2) - 2$$

and

$$\xi_{F_1 \vee F_2} = \xi_{F_1} \vee \xi_{F_2}, \quad \xi_{F_1 \wedge F_2} = \xi_{F_1} \wedge \xi_{F_2}.$$

3.2. Consider the following classification problems:

In 1.2.3 we introduced already the information system MAZES(V) and the classification problem GAP(V, v_0, v_e). The latter classification problem can be considered as a special case of the following classification problem.

Call a V-maze d k-threadable for a given natural number k, if it is threadable and if the path connecting v_0 with v_e has a length smaller or equal to k. Let $t_k = t_k(V, v_0, v_e)$ be the characteristic function of the set of all k-threadable mazes, i.e.,

$$t_k(d) = \text{if } d \text{ is } k\text{-threadable then 1 else 0.}$$

Let $\mathrm{GAP}_k(V, v_0, v_e)$ be the corresponding classification problem. Obviously $\mathrm{GAP}(V, v_0, v_e) = \mathrm{GAP}_k(V, v_0, v_e)$ with $k = \#(V)-1$. It is easy to verify that one has the following equality:

$$t_{k+l} = V\{t_k(V, v_0, v_e) \wedge t(V, v, v_e) \mid v \in V,\ v \neq v_0,\ v \neq v_e\} \vee t_1(V, v_0, v_e).$$

From this equation one gets the following recursion formula for $s(k, n) := \mathrm{size}(\mathrm{GAP}_k(V, v_0, v_e))$ with $n := \#(V)$:

$$\begin{aligned} s(k+l, n) &\leqslant \sum_{i=2}^{n-1} \big(s(k, n)+s(l, n)-2\big)-2(n-3)+4-2 \\ &= (n-2)\big(s(k, n)+s(l, n)-4\big)+4. \end{aligned}$$

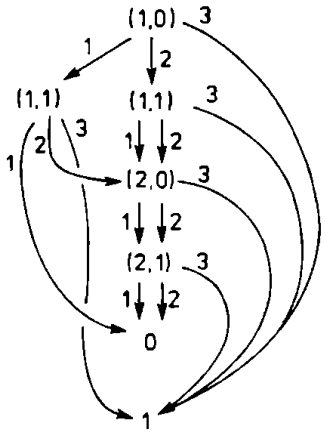
From this formula results:

$$s(2k, n) \leqslant 2(n-2)\big(s(k, n)-2\big)+4 \leqslant 2(n-1)s(k, n)$$

and this gives the following upper bound for $s(n) = \mathrm{size}(\mathrm{GAP}(n))$.

$$s(n) \leqslant 4n(n-1)^{\log n}.$$

3.3. Let us give an example: An optimal solution for the classification problem GAP(3), i.e., a classifying graph for GAP(3) with the minimal number $s(3) = 7$ of nodes is the following classifying graph:



3.4. The above example is the easiest case of the following solution F for $\mathrm{GAP}(n)$ which is in general not optimal but better than 3.2, for n small:

$$\mathrm{act}\, F = \{(U, n) \in 2^V \times N \mid 1 \in U,\ n \in U,\ \#(U)-1 \leqslant n < 2\#(U)\},$$

$\operatorname{term} F = \{0, 1\} = Y$,

$\alpha(U, n) :=$ n-th element of $U \times 2$ in lexicographical order,

$$\delta_{(U,n)} = \begin{cases} (U \cup \{y\}, n+1) & \text{if} \quad n+1 < 2 \# (U \cup \{y\}) \text{ and } y \neq v_e, \\ 1 & \text{if} \quad y = v_e, \\ 0 & \text{if} \quad n+1 = 2 \# (U \cup \{y\}), \end{cases}$$

$$q_0 = (\{v_0\}, 0).$$

It is easy to see, that

$$\operatorname{size} F = 2 + \sum_{i=0}^{n-2} \binom{n-1}{i} (i+2).$$

For $n = 3$ we get $\operatorname{size} F = 7$, and for $n = 4$, $\operatorname{size} F = 14$. The first value is optimal and we believe that also $s_4 = 14$. But already the proof of this fact seems to be hard. The importance of the numbers s_n is demonstrated by the following theorem:

3.5. THEOREM. *Let* $\mathbf{L}$ *be the class of all languages, which can be recognized by a Turing machine with logarithmic tape and let* $\mathbf{NL}$ *be the class of all languages which can be recognized by a nondeterministic Turing machine in logarithmic tape. Obviously* $\mathbf{L} \subseteq \mathbf{NL}$. *In order that* $\mathbf{L} = \mathbf{NL}$ *it is necessary that* s_n *is polynomial in* n.

The proof of this theorem can be found in [2].

4. Classifying trees

4.1. Let $S = (X, A)$ be an (fully faithful) information system. Let $\operatorname{trees}(S)$ be the set of all classifying trees over the given information system S. To every attribute $a \in A$ with $\operatorname{Im} \alpha = \{y_1, \ldots, y_n\}$ there is an n-ary function a: $\operatorname{trees}(S)^n \to \operatorname{trees}(S)$ which is defined in the following way: let $F_1, \ldots, F_n$ be elements of $\operatorname{tree}(S)$ with

$$F_i = (Q_i, Y_{i,i}, \delta_i, q_{i0});$$

then $F := a(F_1, \ldots, F_n) = (Q, Y, \alpha, \delta, q_0)$ is defined as follows:

$$Q = \{a\} \cup \Big(\bigcup_{i=1}^{n} Q_i \times \{i\}\Big), \qquad Y = \bigcup_{i=1}^{n} Y_i, \qquad \alpha(a) := a, \qquad \alpha(q, i) := \alpha_i(q).$$

Then we get $\operatorname{act}(F) = \{a\} \cup \Big(\bigcup_{i=1}^{n} \operatorname{act}(F_i) \times \{i\}\Big)$. The function δ is defined by $\delta_{(q,i)}(y) := (\delta_q(y), i)$ and $\delta_{a(i)} := (q_{i0}, i)$. The initial node q_0 of F is defined by $q_0 := a$, which completes the definition of F. Suppose we are given a fixed set Y. To every element $y \in Y$ we define the following trivial classifying tree:

$$[y] := (\{y\}, \{y\}, 1_{\{y\}}, \emptyset, \{y\}),$$

consisting only of one node. For the next theorem we consider by set theoretic reasons only classifying trees $F = (Q, Y, \alpha, \delta, q_0)$ with a fixed set Y of possible results of the classification.

4.2. THEOREM. *(trees(S), $A)$ is a free algebra and the set of all $[y]$ forms a set of free generators.*

The proof of this theorem is given in [2].

4.3. Consider the following binary relation $\triangleq$ on the set of all classifying trees.

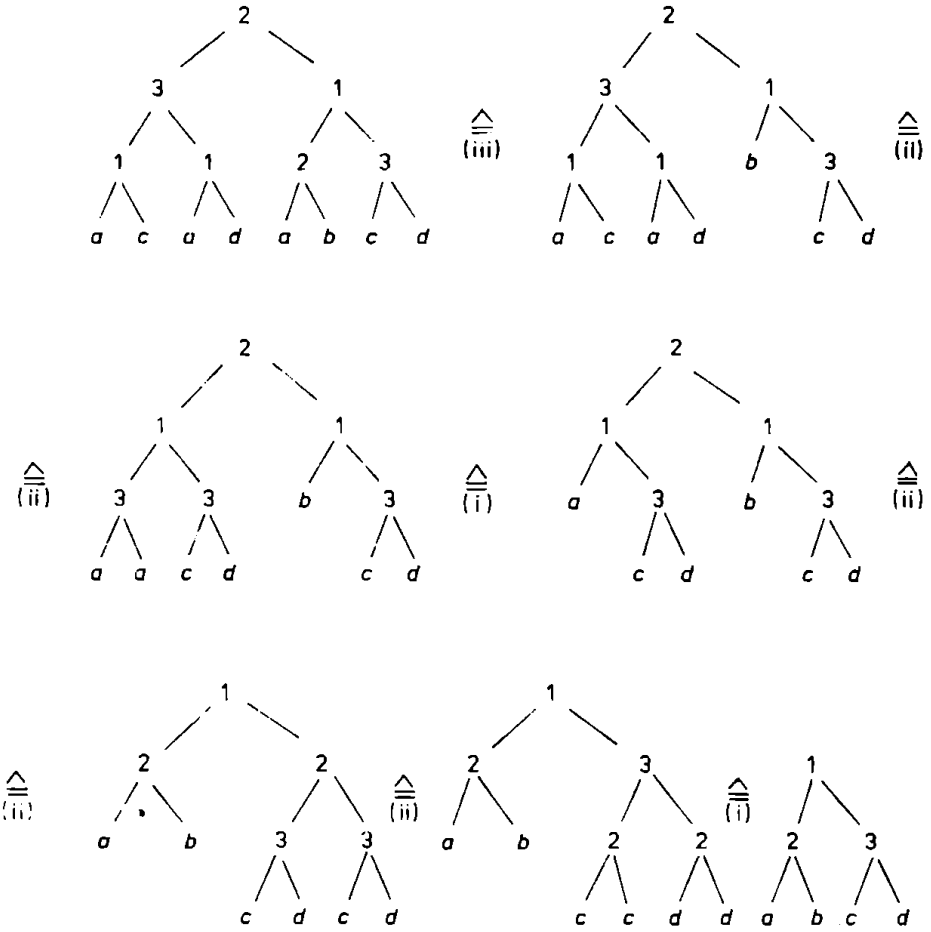
(i) *for $a \in A$, $y \in Y$ holds $a([y], \ldots, [y]) \triangleq [y]$,*

(ii) *if a, $b \in A$ then $a(b(F_{11}, \ldots, F_{1n}), \ldots, b(F_{m1}, \ldots, F_{mn})) \triangleq b(a(F_{11}, \ldots, F_{m1}), \ldots, a(F_{1n}, \ldots, F_{mn}))$;*

(iii) *Let $F = \mu(F_1, \ldots, F_n)$ and assume F_i contains a* subtree $F' = \mu(F'_1, \ldots, F'_n)$. *Let F'' be the tree obtained from F by replacing F' by F'_i. Then $F \triangleq F''$.*

4.4. DEFINITION. The congruence relation $\equiv$ generated by $\triangleq$ will be called the *syntactic congruence* of classifying trees.

Let us consider an example:



4.5. Two classifying trees F_1 and F_2 will be called *semantical equivalent*, $F_1 \sim F_2$, of their classifying functions are identical: $\xi_{F_1} = \xi_{F_2}$. It is obvious that classifying trees which are syntactical equivalent are semantical equivalent too. More interesting is the other direction which will be the main result of the following theorem, the proof of which will be found also in [2].

4.6. THEOREM. *Syntactical and semantical equivalence are equal, i.e., for all classifying trees holds: $F_1 \sim F_2$ if and only if $F_1 \equiv F_2$.*

WARNING. For this theorem the assumption that S is fully faithful is of significant importance. See example 4.13 of [2].

5. Optimal trees for GAP(n)

5.1. In 2.2 we introduced the notion of the "big size" of a classification problem $C = (X, A, f)$:

$$\mathrm{Size}(C) := \min \{\mathrm{size}(F) \mid \xi_F = f \text{ and } \Gamma(F) \text{ is a tree}\}.$$

A classifying tree F is called an *optimal tree solution of* C or for short an *optimal tree for* C if F is a solution of C and if, moreover, $\mathrm{size}(F) = \mathrm{Size}(C)$. Since every classification problem has a solution which is a tree, every classification has an optimal tree solution. The example at the end of 4.4 gives evidence that it might not be easy to find an optimal tree for C and, moreover, it may be rather difficult to decide whether a given classifying tree is an optimal solution. We will find in 7.7 a criterion which allows to answer the second question in certain cases.

Let us first consider the problem GAP(n).

Let $\sigma(n) := \mathrm{Size}(\mathrm{GAP}(n))$. We intend now to give a recursion which allows to compute these numbers and to give lower bounds for $\sigma(n)$.

5.2. Let V be a finite set with two distinguished elements v_0 and v_e. In 1.2.3 we introduced V-mazes as functions $d: V' \times \{0, 1\} \to V$ with $V' = V \setminus \{v_e\}$. A partial V-maze is a partial function $d: V' \times \{0, 1\} \xrightarrow{\supseteq} V$. As for V-mazes we can consider the directed graph $\Gamma(d)$ with V being the set of vertices of $\Gamma(d)$ and E, the set of edges being defined by

$$E := \{(v, d(v, i)) \mid (v, i) \in \mathrm{dom}\, d\}.$$

Let d be a partial V-maze. Let reach d be the set of all $v \in V$ which are connected with v_0 by a path from v_0 to v. We call d a *trunk* if $\mathrm{dom}\, d \subseteq (\mathrm{reach}\, d) \times \{0, 1\}$, and d is called a *complete trunk* if in this inclusion equality holds. We call d to be *connected* if $v_e \in \mathrm{reach}\ d$. We call d *disconnected* if it is not connected, and *stably disconnected* if all extensions $d' \supseteq d$ of d are disconnected.

Let d be any partial maze. Define $\tilde{d}$ by

$$\tilde{d} := d\downarrow((\operatorname{reach} d) \times \{0, 1\}) \cap \operatorname{dom} d.$$

Obviously $\tilde{d}$ is a trunk and it is easy to see that d is stably disconnected if and only if $\tilde{d}$ is a complete disconnected trunk.

Let F be a tree solution of $\mathrm{GAP}(V, v_0, v_e)$ and let q be any node of F. Let

$$q_0 \xrightarrow{x_0} q_1 \xrightarrow{x_1} q_2 \xrightarrow{x_2} \ldots \to q_n = q$$

be the path in F connecting the initial node q_0 of F with q. Since $\alpha(q_t) \in A = V' \times \{0, 1\}$, we have $\alpha(q_t) = (v_t, i_t)$. Consider the set $\{((v_t, i_t), x_t) \mid t = 0, 1, \ldots, n-1\}$ which we call d_q.

5.3. THEOREM. *If F is an optimal tree for $\mathrm{GAP}(V, v_0, v_e)$, then d_q is the graph of a partial function d_q: $V' \times \{0, 1\} \xrightarrow{\supset} V$, i.e., d_q is a partial maze. Moreover:*

(i) *d_q is a trunk for all nodes q of F;*

(ii) *for all $q \in \operatorname{act} F$, d_q is not complete nor connected;*

(iii) *for all $q \in \operatorname{term} F$ with $\alpha(q) = 1$, d_q is connected;*

(iv) *for all $q \in \operatorname{term} F$ with $\alpha(q) = 0$, d_q is complete and disconnected, i.e., stably disconnected.*

5.4. Every partial maze d defines a point in the grid N^2 by

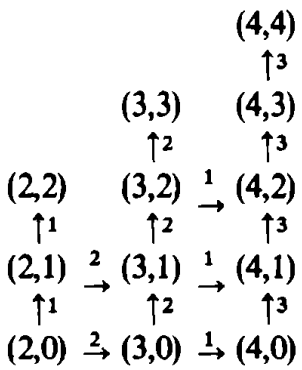
$$\eta(d) := (\#(\operatorname{reach} d) + 1,\ \#(\operatorname{dom} \tilde{d}) - \#(\operatorname{reach} d) + 1).$$

Every node q of as optimal tree for $\mathrm{GAP}(V, v_0, v_e)$ defines a point $\eta(q) := \eta(d_q)$ of N^2.

Consider the following graph $A^n = (V^n, E^n)$:

$$V^n = \{(x, y) \in N^2 \mid 0 \leqslant y \leqslant x,\ 2 \leqslant x \leqslant n\} \cup \{1, 2, \ldots, 2(n-1)\},$$
$$E^n = \{((x, y), i, (x+1, y)) \mid i \in \{1, 2, \ldots, n-x\}\} \cup$$
$$\cup \{((x, y), i, (x, y+1)) \mid i \in \{1, 2, \ldots, x-1\},\ y < x\} \cup$$
$$\cup \{((x, y), 1, x+y-1) \mid y < x\}.$$

Take A^4 as an example where we have omitted in the pictorial representation the edges of the third kind between (x, y) and $x+y-1$:



5.5. THEOREM. *If F is an optimal solution for* GAP(n), *then F is a covering tree of the graph* $\mathbf{1}^n$. *This implies that* $\sigma(n) = \mathrm{Size}(F)$ *is equal to the number of simple pathes in* $\mathbf{1}^n$.

5.6. COROLLARY. $\sigma(n) = \Omega(n^n(n-2)!)$.

For details we refer the reader to [2].

6. Coloured posets

6.1. A finite poset is said to be *pure* if all maximal chains have the same length. A pure poset satisfies the Jordan–Dedekind condition: if x and y are two elements and if $x < y$, then $I_x = \{z|\ z \leqslant x\}$, $V_x = \{z|\ x \leqslant z\}$ and $[x, y] = \{z|\ x \leqslant z \leqslant y\}$ are pure.

The symbol "$\prec$" denotes the covering relation: $x \prec y$ if $x < y$ and if $x < z \leqslant y$ implies $z = y$. If P is a finite pure poset, then a rank function r: $P \to \boldsymbol{N}$ can be defined as follows:

(i) If P has a least element 0, then we define $r(0) = 0$, otherwise we define $r(x) = 1$ for all minimal elements x.

(ii) If $x \prec y$, then $r(y) = r(x) + 1$.

6.2. A finite simplicial complex K is by definition a nonempty family of nonempty subsets called *simplexes* of a set $\{v\}$ of vertices such that

(i) any set consisting of exactly one vertex is a simplex;

(ii) any nonempty subset of a simplex is a simplex.

(For details we refer to [12].) The dimension of a simplex s, $\dim s$, is $\# s - 1$. The dimension of K, $\dim K$, is $\max\{\dim s|\ s \in K\}$. The maximal simplexes, i.e., those simplexes which are maximal under inclusion, are called facets. K is said to be *homogeneously n-dimensional* if every simplex belongs to an n-simplex of K. So all facets are n-dimensional.

Every finite simplicial complex K defines a finite poset $(K, \subseteq)$ the elements of which are the simplexes of K and these are ordered by inclusion. If K is homogeneously n-dimensional, then the corresponding poset is pure. Its rank function ϱ satisfies obviously the following condition: $\varrho(s) = \dim s + 1$.

Let P be an arbitrary ordered set. It defines a simplicial complex $\Delta(P)$ in the following way: The vertices of $\Delta(P)$ are the elements of P and the simplexes of $\Delta(P)$ are nonempty subsets $\{x_0, x_1, \ldots, x_k\}$ of P such that $x_0 < x_1 < \ldots < x_k$. If K is a simplicial complex, then $K' = \Delta(K)$ (K to be considered as poset) is the barycentric subdivision of K.

6.3. EXAMPLE. Let $S = (X, A, V, \varrho)$ be an information system with $N := \# A - 1$. We assume as usual that S is fully faithful. An S-condition is

defined to be a partial function c: $A \xrightarrow{\supset} V$ satisfying $c(a) \in \operatorname{Im} a$ for all $a \in \operatorname{dom} c$. As usual c can be considered as a subset of the product $A \times V$,

$$c = \{(a, v) |\ a \in \operatorname{dom} c,\ v = c(a)\}.$$

Consider the following simplicial complex:

(i) The set of vertices is $A \times V$.

(ii) The set $\operatorname{Cond}(S)$ of simplexes is the set of all S-conditions (considered as subsets of $A \times V$).

The facets of $\operatorname{Cond}(S)$ are the fully defined functions c: $A \to V$. They all are of dimension N. Hence $\operatorname{Cond}(S)$ is a homogeneously N-dimensional simplicial complex.

6.4. Let P be a finite pure poset. A partial function g: $P \xrightarrow{\supset} Y$ is called a *precolouring of* P and (P, g, Y) is called a *precoloured poset* if

(i) all maximal elements of P belong to the domain of g,

(ii) if $x < y$ and $x, y \in \operatorname{dom} g$, then $g(x) = g(y)$.

If x is an element of $\operatorname{dom} g$, then we say that x is *coloured* and $g(x)$ is the *colour of* x.

Let (P_i, g_i, Y_i) $(i = 1, 2)$ be two precoloured posets. A morphism of (P_1, g_1, Y_1) into (P_2, g_2, Y_2) is a pair (π, η) consisting of an order preserving map π: $P_1 \to P_2$ and a function η: $Y_1 \to Y_2$ such that the diagram

$$\begin{array}{ccc} P_1 & \xrightarrow{g_1} & Y_1 \\ {\scriptstyle\pi}\downarrow & & \downarrow{\scriptstyle\eta} \\ P_2 & \xrightarrow[g_2]{} & Y_2 \end{array}$$

is commutative. This means more precisely: Whenever $x \in P_1$ is coloured, $\pi(x)$ is coloured and $g_2 \pi(x) = \eta g_1(x)$.

6.5. Let (P, g, Y) be a precoloured poset. g is a colouring of P and (P, g, Y) is called a *coloured poset* if in addition to properties (i) and (ii) in (6.4) the following property holds:

(iii) If $x \in P$ and if all z, covering x, are coloured and have the same colour $g(z) = y_0 \in Y$, then x is coloured and $g(x) = y_0$.

Let $\max P$ be the set of all maximal elements of P. Condition (iii) is equivalent to either one of the following conditions:

(iii') If $x \in P$ and if all z with $x < z$ are coloured and have the same colour $g(z) = y_0 \in Y$, then x is coloured and $g(x) = y_0$.

(iii'') If $x \in P$ and if all $z \in V_x \cap \max P = \{y \in \max P |\ x \leqslant y\}$ have the same colour $g(z) = y_0 \in Y$, then x is coloured and $g(x) = y_0$.

Obviously every precolouring g can be extended in a unique way to a colouring $\bar{g}$ by the following procedure: $x \in \operatorname{dom} \bar{g}$ iff $\# g(V_x \cap \max P) = 1$. In this case there is a $y_0 \in Y$ with $g(V_x \cap \max P) = \{y_0\}$. Define $\bar{g}(x) := y_0$.

6.6. Every coloured poset (P, g, Y) can be divided into disjoint parts,

$$P = \bigcup_{y \in Y} g^{-1}(y) \cup C\mathrm{dom}\, g,$$

where $C\mathrm{dom}\, g := P \setminus \mathrm{dom}\, g$ is the complement of $\mathrm{dom}\, g$ in P. Obviously all $g^{-1}(y)$ are ascending (open) subsets of P and $C\mathrm{dom}\, g$ is a descending (closed) subset of P.

We define

$$\mathrm{Pure}(g, y) := g^{-1}(y),$$

$$\mathrm{Mix}(g) := C\mathrm{dom}\, g,$$

$$\mathrm{Pure}(g) := \bigcup_{y \in Y} \mathrm{Pure}(g, y) = \mathrm{dom}\, g.$$

6.7. EXAMPLE. Let $K = (S, Y, f)$ be a classification problem and let $S = (X, A, V, \varrho)$ be the underlying information system. As we have already seen in 6.3, S defines a poset $\mathrm{Cond}(S)$. The maximal elements of $\mathrm{Cond}(S)$ are the facets of $\mathrm{Cond}(S)$ which in turn are the elements of $\bigtimes_{a \in A} \mathrm{Im}\, a$. Therefore we can define the function

$$g\colon \max \mathrm{Cond}(S) = \bigtimes_{a \in A} \mathrm{Im}\, a \xrightarrow{\varrho^{-1}} X \xrightarrow{f} Y,$$

i.e., a precolouring of $\mathrm{Cond}(S)$ which defines in turn by 6.5 a colouring $\bar{g}$ which we denote by $\mathring{f}$.

For every $y \in Y$ we define $\mathrm{Pure}(K, y) := \mathrm{Pure}(f, y) := \mathrm{Pure}(\mathring{f}, y)$ and $\mathrm{Pure}(K) := \mathrm{Pure}(f) := \mathrm{Pure}(\mathring{f})$ which are partially ordered sets and define therefore via Δ complexes which are subcomplexes of $\mathrm{Cond}'(S)$, the barycentric subdivision of $\mathrm{Cond}(S)$.

6.8. Let $S_i = (X_i, A_i, V_i, \varrho_i)$ $(i = 1, 2)$ be two information systems. A triple $\varkappa\colon X_1 \to X_2$, $\alpha\colon A_2 \to A_1$, $\nu\colon V_1 \to V_2$ is called a *homomorphism* $\sigma = (\varkappa, \alpha, \nu)\colon S_1 \to S_2$ if for all $x \in X_1$, $a \in A_2$ the following equality holds:

$$\varrho_2(\varkappa(x), a) = \nu\varrho_1(x, \alpha(a))$$

or if we consider attributes as functions:

$$a\varkappa(x) = \nu\alpha(a)(x).$$

$\mathrm{Mix}(K)$, which is defined by $\mathrm{Mix}(K) := \mathrm{Mix}(f) := \mathrm{Mix}(\mathring{f})$ is a simplicial complex, subcomplex of $\mathrm{Cond}(S)$.

The notions of Pure and Mix are motivated by the following: Let c be a condition and x an object of X. We say x *satisfies* c if for all $a \in \mathrm{dom}\, c$ holds $a(x) = c(x)$. Let $\mathrm{Sat}(c)$ be the set of all objects which satisfy c. Then $c \in \mathrm{Pure}(f)$ iff $\# f(\mathrm{Sat}(c)) = 1$, i.e., all x satisfying c are of "the same colour". Otherwise: $c \in \mathrm{Mix}(f)$ iff there are objects x and y satisfying c with $f(x) \neq f(y)$.

Note that neither Pure nor Mix are in general functorial, in many important cases, however, they are:

6.9. PROPOSITION. *Let* $\lambda = (\varkappa, \alpha, \nu, \eta)$: $K_1 \to K_2$ *be a homomorphism of classification problems* $K_i = (S_i, Y_i, f_i)$, $S_i = (X_i, A_i, V_i, \varrho_i)$ *the underlying information systems. Suppose one of the following conditions to be satisfied:*

(i) α *is injective and* $\nu(\mathrm{Im}\,\alpha(a)) = \mathrm{Im}\,a$ *for all* $a \in A_2$.

(ii) $\varkappa$ *is surjective.*

(iii) $K_2 = K_1 \times K_1$ *and* λ *is the diagonal map.*

The mapping λ_* *which maps the* S_1*-condition* c *onto* $\nu \subset \alpha$ *is a simplicial map of* $\mathrm{Mix}(K_1)$ *into* $\mathrm{Mix}(K_2)$. *Moreover,* λ_* *is an order preserving map of* $\mathrm{Pure}(K_1)$ *into* $\mathrm{Pure}(K_2)$ *which in turn defines a simplicial map of the corresponding simplicial complexes* $\Delta(\mathrm{Pure}(K_i))$.

7. Topology of $\mathrm{Cond}(S)$, $\mathrm{Pure}(K)$, and $\mathrm{Mix}(K)$

7.1. Let K be a simplicial complex. The geometric realization $|K|$ of K is by definition the set of all functions p defined over the set of vertices of K with values in the interval $[0, 1] \subseteq \boldsymbol{R}$ satisfying the following conditions:

(a) $\mathrm{supp}\,p = \{v|\ p(v) \neq 0\} \in K$;

(b) $\sum_v p(v) = 1$.

7.2. PROPOSITION. *Assume* S *to be completely fully faithful. Then for the* i*-th homology group of* $\mathrm{Cond}(S)$ *with coefficients in* $\boldsymbol{Z}$ *holds*

$$H_i(\mathrm{Cond}(S); \boldsymbol{Z}) \cong \begin{cases} \boldsymbol{Z} & \text{if } i = 0, \\ \boldsymbol{Z}^t & \text{if } i = N, \\ \{0\} & \text{otherwise,} \end{cases}$$

where $t = (m-1)^{N+1}$ *with* $m = \# V$, $N+1 = \# A$.

If $m = 2$ (case of boolean functions), *then* $|\mathrm{Cond}(S)|$ *is homeomorphic to the* N*-dimensional sphere.* (B. Graw [6] proved, moreover, that in general $\mathrm{Cond}(S)$ is shellable and $|\mathrm{Cond}(S)|$ is a bouquet of t N-dimensional spheres.)

7.3. PROPOSITION. *Let* K *be a classification problem with completely fully faithful underlying information system* S. *Then* $\mathrm{Mix}(K)$ *is a pure* $(N-1)$*-dimensional subcomplex of* $\mathrm{Cond}(S)$ *and* $|\Delta(\mathrm{Pure}(K))|$ *is homotopic to the complement of* $|\mathrm{Mix}(K)|$ *in* $|\mathrm{Cond}(S)|$.

7.4. PROPOSITION (Lefschetz Duality). *If* S *is completely fully faithful and* $m = \# V = 2$, *then* $\mathrm{Mix}(K)$ *and* $\mathrm{Pure}(K)$ *are connected by the following isomorphism of the homology groups:*

$$H_i(\mathrm{Pure}(K); \boldsymbol{Z}) = H_{N-i}(\mathrm{Cond}(S), \mathrm{Mix}(K); \boldsymbol{Z}).$$

This theorem allows to compute the homology groups $H_i(\mathrm{Pure}(K); Z)$ knowing $H_i(\mathrm{Mix}(K); Z)$, and vice versa. To do this use the exact homology sequence

$$\ldots \to H_{i+1}(\mathrm{Cond}(S), \mathrm{Mix}(K); Z) \to H_i(\mathrm{Mix}(K); Z) \to H_i(\mathrm{Cond}(S); Z) \to H_i(\mathrm{Cond}(S), \mathrm{Mix}(K); Z) \to \ldots$$

and take into account Proposition 3.2.

Let

$$h_i(\mathrm{Pure}(K)) := \operatorname{rank} H_i(\mathrm{Pure}(K); Z), \quad h_i(\mathrm{Mix}(K)) := \operatorname{rank} H_i(\mathrm{Mix}(K); Z)$$

be the Betti numbers of $\mathrm{Pure}(K)$, $\mathrm{Mix}(K)$ respectively. Then Proposition 7.4 yields the following:

7.5. COROLLARY. *Assume* $N \geqslant 2$. *Under the assumptions of* 7.4 *the following equalities hold:*

$$\begin{aligned} h_0(\mathrm{Pure}(K)) &= h_{N-1}(\mathrm{Mix}(K)) + 1, \\ h_i(\mathrm{Pure}(K)) &= h_{N-1-i}(\mathrm{Mix}(K)) \quad \text{if} \quad N-1 > i > 0, \\ h_{N-1}(\mathrm{Pure}(K)) &= h_0(\mathrm{Mix}(K)) - 1, \\ h_i(\mathrm{Pure}(K)) &= 0 \quad \text{if} \quad i > N-1. \end{aligned}$$

7.6. Let $\mathrm{size}(C)$ be the size of a smallest classifying tree for the classification problem C. The following result gives evidence that classification problems which are difficult from the topological point of view are intractable from the computational standpoint.

7.7. THEOREM.

$$\mathrm{size}(C) \geqslant \frac{h_0(\mathrm{Pure}(C)) - 1}{m-1}.$$

7.8. COROLLARY. *If* $m = 2$, *then*

$$\mathrm{size}(C) \geqslant h_0(\mathrm{Pure}(C)) - 1 = h_{N-1}(\mathrm{Mix}(C)),$$

i.e., classification problems with many "$(N-1)$*-dimensional holes" are of high complexity.*

The following theorem is of great importance for the study of connections between different classification problems.

7.9. THEOREM. *Let* $\lambda\colon C_1 \to C_2$ *be a homomorphism of classification problems satisfying one of conditions* (i), (ii) *or* (iii) *of* 2.8. *Let* D *be any sheaf* (cf. [5]) *of* R*-modules* (R *an arbitrary ring*) *over* $\mathrm{Cond}(C_1)$. *There are two spectral sequences:*

$$I_2^{pq} = H^p(\mathrm{Pure}(C_2), R^q \lambda_* D) \Rightarrow I^n = H^n(\mathrm{Pure}(C_1), D)$$

and

$$II_2^{pq} = H^p(\mathrm{Mix}(C_2),\ R^q\lambda_* D) \Rightarrow II^n = H^n(\mathrm{Mix}(C_1),\ D).$$

If c is a condition of C_1, then the fiber of $R^q\lambda_* D$ in c is given by

$$(R^q\lambda_* D)_c = H^q(\lambda_*/c,\ D) = H^q(\lambda_*^{-1}(V_c),\ D),$$

where $V_c = \{d \mid d$ a condition of $\mathrm{Pure}(C_2)$, $\mathrm{Mix}(C_2)$, respectively, and c a subcondition of $d\}$.

This theorem will be used in the next section to compute the Euler–Poincaré characteristic.

8. The Euler–Poincaré characteristic

8.1. Let K be a finite simplicial complex. The alternating sum

$$\chi(K) = \sum_{i=0}^{\infty} (-1)^i \,\#\,\{s \in K \mid \dim s = i\} = \sum_{s \in K} (-1)^{\dim s},$$

i.e., the number of simplexes of even dimension minus the number of simplexes of odd dimension is a topological invariant because $\chi(K) = \sum_{i=0}^{\infty} (-1)^i \operatorname{rank} H_i(K; \mathbf{Z})$. This invariant is called the *Euler–Poincaré characteristic*. For a poset P one defines $\chi(P) = \chi(\Delta(P))$. The Lefschetz duality (cf. 7.4) has the following straightforward consequence.

8.2. PROPOSITION. *Let C be a classification problem and $S = (X, A, V, \varrho)$ the underlying information system which is assumed to be completely fully faithful, and let $\# V = 2$, $\# A - 1 = N$. Then*

$$\chi(\mathrm{Mix}(C)) - 1 = (-1)^{N+1}\big(\chi(\mathrm{Pure}(C)) - 1\big).$$

8.3. PROPOSITION. *The Euler–Poincaré characteristic of $GAP(N)$ satisfies the following properties:*

(i) $\chi\big(\mathrm{Pure}(\mathrm{GAP}(N),\ 0)\big) = h_0\big(\mathrm{Pure}(\mathrm{GAP}(N),\ 0)\big) = h_0\big(\mathrm{Pure}(\mathrm{GAP}(N))\big) - 1 = \Omega\big((N-2)!\,(n-1)^N\big)$;

(ii) $\mathrm{Mix}(\mathrm{GAP}(N))$ *is shellable* (cf. [6]) *and therefore* $\chi\big(\mathrm{Mix}(\mathrm{GAP}(N))\big) = 1 + \operatorname{rank} H_{2(N-2)}\big(\mathrm{Mix}(\mathrm{GAP}(N));\ \mathbf{Z}\big)$.

(For $N = 3$ one gets $\chi\big(\mathrm{Pure}(\mathrm{GAP}(3))\big) = 13$.)

The next theorem studies the relationship of the Euler–Poincaré characteristic along 'nice' homomorphisms.

8.4. THEOREM. *Let $\lambda\colon C_1 \to C_2$ be a homomorphism of classification problems satisfying one of conditions* (i), (ii) *or* (iii) *of* 6.9. *Let c be an S_2-condition and define λ_*/c, $\lambda_* \backslash c$ in the following way:*

$$\lambda_*/c := \{d \in \mathrm{Mix}(C_1) |\ \lambda_*(d) \supseteq c\} \quad (\textit{defined if } c \in \mathrm{Mix}(C_2)),$$

$$\lambda_* \setminus c := \{d \in \mathrm{Pure}(C_1) |\ \lambda_*(d) \subseteq c\} \quad (\textit{defined if } c \in \mathrm{Pure}(C_2)).$$

λ_*/c and $\lambda_* \setminus c$ are called the fibers of c along λ. Then

$$\chi(\mathrm{Mix}(C_1)) = \sum_{c \in \mathrm{Mix}(C_2)} (-1)^{\dim c} \chi(\lambda_*/c),$$

$$\chi(\mathrm{Pure}(C_1)) = \sum_{c \in \mathrm{Pure}(C_2)} (-1)^{\mathrm{codim}\, c} \chi(\lambda_* \setminus c)(m-1)^{\mathrm{codim}\, c}.$$

This theorem allows to compute the Euler–Poincaré characteristic of Mix and Pure if one finds 'nice' homomorphisms onto easier classification problems, the fiber of which is also computable.

8.5. The next theorem studies the behaviour of classification problems if one puts one additional question: Let $C = (S, Y, f)$ be a classification problem, $y \in Y$, and let a be an attribute of the underlying information system. If one extends f by adding question a in case of y one gets the following classification $C \dotplus_y a = (S, Y \perp\!\!\!\perp \mathrm{Im}\, a, f \dotplus_y a)$, where range $(f \dotplus_y a) = Y \perp\!\!\!\perp \mathrm{Im}\, a$, and for an object x one defines

$$(f \dotplus_y a)(x) := \begin{cases} f(x) & \text{if} \quad f(x) \neq y, \\ a(x) & \text{if} \quad f(x) = y. \end{cases}$$

Let U be an ascending subset of $\mathrm{Cond}(S)$. Then we define $U \setminus a := \{c \in U |\ a \notin \mathrm{dom}\, c\}$. $U \setminus a$ is obviously a descending subset of U (not of $\mathrm{Cond}(S)$ in general). One gets the following theorem:

8.6. THEOREM. *If $\chi(C) := \chi(\mathrm{Pure}(C))$, then*

$$\chi(C \dotplus_y a) = \chi(C) + \chi(\mathrm{Pure}(C, y) \setminus a).$$

8.7. It is possible to define a classification f relative to a classification g in such a way that

$$\chi(f) = \chi(g) + \chi(f|g)$$

if f classifies finer than g. Moreover, one gets for $C_1 \times C_2$ the formula

$$\chi(C_1 \times C_2) = \chi(C_1) \cdot \chi(C_2).$$

These formulas resemble the well-known formulas of the classical Shannon entropy so that cum grano salis the Euler–Poincaré characteristic can be considered as kind of structural information.

References

[1] L. Budach, *Two pebbles don't suffice*, In: *Mathematical Foundations of Computer Science* 1981; Proceedings, 10th Symposium Strebské Pleso, Czechoslovakia, 1981 (Eds.: J. Gruska, M. Chytil); Lecture Notes in Computer Science, 118; Berlin–Heidelberg–New York 1981; 578–589.

[2] –, *Klassifizierungsprobleme und das Verhältnis von deterministischer zu nichtdeterministischer Raumkomplexität*, Seminarbericht Nr. 68, Sektion Mathematik der Humboldt-Universität, 1985, 1–64.
[3] –, *Information und Rechnen*, In: *Zur Bedeutung der Information für Individuum und Gesellschaft*; Berichtsband der Wissenschaftlichen Konferenz zum Leibniz-Tag der Akademie der Wissenschaften der DDR, Berlin, 29–30.6.1983; 191–208.
[4] –, *A Lower Bound for the Number of Nodes in a Decision Tree*, EIK (to appear).
[5] R. Godement, *Topologie algébrique et théorie des faisceaux*, Hermann, Paris 1958.
[6] B. Graw, Personal communication, 1983.
[7] J. Kahn, M. Saks, *A topological approach to evasivenes*, Manuscript (1983), 1–37.
[8] W. Marek, Z. Pawlak, *Information storage and retrieval systems*, Mathematical Foundations. Theoret. Compt. Sci. 1 (1976), 331–354.
[9] C. F. Picard, *Theorie der Fragebogen*, Akademie-Verlag, Berlin 1973.
[10] P. Pudlàk, S. Žak, *Space complexity of computations*, Manuscript (1983), 1–30.
[11] W. Savitch, *Relations between nondeterministic and deterministic tape complexities*, Journal of Computer and System Science 4 (1970), 177–192.
[12] E. H. Spanier, *Algebraic Topology*, McGraw-Hill, 1966.
[13] R. P. Stanley, *Combinatorics and Commutative Algebra*, Birkhäuser, Boston Basel, Stuttgart 1983.

Presented to the semester
Mathematical Problems in Computation Theory
September 16–December 14, 1985