KYBERNETIKA ČÍSLO 1, ROČNÍK 3/1967

Quantification Method of Classification Processes

Concept of Structural a-Entropy

JAN HAVRDA, FRANTIŠEK CHARVÁT

The aim of this paper is to form a quantificatory theory of classificatory processes. A concept of structural a-entropy is defined and its form is derived.

Definition 1. Let B be a non-empty set with a normed measure (it is a measure defined on the set of all subsets of B such that the measure of B is 1). Let $\{\mathscr{K}_\nu\}_{\nu \in \mathscr{I}}$ be an indexed set of finite families $\mathscr{K}_\nu$ of propositional functions on B ($\mathscr{K}_\nu = \{{}^\nu p_1, \ldots, {}^\nu p_{N_\nu}\}$, where N_ν is a positive integer) such that

$$\bigcup_{i=1}^{N_\nu} M_i(\mathscr{K}_\nu) = B\,, \quad M_i(\mathscr{K}_\nu) \cap M_j(\mathscr{K}_\nu) = \emptyset \quad \text{for} \quad i \neq j\,, \quad i, j = 1, 2, \ldots, N_\nu\,,$$

and for every $\nu \in \mathscr{I}$ where $M_i(\mathscr{K}_\nu) = \{x : x \in B \text{ and } {}^\nu p_i(x) \text{ holds}\}$. The family $\{M_i(\mathscr{K}_\nu)\}_1^{N_\nu}$ the set B, and the family $\mathscr{K}_\nu$ are said classification, base of classification, and classificatory criteria, respectively.

In the sequel we will denote the classification only by $\mathscr{R}(B) = \{M_i\}_1^N$ because we shall not distinguish among classificatory criteria. Let us discuss Definition 1 in more detail: the classification was defined on the sets with normed measure and, conseqnetly, we have simultaneously introduced a quantification of the base of classification. However, it is purposeful to quantificative the classifications of given base. According to this purpose we shall give some formal considerations and denotations: every element of $\mathscr{R}(B)$ we call element of classification: every element $M_i \in \mathscr{R}(B)$ has a measure $\mu(M_i)$, $i = 1, \ldots, N$. The measures $\mu(M_i)$ will serve here as foundation means for quantification of classification and therefore we shall write the classification in the sequel as $\mathscr{R}(B) = \{M_1, \ldots, M_N, \mu_1, \ldots, \mu_N\}$, where $\mu_i = \mu(M_i)$.

In this paper we introduce axiomatically a real function of classifications, so called structural a-entropy, which can serve as a quantitative measure of classification. It will be shown, that there is an analogy between a-entropy and the usual entropy from information theory.

Definition 2. Let $\mathscr{R}(B) = \{M_1, \ldots, M_N, \mu_1, \ldots, \mu_N\}$ be a classification. A function $S(\mu_1, \ldots, \mu_N; a)$ will be said structural a-entropy if

a) $S(\mu_1, \ldots, \mu_N; a)$ is continuous in the region $\mu_i \geqq 0, \sum_{i=1}^{i} \mu_i = 1, a > 0$;

b) $S(1; a) = 0,\ S(\frac{1}{2}, \frac{1}{2}; a) = 1$;

c) $S(\mu_1, \ldots, \mu_{i-1}, 0, \mu_{i+1}, \ldots, \mu_N; a) = S(\mu_1, \ldots, \mu_{i-1}, \mu_{i+1}, \ldots, \mu_N; a)$ for every $i = 1, 2, \ldots, N$;

d) $S(\mu_1, \ldots, \mu_{i-1}, \nu_{i_1}, \nu_{i_2}, \mu_{i+1}, \ldots, \mu_N; a) = S(\mu_1, \ldots, \mu_{i-1}, \mu_i, \mu_{i+1}, \ldots, \mu_N; a) + \alpha\mu_i^a S\left(\frac{\nu_{i_1}}{\mu_i}, \frac{\nu_{i_2}}{\mu_i}; a\right)$ for every $\nu_{i_1} + \nu_{i_2} = \mu_i > 0,\ i = 1, 2, \ldots, N, \alpha > 0$.

The meaning of axioms a)–c) is clear. What concerns axiom d), an increase of the structural a-entropy provided that the classification is "refined" depends on the parameter a which will be said characteristic parameter.

Theorem 1. *Axioms* a) –d) *determine the structural a-entropy unambiguously by*

$$S(\mu_1, \ldots, \mu_N; a) = \frac{2^{a-1}}{2^{a-1} - 1}\left(1 - \sum_{i=1}^{N} \mu_i^a\right) \quad \text{for} \quad a > 0, \quad a \neq 1,$$

$$S(\mu_1, \ldots, \mu_N; 1) = -\sum_{i=1}^{N} \mu_i \log \mu_i,$$

where log *is here and in the sequel taken to the base* 2.

Proof of this theorem will be based on the following lemmas:

Lemma 1. $\alpha = 1$.

Proof. According to d)

$$S(\tfrac{1}{2}, \tfrac{1}{2}; a) = S(1; a) + \alpha S(\tfrac{1}{2}, \tfrac{1}{2}; a),$$

which immediately implies the desired assertion (cf. b)).

Lemma 2. *If* $\nu_k \geqq 0,\ k = 1, \ldots, m, \sum_{k=1}^{m} \nu_k = \mu_i > 0,$ *then*

$$S(\mu_1, \ldots, \mu_{i-1}, \nu_1, \ldots, \nu_m, \mu_{i+1}, \ldots, \mu_N; a) = S(\mu_1, \ldots, \mu_N; a) + \mu_i^a S\left(\frac{\nu_1}{\mu_i}, \ldots, \frac{\nu_m}{\mu_i}; a\right).$$

Proof. To prove this Lemma we argue by induction. For $m = 2$ the desired statement holds (cf. d) and Lemma 1). Using Lemma 1, d) and the induction premise we obtain the following result

$$S(\mu_1, \ldots, \mu_{i-1}, \nu_1, \ldots, \nu_{m+1}, \mu_{i+1}, \ldots, \mu_N; a) =$$
$$= S(\mu_1, \ldots, \mu_{i-1}, \nu_1, \bar{\mu}, \mu_{i+1}, \ldots, \mu_N; a) +$$
$$+ \bar{\mu}^a S\left(\frac{\nu_2}{\bar{\mu}}, \ldots, \frac{\nu_{m+1}}{\bar{\mu}}; a\right) = S(\mu_1, \ldots, \mu_N; a) + \mu_i^a S\left(\frac{\nu_1}{\mu_i}, \frac{\bar{\mu}}{\mu_i}; a\right) +$$
$$+ \bar{\mu}^a S\left(\frac{\nu_2}{\bar{\mu}}, \ldots, \frac{\nu_{m+1}}{\bar{\mu}}; a\right),$$

where $\bar{\mu} = \nu_2 + \ldots + \nu_{m+1}$. One more application of the induction premise yields

$$S\left(\frac{\nu_1}{\mu_i}, \ldots, \frac{\nu_{m+1}}{\mu_i}; a\right) = S\left(\frac{\nu_1}{\mu_i}, \frac{\mu}{\bar{\mu}_i}; a\right) + \left(\frac{\bar{\mu}}{\mu_i}\right)^a S\left(\frac{\nu_2}{\bar{\mu}}, \ldots, \frac{\nu_{m+1}}{\bar{\mu}}; a\right)$$

and hence, in view of the preceding equality, the statement of Lemma 2 holds.

The following Lemma is an obvious consequence of Lemma 2.

Lemma 3. *If* $\nu_{ij} \geqq 0$, $j = 1, 2, \ldots, m_i$, $\sum_{j=1}^{m_i} \nu_{ij} = \mu_i > 0$, $i = 1, 2, \ldots, n$, $\sum_{i=1}^{n} \mu_i = 1$, *then*

$$S(\nu_{11}, \ldots, \nu_{1m_1}, \ldots, \nu_{n1}, \ldots, \nu_{nm_n}; a) =$$
$$= S(\mu_1, \ldots, \mu_n; a) + \sum_{i=1}^{n} \mu_i^a S\left(\frac{\nu_{i1}}{\mu_i}, \ldots, \frac{\nu_{im_i}}{\mu_i}; a\right).$$

If we replace in Lemma 3 m_i by m and ν_{ij} by $1/mn$, $i = 1, \ldots, n$, $j = 1, 2, \ldots, m$, where m and n are positive integers, then we obtain the following

Lemma 4. *If* $F(n, a) = S\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}; a\right)$, *then*

$$F(mn, a) = F(m, a) + \frac{1}{m^{a-1}} F(n, a) = F(n, a) + \frac{1}{n^{a-1}} F(m, a),$$

for every positive integers m, n.

This equality implies

Lemma 5. *If* $a \neq 1$, *then* $F(n, a) = c(a)(1 - 1/n^{a-1})$, *where* $c(a)$ *is a function of the characteristic parameter.*

The tools are now at hand to prove Theorem 1. If n and r_i's are positive integers, $\sum_{i=1}^{m} r_i = n$ and if we put $\mu_i = r_i/n$, $i = 1, 2, \ldots, m$, then an application of Lemma 3 gives

$$S\Big(\underbrace{\frac{1}{n}, \ldots, \frac{1}{n}}_{r_1}, \ldots, \underbrace{\frac{1}{n}, \ldots, \frac{1}{n}}_{r_m}; a\Big) = S(\mu_1, \ldots, \mu_m; a) + \sum_{i=1}^{m} \mu_i^a S\left(\frac{1}{r_i}, \ldots, \frac{1}{r_i}; a\right),$$

or

$$F(n, a) = S(\mu_1, \ldots, \mu_m; a) + \sum_{i=1}^{m} \mu_i^a F(r_i, a)$$

this together with Lemma 5 for $a \neq 1$ implies that

$$S(\mu_1, \ldots, \mu_m; a) = c(a)(1 - 1/n^{a-1}) - \sum_{i=1}^{m} \mu_i^a c(a)(1 - 1/r_i^{a-1}) =$$

$$= c(a)(1 - \sum_{i=1}^{m} \mu_i^a) .$$

In view of axiom a), the later equality holds also for irrational μ_i's. Using axiom b) we get

$$c(a) = \frac{2^{a-1}}{2^{a-1} - 1} .$$

That is, for $a \neq 1$ we have obtained the desired result

$$S(\mu_1, \ldots, \mu_N; a) = \frac{2^{a-1}}{2^{a-1} - 1} (1 - \sum_{i=1}^{N} \mu_i^a) .$$

The equality

$$S(\mu_1, \ldots, \mu_N; 1) = - \sum_{i=1}^{N} \mu_i \log \mu_i$$

is a consequence of the fact that the structural a-entropy is a continuous function of a.

Remark. It is to be noted that the validity of Theorem 1 does not depend ultimately on the assumption of continuity of S in variable a. If this continuity is not required, the proof of Theorem 1 remains unaltered if $a \neq 1$ and for $a = 1$ it can be modified by means of results of [1]. Consequently, the requirement of the continuity mentioned above is not necessary (cf. axiom a)).

In the sequel we list some basic properties of the structural a-entropy.

Theorem 2. $S(\mu_1, \ldots, \mu_N; a)$ *is in the region* $\mu_i \geqq 0,\ i = 1, 2, \ldots, N,\ \sum_{i=1}^{N} \mu_i = 1$ *concave function achieving maximum for* $\mu_i = 1/N,\ i = 1, 2, \ldots, N$.

Proof. Concavity follows from the fact that the matrix of second derivatives of $S(\mu_1, \ldots, \mu_N; a)$ is in the given region negative semidefinite. The proof of the second assertion will be given in the following two steps:

1. Suppose first that $a \neq 1$. As $(2^{a-1} . x^a)/(2^{a-1} - 1)$ is for $0 \leqq x \leqq 1$ convex fucntion, we can write for μ_i under consideration

$$\frac{2^{a-1}}{2^{a-1} - 1} \left(\sum_{i=1}^{N} \frac{1}{N} \mu_i \right)^a \leqq \frac{2^{a-1}}{2^{a-1} - 1} \sum_{i=1}^{N} \frac{1}{N} \mu_i^a ,$$

which yields the desired result.

2. Let now $a = 1$. As $x \log x$ is for $0 \leqq x \leqq 1$ convex function, we can write

$$\left(\sum_{i=1}^{N} \frac{1}{N} \mu_i\right) \log \left(\sum_{i=1}^{N} \frac{1}{N} \mu_i\right) \leqq \sum_{i=1}^{N} \frac{1}{N} \mu_i \log \mu_i$$

and the conclusion of the proof is clear.

The following property of the structural a-entropy seems to be useful for applications:

Theorem 3. *If* $\mu_j \geqq 0, j = 1, \ldots, N, \sum_{j=1}^{N} \mu_j = 1, \mu_{i-1} < \mu_i$ *for* $i = 2, \ldots, N$ *and if* $0 < \varepsilon < (\mu_i - \mu_{i-1})/2$, *then*

$$S(\mu_1, \ldots, \mu_N; a) < S(\mu_1, \ldots, \mu_{i-1} + \varepsilon, \mu_i - \varepsilon, \ldots, \mu_N; a) .$$

Proof. This Theorem obviously follows from Theorem 2.

In closing this paper let us note that the normed measure used in our considerations does not need to be interpreted as a probability measure. The structural a-entropy may be considered as a new generalization of the Shannon's entropy which differs from the generalization given by Rényi [2].

The authors thank for advices and suggestions given to them by Dr. Perez at the consultation about this paper.

(Received June 6th, 1966.)

REFERENCES

[1] Д. К. Фадеев: К понятию энтропии конечной вероятностной схемы. Успехи мат. наук *11* (1956), № 1.

[2] A. Rényi: Wahrscheinlichkeitsrechnung min einen Anhang über Informationstheorie. VEB Deutscher Verlag der Wissenschaften, Berlin 1962.

Kvantifikační metoda klasifikačních procesů

Pojem strukturální α-entropie

JAN HAVRDA, FRANTIŠEK CHARVÁT

Práce je věnována vytvoření jisté kvantifikační metody klasifikačních procesů, přičemž pojem klasifikace je zaveden v definici 1. Problém kvantifikace klasifikace spočívá v axiomatickém zavedení jisté funkce, tzv. strukturální α-entropie na množině všech klasifikací dané množiny s normovanou mírou.

Axiomatické zavedení strukturální α-entropie uvedeným způsobem vede k jednoznačnému určení tvaru strukturální α-entropie. Dále jsou uvedeny základní vlastnosti strukturální α-entropie a ukázána možnost pravděpodobnostní interpretace získaných výsledků, která vede k jistému zobecnění Shannonovy entropie.

Jan Havrda, František Charvát, Katedra matematiky elektrotechnické fakulty ČVUT, Praha 6, Technická 2/1902.