

Jan Havrda; František Charvát

Quantification method of classification processes. Concept of structural -entropy

Kybernetika, Vol. 3 (1967), No. 1, (30)--35

Persistent URL: <http://dml.cz/dmlcz/125526>

Terms of use:

© Institute of Information Theory and Automation AS CR, 1967

Institute of Mathematics of the Academy of Sciences of the Czech Republic provides access to digitized documents strictly for personal use. Each copy of any part of this document must contain these

Terms of use.



This paper has been digitized, optimized for electronic delivery and stamped with digital signature within the project *DML-CZ: The Czech Digital Mathematics Library*
<http://project.dml.cz>

Quantification Method of Classification Processes

Concept of Structural α -Entropy

JAN HAVRDA, FRANTIŠEK CHARVÁT

The aim of this paper is to form a quantificatory theory of classificatory processes. A concept of structural α -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 $\{\mathcal{K}_v\}_{v \in \mathcal{J}}$ be an indexed set of finite families \mathcal{K}_v of propositional functions on B ($\mathcal{K}_v = \{^v p_1, \dots, ^v p_{N_v}\}$, where N_v is a positive integer) such that

$$\bigcup_{i=1}^{N_v} M_i(\mathcal{K}_v) = B, \quad M_i(\mathcal{K}_v) \cap M_j(\mathcal{K}_v) = \emptyset \quad \text{for } i \neq j, \quad i, j = 1, 2, \dots, N_v,$$

and for every $v \in \mathcal{J}$ where $M_i(\mathcal{K}_v) = \{x : x \in B \text{ and } ^v p_i(x) \text{ holds}\}$. The family $\{M_i(\mathcal{K}_v)\}_{i=1}^{N_v}$ the set B , and the family \mathcal{K}_v are said classification, base of classification, and classificatory criteria, respectively.

In the sequel we will denote the classification only by $\mathcal{B}(B) = \{M_i\}_1^N$ 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, consequently, 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 $\mathcal{B}(B)$ we call element of classification: every element $M_i \in \mathcal{B}(B)$ has a measure $\mu(M_i)$, $i = 1, \dots, N$. 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 $\mathcal{B}(B) = \{M_1, \dots, M_N, \mu_1, \dots, \mu_N\}$, where $\mu_i = \mu(M_i)$.

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

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

- a) $S(\mu_1, \dots, \mu_N; a)$ is continuous in the region $\mu_i \geq 0, \sum_{i=1}^N \mu_i = 1, a > 0$;
- b) $S(1; a) = 0, S(\frac{1}{2}, \frac{1}{2}; a) = 1$;
- c) $S(\mu_1, \dots, \mu_{i-1}, 0, \mu_{i+1}, \dots, \mu_N; a) = S(\mu_1, \dots, \mu_{i-1}, \mu_{i+1}, \dots, \mu_N; a)$ for every $i = 1, 2, \dots, N$;
- d) $S(\mu_1, \dots, \mu_{i-1}, v_{i_1}, v_{i_2}, \mu_{i+1}, \dots, \mu_N; a) = S(\mu_1, \dots, \mu_{i-1}, \mu_i, \mu_{i+1}, \dots, \mu_N; a) + \alpha \mu_i^\alpha S\left(\frac{v_{i_1}}{\mu_i}, \frac{v_{i_2}}{\mu_i}; a\right)$ for every $v_{i_1} + v_{i_2} = \mu_i > 0, i = 1, 2, \dots, N, \alpha > 0$.

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

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

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

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)

$$S\left(\frac{1}{2}, \frac{1}{2}; a\right) = S(1; a) + \alpha S\left(\frac{1}{2}, \frac{1}{2}; a\right),$$

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

Lemma 2. If $v_k \geq 0, k = 1, \dots, m, \sum_{k=1}^m v_k = \mu_i > 0$, then

$$S(\mu_1, \dots, \mu_{i-1}, v_1, \dots, v_m, \mu_{i+1}, \dots, \mu_N; a) = S(\mu_1, \dots, \mu_N; a) + \mu_i^\alpha S\left(\frac{v_1}{\mu_i}, \dots, \frac{v_m}{\mu_i}; a\right).$$

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

$$\begin{aligned}
& S(\mu_1, \dots, \mu_{i-1}, v_1, \dots, v_{m+1}, \mu_{i+1}, \dots, \mu_N; a) = \\
& = S(\mu_1, \dots, \mu_{i-1}, v_1, \bar{\mu}, \mu_{i+1}, \dots, \mu_N; a) + \\
& + \bar{\mu}^a S\left(\frac{v_2}{\bar{\mu}}, \dots, \frac{v_{m+1}}{\bar{\mu}}; a\right) = S(\mu_1, \dots, \mu_N; a) + \mu_i^a S\left(\frac{v_1}{\mu_i}, \frac{\bar{\mu}}{\mu_i}; a\right) + \\
& + \bar{\mu}^a S\left(\frac{v_2}{\bar{\mu}}, \dots, \frac{v_{m+1}}{\bar{\mu}}; a\right),
\end{aligned}$$

where $\bar{\mu} = v_2 + \dots + v_{m+1}$. One more application of the induction premise yields

$$S\left(\frac{v_1}{\mu_i}, \dots, \frac{v_{m+1}}{\mu_i}; a\right) = S\left(\frac{v_1}{\mu_i}, \frac{\mu}{\mu_i}; a\right) + \left(\frac{\bar{\mu}}{\mu_i}\right)^a S\left(\frac{v_2}{\bar{\mu}}, \dots, \frac{v_{m+1}}{\bar{\mu}}; a\right)$$

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 $v_{ij} \geq 0$, $j = 1, 2, \dots, m_i$, $\sum_{j=1}^{m_i} v_{ij} = \mu_i > 0$, $i = 1, 2, \dots, n$, $\sum_{i=1}^n \mu_i = 1$, then

$$\begin{aligned}
& S(v_{11}, \dots, v_{1m_1}, \dots, v_{n1}, \dots, v_{nm_n}; a) = \\
& = S(\mu_1, \dots, \mu_n; a) + \sum_{i=1}^n \mu_i^a S\left(\frac{v_{i1}}{\mu_i}, \dots, \frac{v_{im_i}}{\mu_i}; a\right).
\end{aligned}$$

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

Lemma 4. If $F(n, a) = S\left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}; a\right)$, 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 $\mu_i = r_i/n$, $i = 1, 2, \dots, m$, then an application of Lemma 3 gives

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

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

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

$$\begin{aligned} S(\mu_1, \dots, \mu_m; a) &= c(a)(1 - 1/n^{a-1}) - \sum_{i=1}^m \mu_i^a c(a)(1 - 1/r_i^{a-1}) = \\ &= c(a) \left(1 - \sum_{i=1}^m \mu_i^a\right). \end{aligned}$$

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

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

The equality

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

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

Proof. Concavity follows from the fact that the matrix of second derivatives of $S(\mu_1, \dots, \mu_N; a)$ 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 $(2^{a-1} \cdot x^a)/(2^{a-1} - 1)$ is for $0 \leq x \leq 1$ convex function, we can write for μ_i under consideration

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

which yields the desired result.

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

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

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 $\mu_j \geq 0, j = 1, \dots, N, \sum_{j=1}^N \mu_j = 1, \mu_{i-1} < \mu_i$ for $i = 2, \dots, N$ and if $0 < \varepsilon < (\mu_i - \mu_{i-1})/2$, then

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

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 mit einem Anhang über Informationstheorie. VEB Deutscher Verlag der Wissenschaften, Berlin 1962.

Kvantifikační metoda klasifikačních procesů

Pojem strukturální a -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 a -entropie na množině všech klasifikací dané množiny s normovanou mírou.

Axiomatické zavedení strukturální a -entropie uvedeným způsobem vede k jednoznačnému určení tvaru strukturální a -entropie. Dále jsou uvedeny základní vlastnosti strukturální a -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.