Estimation of samples relevance by their histograms.

Mikhail A. Antonets

1. Introduction.

Main object, discussed in the paper is a set $R$ of the samples which are the functions on a finite set $T = \{1, 2, ..., |T|\}$ with values from a finite set $V$. We suppose, that these samples are the result of multiple registration of a dynamic system characteristics that evolve in a mode of operation. In the case, when the function $f$ on the set $T$ was got in a such manner but under conditions, that the mode of operation was not established exactly the following question may be formulated: is the function $f$ relevant to the mode of operation presented by the set of samples $R$.

We suggest an answer to this question using the defined below set $M$ of histograms:

for any function $r$ from the set $R$ its histogram $m$ is the function on the set $V$ defined by the relation

$$m(v) = |\{t: t \in T, r(s) = v\}|, v \in V$$

Any histogram $m$ from the set satisfies $M$ the equality

$$\sum_{v \in V} m(v) = |T|$$

(1)

Our solutions is based on the supporting and covering weights for the set $M$ introduced in [1].

2. Supporting weights and covering weights .

Definition 1. A weight $x$ on the set $V$ is a nonnegative function such that

$$|x| = \sum_{v \in V} x(v) = 1$$

(2)

The set of all weights on the set $V$ will be denoted by $\Delta_V$.

For any functions $x, m$ defined on the set $V$ we set

$$(x, m)_V = \sum_{v \in V} x(v)m(v)$$

Definition 2. A weight $x^M$ on the set $V$ is supporting weight for the set $M$ of the function on the set $V$ if for any weight $x$ on the set $V$ the following inequalities hold

$$\min_{m \in M}(x^M, m)_V \geq \min_{m \in M}(x, m)_V$$

The set of supporting weights $\{x^M\}$ coincides with the set of solutions of the variational problem

$$\min_{m \in M}(x^M, m)_V = \max_{x \in \Delta_V} \min_{m \in M}(x, m)_V$$

(3)

As a quantitative estimation of the degree of relevance, i.e. the correspondence of the tested function $f$ to the criteria for the formation of the set of functions $R$ is based on, we propose the value
where \( \tilde{x}^M \) is the constructed weight and \( m_f \) is histogram of the function \( f \).

**Definition 2.** A weight \( \tilde{x}^M \) on a set \( V \) will be called covering one for a set of functions \( M \) if for any weight \( x \) on the set \( V \) it satisfies the inequality

\[
\max_{m \in M}(x^M, m)_V \leq \max_{m \in M}(\tilde{x}^M, m)_V
\]

The set of covering weights \( \{\tilde{x}^M\} \) coincides with the set of solutions of the variational problem

\[
\max_{m \in M}(x^M, m)_V = \min_{x \in \Delta_V} \max_{m \in M}(x, m)_V
\]

The weight \( \tilde{x}^M \) characterizes the irrelevance of the function \( f \) to the set of functions \( R \) by means of the value

\[
\tilde{s}^M(f) = \sum_{v \in V} m_f(v) \tilde{x}^M(v)
\]

which decreases with the relevance’s growth.

The below-formulated well-known Theorem 1 and Theorem 2 (see [2,3]) states that the discussed variational problems can be reduced to the following linear programming problems:

**Theorem 1.** Any supporting weight \( \bar{x}^M \) is a solution of the following problem:

find all the pairs \( \{\alpha^M, \tilde{x}^M\}, \quad \alpha^M \in \mathbb{R}, \quad \tilde{x}^M \in \Delta_V \), that maximize the value \( \alpha \) under the conditions

\[
x(v) \geq 0, \quad v \in V,
\]

and the inequalities

\[
\alpha - (x, m)_V \leq 0, \quad \forall m \in M
\]

**Theorem 2.** Any covering weight \( \bar{x}^M \) is a solution of the following problem:

Find all the pairs \( \{\bar{\alpha}^M, \bar{x}^M\}, \quad \bar{\alpha}^M \in \mathbb{R}, \quad \bar{x}^M \in \Delta_V \), that minimize the value \( \alpha \) under the conditions

\[
\alpha - (x, m)_V \geq 0, \quad \forall m \in M
\]

Since for the uniformly distributed weight \( x_0(v) \equiv \frac{1}{|V|} \) on the set \( V \) the following equalities holds

\[
\sum_{v \in V} m(v)x_0(v) = \frac{|\Gamma|}{|V|}
\]

then the following inequality hold
\[ \alpha^M \geq \frac{|\mathcal{V}|}{|\mathcal{V}|-1} \]  
\[ \bar{\alpha}^M \leq \frac{|\mathcal{V}|}{|\mathcal{V}|-1} \]  

3. Reduction of dimension of variation problems

**Theorem 3.** For an element \( w \) from the set \( \mathcal{V} \) and an arbitrary finite set \( M \) of nonnegative functions on the set \( \mathcal{V} \) the following statements hold:

1) if for any function \( m \) from the set \( M \) the following inequality take place

\[ m(w) < \frac{1}{|\mathcal{V}|-1}\sum_{v \in \mathcal{V}, v \neq w} m(v) \]  
then for any supporting weight \( \chi^M \) the following equality holds

\[ \chi^M (w) = 0 \]

2) if for any function \( m \) from the set \( M \) the following inequality take place

\[ m(w) > \frac{1}{|\mathcal{V}|-1}\sum_{v \in \mathcal{V}, v \neq w} m(v) \]  
then for any covering weight \( \bar{\chi}^M \) the following equality holds

\[ \bar{\chi}^M (w) = 0 \]

**Proof.** To prove the assertion 1) for given supporting weight \( \chi^M \) let us construct the weight \( \bar{\chi} \), assuming that

\[ \bar{\chi}(w) = 0 \]

and

\[ \bar{\chi}(v) = \chi^M (v) + \frac{\chi^M (w)}{|\mathcal{V}|-1} \]

for any \( v, v \neq w \).

Then

\[ \sum_{v \in \mathcal{V}, v \neq w} \bar{\chi}(v) = 1 \]

and for arbitrary function \( m \) from the set \( M \) the following relations hold

\[ (\bar{\chi}, m)_V = (\chi^M , m)_V - \sum_{v \in \mathcal{V}, v \neq w} m(v)(\chi^M (v) + \frac{\chi^M (w)}{|\mathcal{V}|-1}) - \sum_{v \in \mathcal{V}} \chi^M (v)m(v) = \]

\[ = \chi^M (w)(\sum_{v \in \mathcal{V}, v \neq w} m(v)\frac{1}{|\mathcal{V}|-1} - m(w)) \]
and by virtue inequality (11) the inequality

\[(\bar{x}, m)_V - (\bar{x}^M, m)_V > 0\]  \hspace{1cm} (13)

takes place when

\[\bar{x}^M(w) \neq 0\]

in contradiction with supporting weight definition. The proof of the assertion 2) is similarly. The theorem is proved.

**Corollary.** In the case when the set \(M\) is a set of histogram, generated by functions on the set \(T\) with the values in the set \(V\) there is the equality

\[\frac{1}{|V| - 1} \sum_{v \in V, w \in W} m(v) - m(w) = \frac{1}{|V| - 1} (|T| - m(w)) - m(w) = \frac{1}{|V| - 1} (|T| - |V| m(w))\]

and inequality (11) takes the form

\[m(w) < \frac{|T|}{|V|}\]  \hspace{1cm} (13)

and inequality (12) takes the form

\[m(w) > \frac{|T|}{|V|}\]  \hspace{1cm} (14)

4. **Classification of the set of histograms for** \(V = \{0, 1\}\).

The sets \(R\) of samples with values from the set \{0,1\} gives the most simple examples of the histogram. Any histogram in this case is the pair of natural numbers \(\{m(0), m(1)\}\) satisfying the equality

\[m(0) + m(1) = |T|\]  \hspace{1cm} (15)

The weights \(\bar{x}^M\), \(\bar{x}^M\) and its dual weights \(\bar{x}^\mathcal{V}\), \(\bar{x}^\mathcal{V}\), \(\mathcal{V} = \{\mathcal{v} : \mathcal{v} \in V\}\), where

\[\mathcal{v} : M \rightarrow V, \mathcal{v}(m) = m(\mathcal{v})\]

(see [1]) may be constructed in the explicit form.

**Theorem 4.** For any set of histogram \(M\) generated by the set \(R\) of all functions on the finite set \(T\) and the set of value \(V = \{0, 1\}\) the following assertions are:

1) Let for any histogram \(m\) from the set \(M\) the following inequality hold

\[m(0) > m(1)\]  \hspace{1cm} (16)

then the following equality take place.

\[\bar{a}^M = \min_{m \in M} m(0)\]  \hspace{1cm} (17)
\( x^M = \{1,0\} \) \hspace{1cm} (18)

\( \overline{\alpha}^M = \max_{m \in M} m(1) \) \hspace{1cm} (19)

\( \overline{x}^M = \{0,1\} \) \hspace{1cm} (20)

2) if the set \( M \) contains such histograms \( m', m'' \) that the following inequality take place

\[
\begin{align*}
m'(1) & \geq m'(0) \quad (21) \\
m''(1) & \leq m''(0) \quad (22)
\end{align*}
\]

then the following equalities hold

\[
\begin{align*}
\overline{\alpha}^M & = \overline{\alpha}^M = \frac{|T|}{2} \hspace{1cm} (23) \\
\overline{x}^M & = \overline{x}^M = \left\{ \frac{1}{2}, \frac{1}{2} \right\} \hspace{1cm} (24)
\end{align*}
\]

**Proof.** In the case 1) it follow from the condition (16) that to maximize the sum

\[
m(0)x(0) + m(1)x(1)
\]

for any function \( m \) from the set \( M \) it is necessary put \( x(1) = 0 \) that lead us to equality (17).

In the case 2) by virtue inequalities (21), (22) it follows that the value \( (x, m')_V \) no decrease and the value \( (x, m'')_V \) no increase when the values \( x(1) \) grows. Therefore the equalities (15) and

\[
\left( \left\{ \frac{1}{2}, \frac{1}{2} \right\}, m' \right) = \left( \left\{ \frac{1}{2}, \frac{1}{2} \right\}, m'' \right) = \frac{|T|}{2}
\]

lead to relations (23), (24). The theorem is proved.

The principle of complementary slackness for the weight \( \overline{x}_V^\hat{\beta} \) gives in the case 1) the relation

\[
\sum_{m \in M} m(0)\overline{x}_V^\hat{\beta}(m) = \min_{m \in M} m(0)
\]

That implies that the set of covering weights for the set of functions \( \hat{\nu} \) consist of all weights vanishing outside the set of all functions from the set \( M \) that reach the value \( \min_{m \in M} m(0) \).

Similarly we get that the set of supporting weights for the set of functions \( \hat{\nu} \) consist of all weights vanishing outside the set of all functions from the set \( M \) that reach the value \( \max_{m \in M} m(1) \).

In the case 2) the principle of complementary slackness for the weight \( \overline{x}_V^\hat{\beta} \) gives the relation

\[
\sum_{m \in M} m(j)\overline{x}_V^\hat{\beta}(m) = \frac{|T|}{2} \hspace{1cm} j = 0,1
\]

For any pair functions \( u = \{m', m''\}, m', m'' \in M \) satisfying the condition (21), (22), there exists a solution \( \overline{x}_V^\hat{\beta}, u \) of the system of two linear equations (25), belonging to the set of covering weight for the set \( \hat{\nu} \):
\[ \bar{x}_u^\varphi (m) = 0 \text{ for all } m \text{ distinct from } m', m'' \]  
\[ \bar{x}_u^\varphi (m') = \frac{m''(0) - \frac{1}{2}\overline{T}}{m''(0) - m'(0)} = \frac{\frac{1}{2}|T|-m''(1)}{m'(1) - m''(1)} \]  
\[ \bar{x}_u^\varphi (m'') = \frac{\frac{1}{2}|T|-m'(0)}{m''(0) - m'(0)} = \frac{m'(1) - \frac{1}{2}|T|}{m'(1) - m''(1)} \]

References.

1. Antonets, M.A.; Kogan, G.P. (2016) The variational principle for weights characterizing the relevance. arXiv:1609.01533

2. Dantzig G.B. A proof of the equivalence of the programming and the game problem. Activity Analysis of Production and Allocation, ed. By Koopmans T. C., Cowles Commission Monograph, № 13, New York, Wiley, 1951. p.330-335.

3. Ferguson T.S. Game theory http://www.gametheory.net/books/online.html