Aplikace matematiky

Josef Fučík

An agglomerative method for automatic forming of hierarchical classification

Aplikace matematiky, Vol. 20 (1975), No. 3, 186–205

Persistent URL: http://dml.cz/dmlcz/103583

AN AGLOMERATIVE METHOD FOR AUTOMATIC FORMING OF HIERARCHICAL CLASSIFICATION

JOSEF FUČÍK

(Received January 21, 1974)

INTRODUCTION

1. It is a frequent problem that one needs to investigate a set of elements, each of them being described by a group of properties. It is always necessary to get a general idea of the set, to find the relations between the elements which are the consequence of the qualitative or quantitative connections of their properties. A wide scale of methods has been developed in mathematics to solve such problems. One of them, part of which is also the topics of this paper, is to classify the set. The aim is to divide the set under investigation in to disjoint nonempty subsets (to form the classification), in order to be able to judge in the highest possible degree from the location of an element (object) in the classification of its relation to the other elements of the set and vice versa.

One of the basic concepts for the classification of a set is the similarity of the objects which is expressed mathematically by means of the similarity-function. We shall solve neither the problem how to introduce the function nor that of a suitable choice of the particular similarity-function (see e.g. [1]). For our purpose it is sufficient to suppose that we have already defined a similarity-function on the non-ordered pairs of elements of the set under classification.

If we obtain somehow a classification, we can use it as the basic set and classify it once more. If we repeat this process (called aglomeration) until we group in one class all objects of the original set, we form a sequence of classifications (hierarchical clasification) which offers a more perfect description of the set.

In [1] the so called lexicographic algorithm is described constructing a hierarchical classification by means of transitive closures of the components of quasiordering on the non-ordered pairs of the set under classification. In some cases this algorithm does not yield good results as the aglomeration groups in the same class any two objects from a pair belonging to the given component of the quasiordering. Thus the main features similarity of aglomerated classes are not grasped which results in the following drawbacks:

1) From the very beginning the components of the quasiordering may contain pairs with transitive closure of large cardinality while the algorithm aglomerates objects of very weak similarity. The sets A, B may happen to be aglomerated on the basis of great similarity of a single pair $\{x, y\}$ (where $x \in A$, $y \in B$), although other objects of A, B show only little similarity.
2) Hierarchical classification obtained by means of the lexicographic algorithm is uniguely determined by the quasiordering. Aglomerating, we do not take into account values of the similarity-function corresponding to the components of the quasiordering. That is to say, we neglect the distance of the components.

The algorithm for forming hierarchical classification which will be presented later avoids these defects in the following way:

1) Suitable transitive subsets are selected instead of the transitive closure of components and thus the objects that cause the undesirable properties of the hierarchical classification are omitted.
2) It works directly with the values of the similarity-function so that greater amount of information is taken into account.

2. Now the necessary notation and concepts will be introduced. First of all we need some basic concepts of the theory of sets.

Denoting by X, Y sets, then $\mathscr{P}_0(X)$ is the set of all nonempty subsets of X (power of X without the empty set), $\bigcup X$ is the set of all elements contained in some of the elements of X (union of X) and $|X|$ is the cardinality of X. $\emptyset$ is the symbol for the empty set, $\{x, y\}$ is the non-ordered pair and $\langle x, y\rangle$ is the ordered pair of the elements x and y. We shall use very often the set[1]) of non-ordered pairs of mutually different elements, one of them being an element of X and the other of Y, and we shall denote it by $[X, Y]$. Under $[X]^2$ we understand the set $[X, X]$. The concepts necessary for classification are denoted in accordance with Lerman [1]. E denotes the basic set to be classified; its elements are called objects. S is the symbol for the similarity-function with the domain $[E]^2$. Classification is usually denoted by P and can be introduced more precisely by

Definition 1. *Classification of E is the set of nonempty disjoint subsets (the so-called classes of classification) the union of which is E.*

In particular, denote by $P_{\min}$ the classification with one-element classes and by $P_{\max}$ the classification with one and only one class.

---

[1]) $[X, Y] = \{\{x, y\};\ x \in X \,\&\, y \in Y \,\&\, x \neq y\}$.

Let P_1, P_2 be classifications of E.

(1) P_1 is a refinement of P_2 if
$P_1 \neq P_2 \,\&\, (\forall N_1 \in P_1)(\exists N_2 \in P_2)(N_1 \subseteq N_2)$.

(2) P_1, P_2 are comparable if $P_1 = P_2$ or P_1 is a refinement of P_2 or P_2 is a refinement of P_1.

As we have indicated in the introduction, we shall deal with an aglomerative method for generating hierarchical classifications which is an improvement of the known procedure.

Definition 2. *Sequence $P_0 \dots P_n$ of classifications of E is hierarchical classification of E if $P_0 = P_{\min}$, $P_n = P_{\max}$ and P_k is a refinement of P_{k+1} for every positive integer k less then n.*

The aglomerative process starts from the classification $P_{\min}$ and formes the separate levels of hierarchical classfication by grouping together classes of the preceding classification. In order to be able to compare not only pairs of the elements of E but also the classes of classification of E, we extend the similarity-function to the set $[\mathscr{P}_0(E)]^2$.

Definition 3. *A function $\bar{S}$ with the domain $[\mathscr{P}_0(E)]^2$ is called the extension of the function S if it satisfies the following condictions*:

(1) *For $\{N, M\} \in [\mathscr{P}_0(E)]^2$ and $N \cap M = \emptyset$ it is*

$$\min_{\{x,y\}\in[N,M]} \bar{S}(\{x\},\{y\}) \leqq \bar{S}(N, M) \leqq \max_{\{x,y\}\in[N,M]} \bar{S}(\{x\},\{y\}).$$

(2) *For $\{x, y\} \in [E]^2$ it is $\bar{S}(\{x\},\{y\}) = S(x, y)$.*

Examples: For arbitrary disjoint sets $N, M \in \mathscr{P}_0(E)$ we define

a) extension of the function *by the mean value*

$$\bar{S}(N, M) = \frac{1}{|N| \,.\, |M|} \sum_{\{x,y\}\in[N,M]} S(x, y)$$

b) extension of the function *by the minimum*

$$\bar{S}(N, M) = \min_{\{x,y\}\in[N,M]} S(x, y)$$

In concrete cases it is convenient to choose the extension of the similarity-function (in the same manner as the similarity-function) according to the criteria ensuing from the purpose of the classification. The extension by mean value seems to be the best and that is why we present an algorithm for this extension. The algorithm can be probably modified also for other type of extension. From now throughout this paper we denote by $\bar{S}$ only the extension of the function S by mean value.

The formula $\{x, y\} \leqq \{z, v\} \leftrightarrow S(x, y) \geqq S(z, v)$ defines a similarity relation (quasiordering) on $[E]^2$, which is reflexive and transitive. The relation of equivalence $\sim$ can be defined in a natural manner:

$$\{x, y\} \sim \{z, v\} \leftrightarrow S(x, y) = S(z, v) .$$

The classes of the equivalence determined by the relation $\sim$ are called the components of the quasiordering $\leqq$. In particular, the set $\mathscr{K}(E) = \{\{x, y\}; \{x, y\} \in [E]^2 \; \& \; S(x, y) = \max_{\{z,v\} \in [E]^2} S(z, v)\}$ is called the first component of the quasiordering. The degree of similarity of objects in the classes of classification is aptly described by their similarity average.

Definition 4. *Let P be a classification of E and $P \neq P_{\min}$. Let R be the set $\{\{x, y\};$ $(\exists N \in P)(\{x, y\} \in [N]^2)\}$. The similarity-measure $\mathscr{V}(P)$ of the classification P (with respect to S) is defined by*

$$\mathscr{V}(P) = \frac{1}{|R|} \sum_{\{x,y\} \in R} S(x, y) .$$

3. By means of the similarity-measure and the extended similarity-function we shall try to express mathematically the intuitive good properties of hierarchical classifications.

1) First of all we group the pairs with the largest degree of similarity. That is, we look for a classifications possessing the similarity-measure as large as possible. However, it is desirable at the same time to group together as many objects as possible. That is why the most important classifications are those for which grouping together any further pair of classes should mean considerable decrease in the similarity-measure.

2) Not to form redundently fine hierarchical classifications we also require the value of the maximum similarity between the classes of the classification to decrease when passing to the next level.

Definition 5. *A hierarchical classification $P_0 \ldots P_n$ of E is called proper with respect to S if the following conditions are satisfied.*

(1) $(\forall P)(\forall k)(P_k$ *is a refinement of* $P \; \& \; 0 < k < n \rightarrow \mathscr{V}(P) < \mathscr{V}(P_k))$,

(2) $(\forall i)(\forall j)(0 \leqq i < j < n \rightarrow \max_{\{N,M\} \in [P_i]^2} \bar{S}(N, M) > \max_{\{K,L\} \in [P_j]^2} \bar{S}(K, L))$.

We shall show that the conditions (1) and (2) are independent.

Example 1. Condition (2) does not imply (1).

Values of the similarity-function on the set $E = \{x, y, z, u, v, w\}$ are given in table 1.

	y	z	u	v	w
x	4	2	2	2	2
y		2	2	2	2
z			1	1	3
u				1	3
v					3

Let us consider the hierarchical classification $P_{\min}, P_1, P_2, P_{\max}$ where

$$P_1 : \{x, y\} \{z, u, v\} \{w\} ,$$
$$P_2 : \{x, y\} \{z, u, v, w\} ,$$

$\mathscr{V}(P_1) = 1/4(S(x, y) + S(u, v) + S(u, z) + S(v, z)) = 7/4$. Every classification such that P_1 is its refinement has similarity-measure larger than $\mathscr{V}(P_1)$.

In particular, P_2 does not satisfy (1) even if (2) holds.

Indeed, $\mathscr{V}(P_2) = 16/7 > \mathscr{V}(P_1) = 7/4$

$$\max_{\{N,M\}\in[P_1]^2} \bar{S}(N, M) = \bar{S}(\{u, v, z\}, \{w\}) = 3 > 2 =$$
$$= \bar{S}(\{x, y\}, \{u, v, z, w\}) = \max_{\{N',M'\}\in[P_2]^2} \bar{S}(N', M') \quad \text{and} \quad \max_{\{N,M\}\in[P_{\min}]^2} \bar{S}(N, M) = 4 .$$

Example 2. Condition (1) does not imply (2).

Similarity-function on the set $E = \{x, y, z, u, v\}$ is given in table 2.

	y	z	u	v
x	3	2	1	1
y		2	1	1
z			1	1
u				2

Let us have the hierarchical classification $P_{\min}, P_1, P_2, P_{\max}$ where

$$P_1 : \{x, y\} \{z\} \{u\} \{v\}$$
$$P_2 : \{x, y\} \{z\} \{u, v\} .$$

Condition (1) holds: $\mathscr{V}(P_1) = 3 > \mathscr{V}(P_2) = 5/2 > \mathscr{V}(P_{\max}) = 3/2$.

Condition (2) does not hold: $\max_{\{N,M\}\in[P_1]^2} \bar{S}(N, M) = 2 = \max_{\{N',M'\}\in[P_2]^2} \bar{S}(N', M')$.

The concept of proper hierarchical classification defines the set of hierarchical classifications which give the best possible information about the objects under classification. This will be more apparent in the following when hierarchical classifications will be constructed independently of this definition and satisfying both conditions. Forming these hierarchical classifications or the individual levels in hierarchical classification we shall take into account above all two rather contradictory criteria, namely, slight decrease of the similarity-measure and, on the other hand, grouping together of great number of objects.

TRANSITIVE SETS

1. Let X be a set. We say that $\mathscr{C} \subseteq [X]^2$ is *transitive* iff the relation $U = \{\langle x, y\rangle; \{x, y\} \in \mathscr{C} \vee x = y\}$ is transitive.

Lemma 1. *Transitive sets on $[X]^2$ and classifications of X are in the following one-to-one correspondence:*

1) *If $\mathscr{C}$ is a transitive set on $[X]^2$, then we can define a classification $\mathscr{C}(X)$ of X in this way:*

$$(\forall x, y \in X)\,(x, y \textit{ are in the same class of the classification } \mathscr{C}(X) \leftrightarrow \{x, y\} \in \mathscr{C} \vee x = y)\,.$$

The classification $\mathscr{C}(X)$ is said to be induced by the set $\mathscr{C}$.

2) *To every classification of X there exists a transitive set $\mathscr{C} \subseteq [X]^2$ which induces it.*

Lemma 2. *Let $\mathscr{C}, \mathscr{D} \subseteq [X]^2$ be transitive sets. Then $\mathscr{C}(X)$ is a refinement of $\mathscr{D}(X)$ iff $\mathscr{C} \subset \mathscr{D}$.*

Let particularly X be a classification of E (denote it by P). Any classification P_1 of P defines a classification P_2 of E, where $P_2 = \{\bigcup N;\, N \in P_1\}$. Denote by $\mathscr{C}$ the transitive set on $[P]^2$ which induces the classification P_1 of P. Then P_2 is called the classification *biinduced* by the set $\mathscr{C}$ and in the following we shall denote it by $P_{\mathscr{C}}$. $P_{\mathscr{C}}$ can be defined directly from $\mathscr{C}$ as follows:
$(\forall\, x, y \in E)\,(x, y$ are in the same class of the classification

$$P_{\mathscr{C}} \leftrightarrow (\exists M, N \in P)\,(x \in M \;\&\; y \in N \;\&\; (\{M, N\} \in \mathscr{C} \vee M = N)))\,.$$

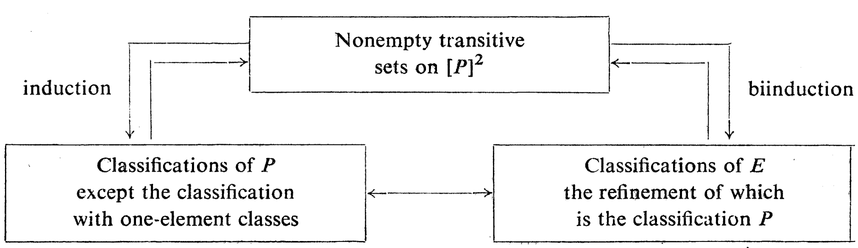


Fig. 1

Let us suppose now that we form a hierarchical classification by means of an aglomerative method. Denote by P the last level (classification) constructed. We obtain the next level by grouping together classes of the classification P. This means that we select one classification from the set of all classifications the refinement of which is P. This is equivalent to the choice of a transitive set from all nonempty transitive sets on $[P]^2$ (see Lemma 1). In order to make the best of this relation for the construction, it is necessary to investigate in more detail the relationship between transitive sets on $[P]^2$ and adjoin biinduced classifications.

2. We shall show that transitive subsets of the first component of the quasiordering defined by the function $\bar{S}$ are of the greatest importance for the aglomeration since the decrease in similarity-measure is smaller. In this way we obtain finer hierarchical classification and stronger similarity of objects in the same classes.

Proof for the first degree of the aglomeration is easy.

Theorem 1. *Let $\mathscr{K}(P_{\min})$ be the first component on $[P_{\min}]^2$, let $P_{\mathscr{C}}$ be the classification biinduced by a nonempty transitive set $\mathscr{C} \subseteq \mathscr{K}(P_{\min})$. If $P_{\mathscr{D}}$ is the classification biinduced by a nonempty transitive set $\mathscr{D} \subseteq [P_{\min}]^2$, then $\mathscr{V}(P_{\mathscr{C}}) \geqq \mathscr{V}(P_{\mathscr{D}})$ and if in addition $\mathscr{D} \nsubseteq \mathscr{K}(P_{\min})$, then $\mathscr{V}(P_{\mathscr{C}}) > \mathscr{V}(P_{\mathscr{D}})$.*

Proof. For each two objects x, y of the same class of the classification $P_{\mathscr{C}}$ (i.e. $\{x, y\} \in R$), $S(x, y) = r$ holds where $r = \max_{\{\{z\},\{v\}\} \in [P_{\min}]^2} S(z, v)$, since $\mathscr{C} \subseteq \mathscr{K}(P_{\min})$. From the definition of the similarity-measure it is $\mathscr{V}(\mathscr{P}_{\mathscr{C}}) = r$. If $\mathscr{D} \nsubseteq \mathscr{K}(P_{\min})$ then there exists such a pair $\{\{x\}, \{y\}\} \in \mathscr{D}$ that $S(x, y) < r$. Consequently, $\mathscr{V}(P_{\mathscr{C}}) > \mathscr{V}(P_{\mathscr{D}})$.

In general case when we aglomerate the classes of classification P which is the last level of the hierarchical classification constructed, it is necessary to assume in Theorem $\mathscr{V}(P) \geqq r$ where r is the value of the function $\bar{S}$ on the set $\mathscr{K}(P)$ ($r = \max_{\{N,M\} \in [P]^2} \bar{S}(N, M)$). This condition corresponds to a great extent to the fact that the classification P has been formed in accordance with the criterion to group the pairs of objects with the largest similarity-measure.

Further, it is apparent that the decrease in the similarity-measure depends on the number of objects grouped together. That is why the inequality (1) appears in Theorem 2. This inequality compares the number of pairs of objects grouped together when passing from the classification P to $P_{\mathscr{C}}$ or $P_{\mathscr{D}}$, respectively.

Theorem 2. *Let $\mathscr{V}(P) \geqq r$ and $P_{\mathscr{C}}$ be the classification biinduced by a nonempty transitive set $\mathscr{C} \subseteq \mathscr{K}(P)$. Then every classification $P_{\mathscr{D}}$ biinduced by a nonempty transitive set $\mathscr{D} \subseteq [P]^2$, $\mathscr{D} \nsubseteq \mathscr{K}(P)$ such that*

$$(1) \qquad \sum_{\mathscr{N} \in \mathscr{C}(P)} \sum_{\{N,M\} \in [\mathscr{N}]^2} |N| \cdot |M| \leqq \sum_{\mathscr{M} \in \mathscr{D}(P)} \sum_{\{K,L\} \in [\mathscr{M}]^2} |K| \cdot |L|$$

where $\mathscr{C}(P)$, $\mathscr{D}(P)$ *are the classifications induced by the sets* $\mathscr{C}$ *and* $\mathscr{D}$, *respectively, fulfils* $\mathscr{V}(P_{\mathscr{C}}) > \mathscr{V}(P_{\mathscr{D}})$.

Proof. The similarity-measure $\mathscr{V}(P_{\mathscr{C}})$ is the average of the sequence which contains a-times the value $\mathscr{V}(P)$ (corresponding to the pairs already grouped together in the classification P) and c-times the value r (corresponding to the pairs grouped together when passing from P to $P_{\mathscr{C}}$), where c is the value on the left-hand side of (1). Analogously, $\mathscr{V}(P_{\mathscr{D}})$ is the average of the sequence containing a-times $\mathscr{V}(P)$ and d values which are all less or equal to r (d is the right-hand side of (1)).

Suppose $c \leqq d$. Therefore the inequality

$$\frac{a \,.\, \mathscr{V}(P) + c \,.\, r}{a + c} \geqq \frac{a \,.\, \mathscr{V}(P) + \sum_{k \leqq d} r_k}{a + d},$$

with r_k substituted by r becomes the inequality $a \,.\, r \,.\, (c - d) + a \,.\, \mathscr{V}(P) \,.\, (d - c) \geqq \geqq 0$ which holds provided $r \leqq \mathscr{V}(P)$. Obviously, if some of the members r_k of the sum is lest then r, then sharp inequality holds.

If we manage to observe the requirement $\mathscr{V}(P) \geqq r$ in the course of aglomeration then, when forming the next level of the hierarchical classification, it will be again important to group together pairs of classes from the set $\mathscr{K}(P)$. Let us restrict ourselves to group together only such pairs. Hence we shall work with transitive subsets of the set $\mathscr{K}(P)$.

3. When forming hierarchical classifications, we desire to group together the largest possible number of objects. In accordance with what has been said above the best possible solution is to chosse the transitive subsets of $\mathscr{K}(P)$ which have the largest cardinality. However, computer solution of this task is too pretentious. That is why this requirement will be satisfied only partially; we shall select transitive sets maximal in $\mathscr{K}(P)$. Let us remind that a transitive set $\mathscr{C} \subseteq \mathscr{E} \subseteq [P]^2$ is *maximal* in $\mathscr{E}$ if there exists no transitive set $\mathscr{D} \subseteq \mathscr{E}$ such that $\mathscr{C} \subset \mathscr{D}$.

Theorem 3. *Let* $\mathscr{J} \subseteq \mathscr{K}(P)$ *be the transitive set biinducing the classification* $P_{\mathscr{J}}$, *where* $P_{\mathscr{J}} \neq P_{\max}$. *Then* $\mathscr{J}$ *is maximal in* $\mathscr{K}(P)$ *iff* $\max_{\{N,M\} \in [P]^2} \bar{S}(N, M) > \max_{\{K,L\} \in [P_{\mathscr{J}}]^2} \bar{S}(K\ L)$.

Proof. Denote $r = \max_{\{N,M\} \in [P]^2} \bar{S}(N, M)$. First we suppose that $\mathscr{J}$ is the maximal transitive set in $\mathscr{K}(P)$. It is to be proved that for each pair $\{K, L\} \in \{P_{\mathscr{J}}]^2$, $\bar{S}(K, L) < r$ holds. There are classes $\mathscr{R}$, $\mathscr{L}$ induced by the set $\mathscr{J}$ such that $K = \bigcup \mathscr{R}$ and $L = \bigcup \mathscr{L}$. Then $\bar{S}(K, L)$ can be expressed as

$$\bar{S}(K, L) = \frac{1}{\sum_{\{N,M\} \in [\mathscr{R}, \mathscr{L}]} |N| \,.\, |M|} \sum_{\{N,M\} \in [\mathscr{R}, \mathscr{L}]} \bar{S}(N, M) \,.\, |N| \,.\, |M| \,.$$

The definition of r implies that $\bar{S}(N, M) \leqq r$ for each pair $\{N, M\} \in [\mathcal{R}, \mathcal{L}]$. It is sufficient to prove that for at least one pair the sharp inequality holds. Let $\bar{S}(N, M) = r$ for each pair $\{N, M\} \in [\mathcal{R}, \mathcal{L}]$. We shall show that the set $\mathcal{T} = \mathcal{J} \cup \cup [\mathcal{R}, \mathcal{L}]$ is a transitive subsets of the set $\mathcal{K}(P)$ and $\mathcal{J} \subset \mathcal{T}$. If $\{N, M\}, \{M, T\} \in \mathcal{T}$, then the following cases are possible:

$$\{N, M\}, \{M, T\} \in \mathcal{J} ;$$
$$\{N, M\}, \{M, T\} \in [\mathcal{R}, \mathcal{L}] ;$$
$$\{N, M\} \in \mathcal{J} \,\&\, \{M, T\} \in [\mathcal{R}, \mathcal{L}] .$$

The first two trivially imply $\{N, T\} \in \mathcal{T}$. In the last case $\{M, T\} \in [\mathcal{R}, \mathcal{L}]$ implies $M \in \mathcal{R}$ or $M \in \mathcal{L}$. Then $N \in \mathcal{R}$ or $N \in \mathcal{L}$, as $\mathcal{R}, \mathcal{L}$ are induced by $\mathcal{J}$. $\{N, T\} \in \mathcal{T}$ according to the definition of $\mathcal{T}$. If $\mathcal{T} = \mathcal{J}$, then $\mathcal{J}$ induces the class containing $\mathcal{R} \cup \mathcal{L}$. This, however, is not the case, as $\mathcal{J}$ induces the classes $\mathcal{R}, \mathcal{L}$. Hence $\mathcal{J} \subset \mathcal{T}$. We have shown the set $\mathcal{J}$ not to be maximal – a contradiction. Therefore there exists a pair $\{N, M\} \in [\mathcal{R}, \mathcal{L}]$ for which $\bar{S}(N, M) < r$ holds.

Converse implication. If $\mathcal{J}$ is not maximal, then there exists a transitive set $\mathcal{T}$, $\mathcal{K}(P) \supseteq \mathcal{T} \supset \mathcal{J}$ and a pair $\{N, M\} \in \mathcal{T} \setminus \mathcal{J}$. If $\mathcal{J}(P)$ is the classification induced by the set $\mathcal{J}$, then there exists a pair $\{\mathcal{N}, \mathcal{M}\} \in [\mathcal{J}(P)]^2$ such that $N \in \mathcal{N}$ and $M \in \mathcal{M}$. As the transitive set $\mathcal{T}$ contains the sets $[\mathcal{N}]^2$, $[\mathcal{M}]^2$ and the pair $\{N, M\}$, it must also contains the set $[\mathcal{N}, \mathcal{M}]$. Consequently $\bar{S}(K, L) = r$ for the classes $K = \bigcup \mathcal{N}$, $L = \bigcup \mathcal{M}$ of the classification $P_{\mathcal{J}}$. If $\mathcal{J}$ is not maximal, there is no decrease of the maximum similarity of the classes of the classification.

4. Maximal transitive set $\mathcal{J} \subseteq \mathcal{K}(P) \subseteq [P]^2$ is called a *quasi-kernel* of P.

We shall say that a hierarchical classification $P_0 \ldots P_n$ is *very proper* if for every k, $0 < k \leqq n$, the classification P_k is biinduced by a quasi-kernel of P_{k-1}.

Theorem 4. *Let $P_0 \ldots P_n$ be a very proper hierarchical classification. Denote* $r_k = \max\limits_{\{N,M\} \in [P_k]^2} \bar{S}(N, M)$. *Then*

(1) $(\forall k)(0 < k < n \rightarrow r_{k-1} > r_k)$,

(2) $(\forall P)(\forall k)(0 < k < n \,\&\, P_k$ *is a refinement of* $P \rightarrow \mathcal{V}(P) < \mathcal{V}(P_k))$,

(3) $(\forall k)(0 < k < n \rightarrow \mathcal{V}(P_{k+1}) > r_k)$.

Proof. (1) holds according to Theorem 3.

(2) and (3) will be proved simultaneously by induction with respect to k.

For $k = 1$ it is

$$\mathcal{V}(P_1) = \frac{1}{\sum\limits_{N \in P_1} \binom{|N|}{2}} \sum_{N \in P_1} \sum_{\{x,y\} \in [N]^2} S(x, y) = r_0 .$$

If P_1 is a refinement of P, then P is biinduced by a transitive set $\mathscr{C} \nsubseteq \mathscr{K}(P_0)$. In accordance with Theorem 1 it is $\mathscr{V}(P) < r_0 = \mathscr{V}(P_1)$.

Suppose (2) and (3) are proved for $k - 1$.

Let P_k be a refinement of P and $\mathscr{P}$ a classification of P_k corresponding to P. Then for each class $\mathscr{N} \in \mathscr{P}$ and each pair $\{N, M\} \in [\mathscr{N}]^2$, $\bar{S}(N, M) \leqq r_k$ holds.

The similarity-measure $\mathscr{V}(P)$ can be expressed by means of classes of the classification P_k:

$$\mathscr{V}(P) = \frac{1}{\sum_{N \in P_k} \binom{|N|}{2} + \sum_{\mathscr{N} \in \mathscr{P}} \sum_{\{N,M\} \in [\mathscr{N}]^2} |N| \cdot |M|} \left(\mathscr{V}(P_k) \sum_{N \in P_k} \binom{|N|}{2} + \right.$$
$$\left. + \sum_{\mathscr{N} \in \mathscr{P}} \sum_{\{N,M\} \in [\mathscr{N}]^2} \bar{S}(N, M) \cdot |N| \cdot |M| \right).$$

From the inductive hypothesis $r_k < r_{k-1} < \mathscr{V}(P_k)$ and from the inequality $\bar{S}(N, M) \leqq r_k$ it is evident that $\mathscr{V}(P) < \mathscr{V}(P_k)$. In particular, for $P = P_{k+1}$ it is $\bar{S}(N, M) = r$ for each pair $\{N, M\} \in [\mathscr{N}]^2$ and consequently $\mathscr{V}(P_{k+1}) > r_k$.

Corollary. *Every very proper hierarchical classification is proper.*

5. We shall show that the concept of very proper hierarchical classification excludes first of all those proper hierarchical classifications for which finer proper hierarchical classification exist.

Lemma 3. *Hierachical classification $P'_0 \dots P'_l$ selected from a very proper hierarchical classification $P_0 \dots P_n$ in such a manner that $l < n$ is a proper hierarchical classification but not a very proper one. Hierarchical classification which results from a very proper hierarchical classification by increasing the number of levels is not a proper hierarchical classification.*

Proof. As $l < n$, there exists such a segment $P_u, P_{u+1}, \dots, P_v$ of hierarchical classification $P_0 \dots P_n$ that P_u, P_v are adjacent classifications in the hierarchical classification $P'_0 \dots P'_l$. The classification P_{u+1} is biinduced by a quasi-kernel $\mathscr{J}$ of P_u and every classification refinement of which is P_{u+1} is biinduced by a transitive set $\mathscr{C} \subseteq [P_u]^2$ for which $\mathscr{J} \subset \mathscr{C}$. $\mathscr{C}$ is not a quasi-kernel. P_v is not biinduced by a quasi-kernel of P_u and the hierarchical classification $P'_0 \dots P'_l$ is not very proper although it is proper owing to the fact that the properties of proper hierarchical classification are obviously conserved.

The number of levels of a hierarchical classification can be increased by inserting a classification P between two adjacent classifications P_k, P_{k+1} which is a refinement of P_{k+1} while P_k is a refinement of P. If P_{k+1} is biinduced by a quasi-kernel $\mathscr{J}$ of P_k then P is biinduced by a transitive subset of $\mathscr{J}$ and Theorem 3 implies that the condition (2) in the definition of proper hierarchical classification is not satisfied.

Example 3. Let the similarity-function S on $E = \{x, y, u, v\}$ be given by Table 3.

	y	u	v
x	4	2	2
y		2	2
u			1

We shall construct all proper hierarchical classification. For the classification $P_{\min}$ there exists only one quasi-kernel which biinduces a classification $P_1 : \{x, y\} \{u\} \{v\}$. The values of the similarity-function $\bar{S}$ on P_1 are given in Table 4.

	$\{u\}$	$\{v\}$
$\{x, y\}$	2	2
$\{u\}$		1

It is obvious from Table 4 that there are two quasi-kernel biinducing classifications

$$P_2 : \{x, y, u\} \{v\} ,$$
$$P_2' : \{x, y, v\} \{u\} .$$

Consequently, the only very proper hierarchical classifications of E are $P_{\min}, P_1, P_2, P_{\max}$ and $P_{\min}, P_1, P_2', P_{\max}$. The hierarchical classification $P_{\min}, P, P_{\max}$ where $P : \{x, y\} \{u, v\}$ is proper and is not selected from the very proper hierarchical classification.

KERNEL OF CLASSIFICATION

1. A nonempty transitive set $\mathscr{A} \subseteq [P]^2$ is *connected* if $(\forall x)(\forall y)(\forall u)(\forall v)((\{x, y\} \in \in \mathscr{A} \;\&\; \{u, v\} \in \mathscr{A} \;\&\; x \neq u) \rightarrow \{x, u\} \in \mathscr{A})$. A transitive set $\mathscr{A} \subseteq [P]^2$ is connected iff it induces a single more-element class on P. Denoting this class by $\mathscr{N}$, we have $\mathscr{A} = [\mathscr{N}]^2$.

Lemma 4. *Let $\mathscr{A}, \mathscr{B} \subseteq [P]^2$ be disjoint connected transitive sets. If $\mathscr{A} = [\mathscr{N}]^2$ and $\mathscr{B} = [\mathscr{M}]^2$ then $\mathscr{A} \cup \mathscr{B}$ is transitive iff $\mathscr{N} \cap \mathscr{M} = \emptyset$.*

If, moreover, $\mathscr{A}$ and $\mathscr{B}$ are nonempty and $\mathscr{A} \cup \mathscr{B}$ is transitive then $\mathscr{A} \cup \mathscr{B}$ is not connected.

Lemma 5. *Let $\mathscr{C}$ be a transitive set, $\mathscr{C} \subseteq [P]^2$. There exists one and only one set $\{\mathscr{A}_1 \dots \mathscr{A}_m\}$ the elements of which are transitive connected pairwise disjoint sets and $\mathscr{C} = \bigcup_{i=1}^{m} \mathscr{A}_i$.*

Proof. Let $\mathscr{C}(P)$ be the classification induced by the set $\mathscr{C}$. We shall number the classes of the classification $\mathscr{C}(P)$. Denote $\mathscr{N}_1 \dots \mathscr{N}_m, \mathscr{N}_{m+1} \dots \mathscr{N}_p$, where $\mathscr{N}_{m+1} \dots \dots \mathscr{N}_p$ are just all the one-element classes. Define sets $\mathscr{A}_i$ by $\mathscr{A}_i = [\mathscr{N}_i]^2$ for $i = 1 \dots m$. It is apparent from the definition that are transitive connected pairwise disjoint sets and $\mathscr{C} = \bigcup_{i=1}^{m} \mathscr{A}_i$.

We shall prove the uniqueness. Let one more decomposition of $\mathscr{C}$ exists, $\{\mathscr{B}_1 \dots \mathscr{B}_n\}$, satisfying the conditions of Lemma 5. Consider any set $\mathscr{B}_j$. We shall prove $\mathscr{B}_j = \mathscr{A}_i$ for some $i \leqq m$. Take a pair $\{x, y\} \in \mathscr{B}_j$. Then there is $\mathscr{A}_i$ such that $\{x, y\} \in \mathscr{A}_i$ owing to $\mathscr{C} = \bigcup_{i=1}^{m} \mathscr{A}_i = \bigcup_{j=1}^{n} \mathscr{B}_j$. Let $\mathscr{B}_j = [\mathscr{M}_j]^2 \neq \mathscr{A}_i = [\mathscr{N}_i]^2$, which means e.g. that there exists $z \in \mathscr{M}_j \setminus \mathscr{N}_i$. Then $\{x, z\} \in [\mathscr{M}_j]^2 \setminus [\mathscr{N}_i]^2$ and $\{x, z\} \in [\mathscr{N}_k]^2$ for some k, $k \neq i$, consequently $x \in \mathscr{N}_i \cap \mathscr{N}_k$ and the sets are not disjoint — contradiction.

Remark 1. Investigating transitive sets it is possible to restrict oneself without loss of generality to connected transitive sets.

2. Every connected transitive set (its corresponding biinduced class) is the collection of similar elements. The following Theorem suggests that the greater is the cardinality of this collection, the greater is its "consistence".

Theorem 5. *Let $\mathscr{A}, \mathscr{B} \subseteq \mathscr{K}(P_{\min})$ be connected transitive sets biinducing the classes N, M with $|N| < |M|$. Suppose that there exist objects $x, y \in E$ such that $x \notin N$, $y \notin M$ and*

$$(1) \qquad \bar{S}(N, \{x\}) = \bar{S}(M, \{y\}) .$$

Denote by P and P' the classifications with one and only one more-element class $N \cup \{x\}$ and $M \cup \{y\}$, respectively. Then

$$(2) \qquad \mathscr{V}(P) \leqq \mathscr{V}(P')$$

$$\text{and} \quad \mathscr{V}(P) < \mathscr{V}(P') \leftrightarrow (\exists z \in N) (\{x, z\} \notin \mathscr{K}(P_{\min})) .$$

Proof. We can express the similarity-measure of the classification P by means of the cardinality of its more-element class:

$$\mathscr{V}(P) = \frac{n-1}{n+1} r + \frac{2}{n(n+1)} \sum_{z \in N} S(x, z)$$

where $n = |N|$ and r is the value of the similarity-function S on $\mathscr{K}(P)$.

If we replace n by m with $m = |M|$, we obtain the analogous equality for $\mathscr{V}(P')$. Then by (2) we have

$$(3) \qquad \frac{2}{n(n+1)} \sum_{z \in N} S(x, z) - \frac{2}{m(m+1)} \sum_{z \in M} S(y, z) \leqq \left(\frac{m-1}{m+1} - \frac{n-1}{n+1}\right) r .$$

In accordance with the equality (1) and the assumption $m > n$ (3) is equivalent to the inequality $(1/n) \sum_{z \in N} S(x, z) \leqq r$. Hence the first assertion holds.

Analogously, the inequality $\mathscr{V}(P) < \mathscr{V}(P')$ is equivalent to the condition $(1/n) \sum_{z \in N} S(x, z) < r$ which holds iff there exists an object $z \in N$ and $S(x, z) < r$ (i.e. $\{x, z\} \notin \mathscr{K}(P_{\min})$).

Remark 2. In Theorem 5 the implication $(\exists z \in N)(\{x, z\} \notin \mathscr{K}(P_{\min})) \to \mathscr{V}(P) < \mathscr{V}(P')$ holds even under a weaker assumption, $\bar{S}(N, \{x\}) \leqq \bar{S}(M, \{y\})$.

We have already mentioned that the choice of transitive sets which have the largest cardinality is too computer-time consuming. That is why we shall at least prefer connected transitive sets which are maximal with respect to inclusion.

3. A set $\tilde{\mathscr{D}}^{\mathscr{E}} = \{\{x, y\};\ \{x, y\} \in \mathscr{E} \,\&\, (\exists z \in P)(\{x, z\} \in \mathscr{D} \vee \{y, z\} \in \mathscr{D})\}$ where $\mathscr{E} \subseteq [P]^2$ is called the *closure* of $\mathscr{D}$ in $\mathscr{E}$. A set $\mathscr{J} \subseteq \mathscr{K}(P) \subseteq [P]^2$ is called the kernel of P, if there exists a sequence of sets $\mathscr{A}_1 \ldots \mathscr{A}_m$ satisfying the following conditions:

(1) $\mathscr{J} = \bigcup_{i=1}^{m} \mathscr{A}_i$,

(2) $(\forall i)(1 \leqq i \leqq m \to \mathscr{A}_i$ is maximal connected transitive set in $\mathscr{K}(P) \setminus \bigcup_{j=1}^{i-1} \tilde{\mathscr{A}}_j^{\mathscr{K}(P)}$,

(3) $\mathscr{K}(P) = \bigcup_{i=1}^{m} \tilde{\mathscr{A}}_i^{\mathscr{K}(P)}$.

Theorem 6. *Every kernel of P is also a quasi-kernel of P.*

Proof. Let $\mathscr{A}_1 \ldots \mathscr{A}_m$ be the sequence of sets determining the kernel $\mathscr{J}$. Condition (2) implies that the classes induced by sets $\mathscr{A}_i$ for $i = 1 \ldots m$ are disjoint. Hence obviously $\mathscr{J}$ is transitive.

It remains to prove that $\mathscr{J}$ is maximal. In the opposite case there exists a transitive set $\mathscr{T}$ where $\mathscr{K}(P) \supseteq \mathscr{T} \supset \mathscr{J}$. This means that there exists at least one pair $\{u, v\} \in \mathscr{T}$ which is not in $\mathscr{J}$. Let s be the first index such that there exists a pair $\{u, v\} \in \tilde{\mathscr{A}}_s^{\mathscr{K}(P)} \setminus \mathscr{A}_s$ and $\{u, v\} \in \mathscr{T}$. This set exists with respect to conditions (1) and (3). Define $\mathscr{B} = \mathscr{A}_s \cup [\mathscr{N}, \{u, v\}]$, where $\mathscr{N}$ is the class biinduced by the set $\mathscr{A}_s$. Clearly $\mathscr{B}$ is connected and transitive since $\mathscr{B} = [\mathscr{N} \cup \{u, v\}]^2$. Moreover, it holds $\mathscr{A}_s \subset \mathscr{B}$ as $\{u, v\} \in \mathscr{B}$ and $\{u, v\} \notin \mathscr{A}_s$. The inclusion $\mathscr{B} \subseteq \mathscr{K}(P) \setminus \bigcup_{j=1}^{s-1} \tilde{\mathscr{A}}_j^{\mathscr{K}(P)}$ is equivalent to the assertion $\{x, y\} \in [\mathscr{N} \cup \{u, v\}]^2 \to \{x, y\} \notin \tilde{\mathscr{A}}_t^{\mathscr{K}(P)}$ for all t, $1 \leqq t < s$. The validity of this condition is quaranteed by the choice of the index s.

Hence $\mathscr{A}_s$ is not maximal which is a contradiction.

Corollary. *Let $P_0 \ldots P_n$ be a hierarchical classification of E such that*

$$(\forall k)(0 \leqq k < n \to P_{k+1} \text{ is biinduced by a kernel of } P_k).$$

Then $P_0 \ldots P_n$ is a very proper hierarchical classification.

Example 4. *Not every quasi-kernel is a kernel.* The set $\mathcal{K}(P) = \{\{x, y\} \{y, z\} . . \{x, z\} \{y, u\} \{z, u\}\}$ has two and only two maximal transitive sets $\mathcal{C} = \{\{x, y\} . . \{y, z\} \{x, z\}\}$ and $\mathcal{B} = \{\{y, z\} \{y, u\} \{z, u\}\}$. Evidently, these sets are just all kernels of P. The set $\mathcal{D} = \{\{x, y\} \{z, u\}\}$ is maximal transitive in $\mathcal{K}(P)$. Hence, $\mathcal{D}$ is a quasi-kernel but not a kernel.

4. A transitive set $\mathcal{D}$, $\mathcal{D} \subseteq \mathcal{E} \subseteq [P]^2$ and $\tilde{\mathcal{D}}^{\mathcal{E}} = \mathcal{D}$ is called *closed* in $\mathcal{E}$.

If $\mathcal{D}$ is closed connected subset $\mathcal{E}$ then $\mathcal{D}$ is maximal connected transitive in $\mathcal{E}$.

Lemma 6. *Let $\mathcal{J}$ be a kernel of P and let $\mathcal{A}_1 \dots \mathcal{A}_m$ be the sequence of sets by which $\mathcal{J}$ is given. Then there exists, for every closed connected set $\mathcal{D}$ in $\mathcal{K}(P)$, an index j so that $\mathcal{D} = \mathcal{A}_j$.*

ALGORITHM OF THE METHOD

In the preceding part of this paper we have described a construction of hierarchical classification with "good properties". We shall call the algorithm realizing this construction "algorithm of consecutive classification".

We shall show its complete realization by means of flowcharts. First we shall design two simple algorithms which are its constituents.

1. *Algorithm of selection of maximal connected transitive set $\mathcal{A}$, $\mathcal{A} \subseteq \mathcal{E} \subseteq [P]^2$.* For $x \in P$, denote by G_x the set $\{y; y \in P \,\&\, \{x, y\} \in \mathcal{E}\}$ which is used in the flowchart 1.

It is easy to verify that the output set $\mathcal{A} = [\{x_0 \dots x_i\}]^2$ is maximal connected transitive in $\mathcal{E} \subseteq [P]^2$.

2. *Algorithm of selection of kernel.* Sets $\mathcal{A}_1 \dots \mathcal{A}_m$ are consecutively selected satisfying the following conditions:

(1) $\mathcal{A}_i$ is transitive and connected.

(2) Denote $\mathcal{L}_i = \mathcal{K}(P) \setminus \bigcup_{j=1}^{i-1} \tilde{\mathcal{A}}_j^{\mathcal{K}(P)}$, then $\mathcal{A}_i \subseteq \mathcal{L}_i$ and $\mathcal{A}_i$ is maximal connected in $\mathcal{L}_i$.

(3) $\mathcal{J} = \bigcup_{i=1}^{m} \mathcal{A}_i$.

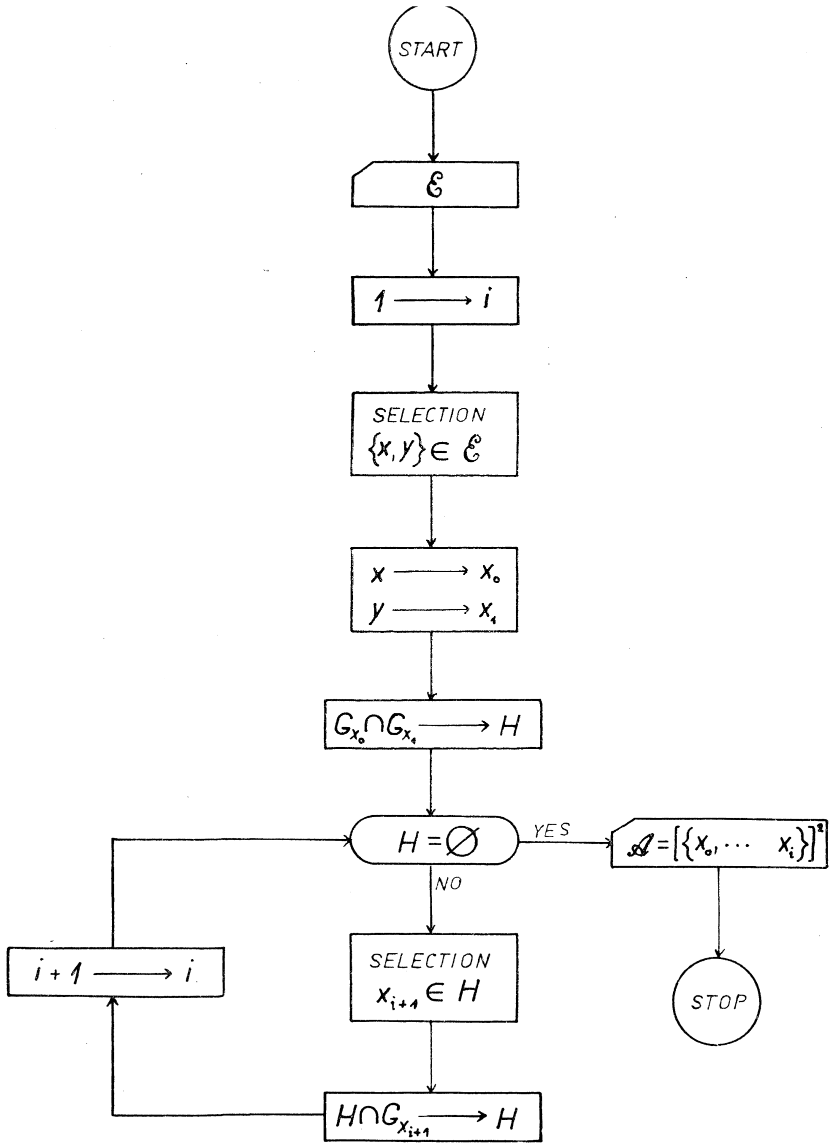
(4) $\mathcal{K}(P) = \bigcup_{i=1}^{m} \tilde{\mathcal{A}}_i^{\mathcal{K}(P)}$.

Evidently, the set $\mathcal{J}$ selected by the algorithm is a kernel.

3. *Algorithm of consecutive classification.*

Description:

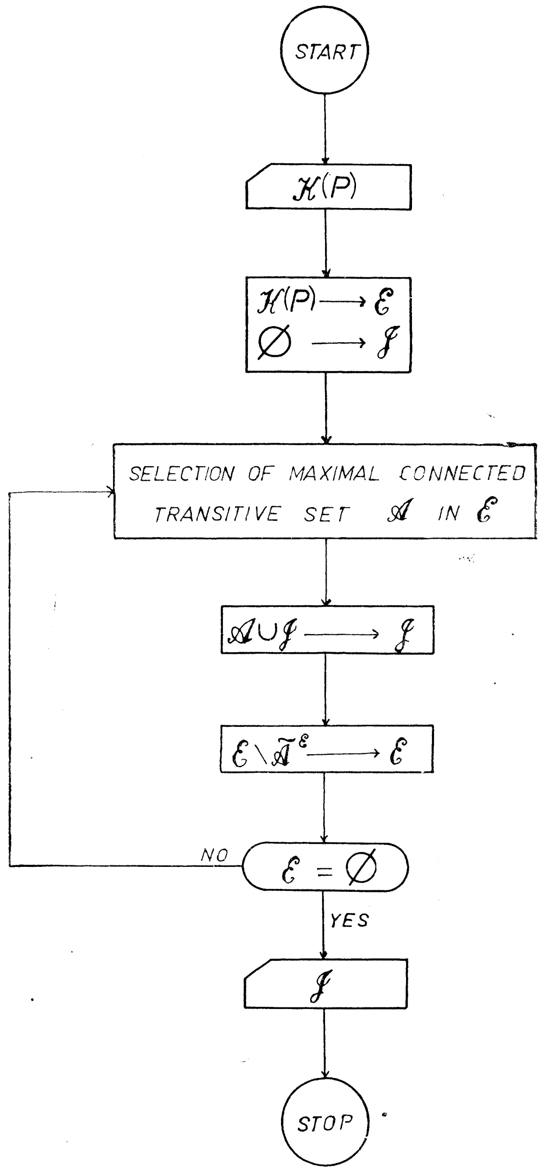
1) The first classification of the desired hierarchical classification is $P_{\min}$ (one-element classes).



Flowchart 1.

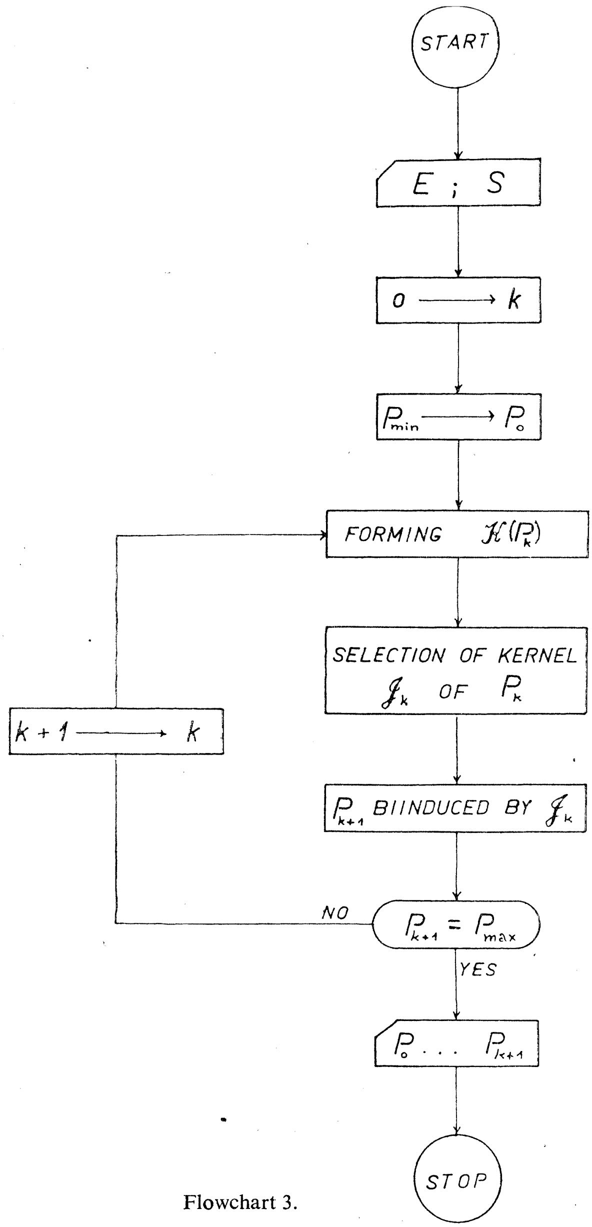
2) Let P_k be the k-th level of the hierarchical classification. We shall select a suitable transitive subset of $[P_k]^2$ (e.g. a kernel, a quasi-kernel) biinducing the classification P_{k+1} of the hierarchical classification taking into account the values of the function $\bar{S}$ on $[P_k]^2$ and contingently also other criteria.

3) If $P_k = P_{\max}$ then the sequence $P_0 \ldots P_k$ obtained is a hierarchical classification.



Flowchart 2.

Given a set E and a similarity-function S with the domain $[E]^2$ then the algorithm of consecutive classification determines a very proper hierarchical classification of E. The selection of this hierarchical classification is given by the selection of kernels in the individual cycles and this one again depends on the selection of non-ordered



Flowchart 3.

pairs forming the maximal connected transitive sets. For example, choosing the pair $\{\{x, y\} \{u\}\}$ or $\{\{x, y\} \{v\}\}$ in Example 3 we obtain the classification P_2 or P'_2, respectively. In a concrete program the relaization of the selection depends on the ordering of the set E and of the values of the function $\bar{S}$. In order to reduce the influence of the subject we may modify the algorithm so that in every cycle all kernels would be found instead of the one preferred by the ordering of the data. It is not evident if it is possible to carry out the modification with sufficient economy.

It seems that in some cases another possibility would be to proceed as follows:

1) We insert other objects which represent a certain hypothesis in the ordered set under classification so that these objects affect already the first selections of kernels. From the point of view of this hypothesis it is interesting to examine the course of hierarchical classification with respect to these objects.

2) We consider a classification P and a similarity-function among its classes and replace the set under classification by P. The algorithm gives a hierarchical classification of P. Analogously we can investigate parts of a hierarchical classification or affect a hierarchical classification by other criteria following from the aim of our investigation.

APLICATION OF ALGORITHM

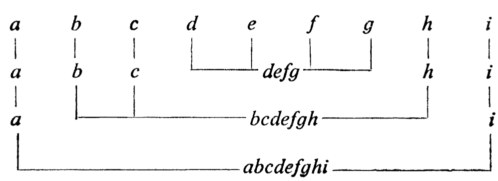
We shall compare the algorithm of consecutive classification and the lexicographic algorithm (see [1]) on a simple example. Purposly, we choose an example exposing clearly the inaccuracy which is removed by the use of the algorithm of consecutive classification.

Example:

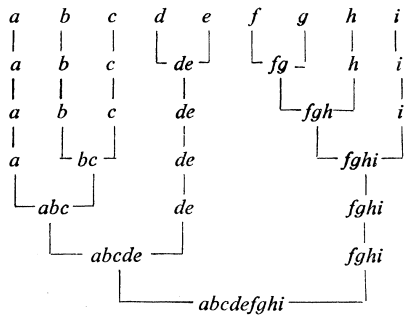
Similarity-function with the domain $[E]^2$, where $E = \{a, b, c, d, e, f, g, h, i\}$, is given in the table.

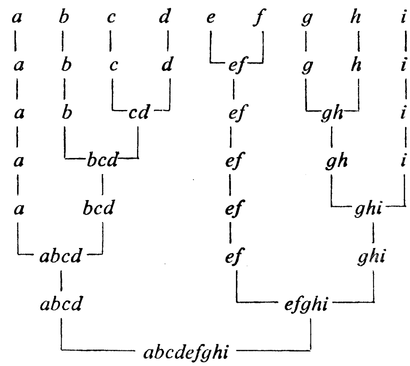
	b	*c*	*d*	*e*	*f*	*g*	*h*	*i*
a	2	1	2	0	0	0	0	0
b		2	3	0	0	0	0	0
c			3	0	0	0	0	0
d				4	0	0	0	0
e					4	0	0	0
f						4	3	2
g							3	2
h								2

The lexicographic algorithm yields the following hierarchical classification:



The result of the algorithm of consecutive classification depends on the selection of kernels. We present only two possible solutions. The other differ from them very little.





CONCLUSION

Analysing hierarchical classifications we obtained an aglomerative procedure resulting in hierarchical classifications with good properties. We believe that it is possible to carry out this procedure (algorithm of consecutive classification) in practice. The research is far from being closed. Let us mention at least some open problems.

1) We have considered only the similarity-function S with the domain $[E]^2$ and its extension by the mean value. We can use also other extension of the similarity-function or directly a similarity-function whose domain is $[\mathscr{P}_0(E)]^2$ and then to investigate properties of hierarchical classifications obtained in this way.

2) We selected the transitive sets from the first component of the quasiordering determined by the function $\bar{S}$. Analogously we could select the maximal transitive sets from a section of quasiordering (a set $\mathscr{L}(E) \subseteq [E]^2$ is called a section of the quasiordering $\leqq$ if the condition $(\forall x, y, z, v)\,(\{x, y\} \in [E]^2 \,\&\, \{z, v\} \in \mathscr{L}(E) \,\&$ $\&\, \{x, y\} \leqq \{z, v\} \rightarrow \{x, y\} \in \mathscr{L}(E))$ holds) and if need be, we could control the choice of the section by decreasing similarity. This procedure is quicker but the hierarchical classifications obtained are less fine.

3) It would be interesting to estimate the number of proper and very proper hierarchical classifications.

4) Algorithm of consecutive classification is an aglomerative method. In a similar way it is possible to form a splitting method which proceeds from the classification $P_{\max}$ and formes the separate levels of the hierarchical classification by splitting the classes of the preceding classification.

Acknowledgement. The author is deeply indebted to RNDr. Petr Štěpánek, CSc., PhDr. Jiří Polívka and RNDr. Petr Hájek, CSc. for valuable comments and suggestions. Special thanks are due to RNDr. P. Hájek, CSc. for the opportunity of reporting the results of the present paper at the Seminar "Applications of Mathematical Logic" at the Faculty of Mathematics and Physics, Charles University.

References

[1] *I. C. Lerman:* Les bases de la classification automatique. Paris, 1970.

[2] *P. Macnaughton - Smith:* Some statistical and other numerical techniques for classifying individuals, Home Office Research Unit Report, s 1—31. London, 1965.

Souhrn

ALGOMERATIVNÍ METODA AUTOMATICKÉHO VYTVÁŘENÍ HIERARCHICKÝCH KLASIFIKACÍ

JOSEF FUČÍK

V článku je navržena nová metoda pro vytváření hierarchických klasifikací. V prvé části jsou zavedeny některé pojmy (rozšíření podobnostní funkce, podobnostní míra, tranzitivní množiny), které umožňují studovat hierarchické klasifikace. V průběhu dalšího zkoumání je implicitně naznačen algoritmus pro získávání hierarchických klasifikací s „dobrými vlastnostmi". Tento klasifikační postup je v závěru přesně popsán a znázorněn blokovými schématy. Teoretické opodstatnění metody je obsaženo především ve větách 1 až 4, které ukazují souvislost mezi vlastnostmi tranzitivních množin a hierarchické klasifikace. Praktické použití metody předpokládá konkrétní zavedení podobnostní funkce a příslušnou formu zadání zkoumaného materiálu.

Author's address; RNDr. *Josef Fučík,* Ústav výpočtové techniky ČVUT, Horská 3, 128 00 Praha 2.