INDEX CONSTRUCTION:

the expert-statistical method

V. V. Strijov, V. V. Shakin

Computer Center of the Russian Academy of Science

Abstract

C-2

There are three ways to construct an index. The first is to make index with a method that used measured data with no expert estimations. The second is to make index where experts estimate the feature significance weights. Both methods use feature linear combination for measured data to make the index.

Abstract

C-3

Unlike the mentioned methods the expert estimations concordance technique uses as measures data as well expert estimations of object quality and feature significance weights. According to the technique the experts extricate the contradictions between measured data and expert estimation.

Notation

$\Upsilon = \{\upsilon_1, ..., \upsilon_m\}$ — object set,

$\Psi = \{\psi_1, ..., \psi_n\}$ — feature set,

$A = \{a_{ij}\}_{i,j=1}^{m,n}$ — source data, $A \in \mathbb{R}^{m \times n}$,

$\mathbf{q} = \langle q_1, ..., q_m \rangle^T : \mathbf{q} \in \mathbb{R}^m$ — object indexes,

$\mathbf{w} = \langle w_1, ..., w_n \rangle^T : \mathbf{w} \in \mathbb{R}^n$ — feature weights.

Definition 1

C-5

The index $q_i \in \mathbb{R}^1$ is a scalar, which corresponds to the feature set $\mathbf{a}_{i\cdot}$ of an i-th object $\upsilon_i \in \Upsilon$.

Useful methods

1. Principal components method: $\mathbf{q}_1 = \tilde{A}\Theta_1$, where $\Theta_1 = [\theta_{11}, ..., \theta_{1m}]^T$ is the the first eigenvector of the source data covariance matrix.

2. Singular vectors method: $\mathbf{q}_2 = \lambda_1 U_1$, where $A = U\Lambda V^T$, matrices U and V are real orthogonal matrices and $\Lambda = \{\lambda_i\}$ subject to $\lambda_1 \geqslant ... \geqslant \lambda_r > \lambda_{r+1} = ... = \lambda_n = 0$, and r is rank(A).

3. Pareto slicing: the descriptions $\{\mathbf{a}_{i\cdot}\}$ represented as $T = \bigcup_{\zeta=1}^{l} S_\zeta$ s.t. $S_\zeta \cap S_\eta = \emptyset$, if $\zeta = \eta$, where the Pareto set of ζ-th slice is $S_\zeta = \{\mathbf{a}_{i\cdot} : i \in \{1, ..., m\}\}$ and l is number of slices for the set $\{\mathbf{a}_{i\cdot}\}$. For all $\zeta = 1, ..., l$ the set S_ζ was defined as the set of non-dominated vectors, wich are not in the set $S_{\zeta-1}$, i.e., $S_\zeta = \{\mathbf{a}_{\xi\cdot} : \mathbf{a}_{\xi\cdot} \notin S_{\zeta-1}$, $\forall i \in \{1, ..., m\}, i \neq \xi \exists j \in \{1, ..., n\} : a_{\xi j} \geqslant a_{ij}\}_{\xi=1}^{m}$. Each vector $\mathbf{a}_{\xi\cdot}$, $\xi = 1, ..., m$ corresponds to an index ζ of the set S_ζ, s.t. $\mathbf{a}_{\xi\cdot} \in S_\zeta$. The index $\mathbf{q}_3 = \{\zeta_\xi\}_{\xi=1}^{m}$.

Source data — the triplet $(\mathbf{q}_0, \mathbf{w}_0, A)$

There are vectors, given by experts

$\mathbf{q}_0 = \langle q_{01}, ..., q_{0m} \rangle^T : \mathbf{q}_0 \in \mathbb{R}^m$ и

$\mathbf{w}_0 = \langle w_{01}, ..., w_{0n} \rangle^T : \mathbf{w}_0 \in \mathbb{R}^n.$

There are triplet $\begin{array}{c|c} & \mathbf{w}_0^T \\ \hline \mathbf{q}_0 & A \end{array}.$

Expert disagreement measure

Linear mapping operators A and A^{+} defines

$\mathbf{q}_1 = A\mathbf{w}_0$ — forward mapping and

$\mathbf{w}_1 = A^{+}\mathbf{q}_0$ — backward mapping,

where A^{+} — operator, pseudo inverse for A.

In general case $\mathbf{q}_1 \neq \mathbf{q}_0$ и $\mathbf{w}_1 \neq \mathbf{w}_0$.

We use Eucledian distance $\|\mathbf{q}_0 - \mathbf{q}_1\|$ и $\|\mathbf{w}_0 - \mathbf{w}_1\|$ as an expert disagreement measure.

Definition 2

C-10

Concorded values of index and feature weights are values $\hat{\mathbf{q}}$ and $\hat{\mathbf{w}}$, so that the next condition is fare:

$$\begin{cases} \hat{\mathbf{q}} = A\hat{\mathbf{w}}, \\ \hat{\mathbf{w}} = A^{+}\hat{\mathbf{q}}. \end{cases} \tag{1}$$

Definition 3

Concordance operator Φ of expert estimations is operator, which maps the triplet $(\mathbf{q}_0, \mathbf{w}_0, A)$ to the triplet $(\hat{\mathbf{q}}, \hat{\mathbf{w}}, A)$, where the vectors $\hat{\mathbf{q}}, \hat{\mathbf{w}}$ meet the condidion (1):

$$\Phi : (\mathbf{q}_0, \mathbf{w}_0, A) \longrightarrow (\hat{\mathbf{q}}, \hat{\mathbf{w}}, A).$$

α-concordance

Let $A : W \longrightarrow Q$, $W \ni \mathbf{w}_0$, $Q \ni \mathbf{q}_0$ and for A there exists pseudoinverse operator

$A^+ : Q \longrightarrow W$, i.e, $A^+A = I_n$,

$AA^+ = I_m$, $A^+AA^+ = A^+$, $AA^+A = A$.

Theorem

There are exists a singular value decomposition $A = U\Lambda V^T$ of a matrix A, where $\Lambda = \mathrm{diag}(\lambda_1, ..., \lambda_R)$, $R = \min(m, n)$, и $U^T U = I_m, VV^T = I_n$.

The matix $A^+ = V^T \Lambda^{-1} U$ is pseudoinverse for the matrix A.

Assume A^+ as $A^+ = V\Lambda_r^{-1} U^T$, where $\Lambda_r^{-1} = \mathrm{diag}(\lambda_1^{-1}, ..., \lambda_r^{-1}, 0, ..., 0)$.

Concordance procedure

Assume $\mathbf{q}_1 = A\mathbf{w}_0$ and $\mathbf{w}_1 = A^+\mathbf{q}_0$. Assign $[\mathbf{q}_1, \mathbf{q}_0] \subset Q$ and $[\mathbf{w}_1, \mathbf{w}_0] \subset W$. Let the concorded estimations are in the segments. Find the linear convex combinations $\mathbf{q}_0, \mathbf{q}_1$, and $\mathbf{w}_0, \mathbf{w}_1$ as

$$\begin{aligned} \{\mathbf{w}_\alpha : \mathbf{w}_\alpha &= (1-\alpha)\mathbf{w}_0 + \alpha A\mathbf{w}_0\} \in [\mathbf{w}_0, \mathbf{w}_1], \\ \{\mathbf{q}_\alpha : \mathbf{q}_\alpha &= \alpha\mathbf{q}_0 + (1-\alpha)A^+\mathbf{q}_0\} \in [\mathbf{q}_0, \mathbf{q}_1], \end{aligned} \tag{2}$$

where $\alpha \in [0, 1]$.

How to define α?

$$\varepsilon^2 = \|\mathbf{q}_\alpha - \mathbf{q}_0\|^2,$$
$$\delta^2 = \|\mathbf{w}_\alpha - \mathbf{w}_0\|^2.$$

As a criterium of parameter α let us chose the minimal distance between the initial and concorded expert estimations in both spaces Q and W.

$$\frac{\varepsilon^2}{m-1} = \frac{\delta^2}{n-1}. \quad (3)$$

How to modify α?

The obtained results can be represented and proposed to experts to discuss in the following way:

init		$\mathbf{w}_0^T$
	fin	$\mathbf{w}_\alpha^T$
$\mathbf{q}_0$	$\mathbf{q}_\alpha$	A

Summary

As a result we have: first, precise valid indices. Second, we have the reasoned expert estimations; we know why expert valued an object and what contribution a feature makes to index. And we have weights to make future indices by using "non-expert"methods.

The methods with or without expert involvement were used for solution different economical, sociological and ecological problems. Here are some of them.

Indexes

1. Russian nature protected areas management effectiveness evaluation.

2. Integral indicator for quality of life in Russian regions.

3. Human Development Index in Russia.

4. Kyoto-index of Ohio power plants, USA. Kyoto-index also is a measure of any industrial system ecological footprint.

The information

C-19

http://cs.ru/cito/index