Aplinkos tyrimai, inžinerija ir vadyba, 2003.Nr.4(26), P.51-55 *ISSN 1392-1649*
Environmental research, engineering and management, 2003.No.4(26), P.51-55

Index Construction: the Expert-Statistical Method

Vadim Strijov, Vsevolod Shakin
Dorodnicyn Computing Centre of the Russian Academy of Sciences

(received in November, 2003; accepted in December, 2003)

The paper deals with the index construction and presents a new technique that involves expert estimations of object indices as well as feature significance weights. The method based on a linear indexing model for a set of objects index is calculated as a linear combination of the object feature descriptions. Well-known methods of index construction with "no teacher" are overviewed to give a comparison with the new method. Experts can take part in the index calculation and verify the results, which are: the first, precise valid indices and the second, we have the reasoned expert estimations. The methods with or without expert involvement were used for solution of listed below different economical, sociological, and ecological problems. The paper is based on the report, presented in OECD seminar, held in Paris, March 5-6, 1996.
Keywords: *indicator method, expert estimation, sustainable development.*

1. Introduction

The indexing technique is based on data reduction and expert-statistical methods. The term was introduced by S. A. Aivazian in 1974. "When we are trying to estimate, in general, production effectiveness of a separate employee, a division or an enterprise (for example, estimation criteria can be the quality of live, development index), estimate quality of a sportsman or sporting team in game sports, every time we solve (often intuitional) the same problem: taking as a initial condition a set of the separate features, where each feature can be measured and describes a separate part of effectiveness criteria, we are assigning weights of a feature significance (the weight of feature influence on general aggregate effectiveness conception) and so we have some scalar, general effectiveness index" [1].

Hereby, idea to construct an index, or integral indicator for a set of objects as a linear combination of the object feature descriptions was proposed. Also we can list methods for index making such as Principle Components, Factor Analysis, Extreme Feature Grouping, Multidimensional Scaling, Maximal Informative Features Selection, et cetera. Those methods use a measured description of objects as information for indexing. Also there are methods that use expert estimations: Pair-wise Comparison, Fussy Relations, and Linguistic Scaling.

On the other hand, V. V. Shakin introduced a method for jury estimations objectification in 1972. The main principle of the method is in the duality of expert estimations: experts can estimate both quality of objects and feature significance weights. The proposed method develops that principle estimations to the expert is concordance technique.

So, there are two 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.

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

Experts play important role in the indexing and set an index criteria, approve the set of comparable objects, observe set of feature descriptions, put estimations. We assume that an expert that takes part in index creating has his own opinion and the opinion

is not biased by public opinion. The expert opinion must result the expert experience and skills. Experts put their opinions and estimations to specially prepared questionnaires. All the questionnaires were created to give experts maximal freedom in expression.

As the result of the expert estimations on concordance technique application we have validated indexes. Also we have the reasoned expert estimations and feature weights to make indexes in future with no expert involvement.

2. The problem

Let us consider a set of objects. One has to compare objects to know what objects are better and what objects are worse. A rating of the objects is the result of the comparison. That is a list of objects sorted according to their quality, or the result is an index set where each index points to each object' quality. Index or integral indicator is a general characteristic of an object (value of object quality, effectiveness, or preference). Usually index is a positive number that is calculated by means of mathematical methods, or models and description of the objects.

To construct an index one has to collect data about the every object. For that purpose one needs to assume a set of features that describe the objects according to some criteria. A criterion exposes principle that joins all the features and shows the goal of index making. After data acquisition we have a table "object-feature" where the rows represent objects and columns represent features.

We can make the required index using the table. For that one chooses one of mathematical methods such as Singular vectors, Principle components, Pareto Slicing, and the others. The methods form index as precise (linear) estimations, object ordering (rating), rough or linguistics estimations. The mentioned methods are also called as "non-expert" methods.

Let us consider a given set $\mathrm{Y} = \{\upsilon_1,...,\upsilon_m\}$ of objects and set $\Psi = \{\psi_1,...,\psi_n\}$ of features. An object υ_i is described by a row vector $\mathbf{a}_{i\cdot} = \langle a_{i1},...,a_{in}\rangle$, $\mathbf{a}_{i\cdot} \in \mathbf{R}^n$. The object set description is represented as a source data matrix $A = \{a_{ij}\}_{i,j=1}^{m,n}$.

Definition 1. *An object υ_i, which has maximal index value (or maximal value of an expert estimation in case the estimation is the index), $q_i = \max\{q_1,...,q_m\}$ is the best. An object υ_i, which has minimal index value $q_i = \min\{q_1,...,q_m\}$ is the worst.*

Definition 2. *A feature ψ_j, which has maximal weight (or maximal value of an expert estimation in case the estimation is the weight), $w_j = \max\{w_1,...,w_n\}$ is the most important for index construction algorithm.*

The next conditions are stated. A maximal value item $a_{\xi\zeta}$ of a feature ψ_ζ with the number ζ means the ξ-th object υ_ξ is the best in case when the feature is one. Similar, a minimal value item $a_{\xi\zeta}$ of a feature ψ_ζ means the object υ_ξ is the worst in the case when the feature is one,

$$\begin{aligned} a_{\xi\zeta} &= \max\{a_{i\zeta}\}_{i=1}^{m} \Rightarrow q_\xi = \max\{q_1,...,q_m\}, \\ a_{\eta\vartheta} &= \min\{a_{i\vartheta}\}_{\vartheta=1}^{m} \Rightarrow q_\eta = \min\{q_1,...,q_m\}. \end{aligned} \quad (1)$$

Vectors $\mathbf{a}_{\cdot j} = \langle a_{1j},...,a_{mj}\rangle^T$, $\mathbf{a}_{\cdot j} \in \mathbf{R}^m$ are standardized so that the next equation is fair:

$$a_{ij} = 1 - \frac{|\, a_{ij} - a_j^{opt}\,|}{\max\{(a_j^{opt} - \min\{a_{\cdot j}\}),(\max\{a_{\cdot j}\} - a_j^{opt})\}}, \quad (2)$$

$i = 1,...,m, j = 1,...,n$, where the optimal values a_j^{opt} are given.

3. Making index with "no teacher"

There are several methods to find object indices with "no teacher": principal components method, singular vectors method and Pareto slicing. Very often the weighted sum $\mathbf{q} = A\mathbf{w}_0$, where the weights $\mathbf{w}_0 = \langle w_{01},...,w_{0n}\rangle^T, \mathbf{w}_0 \in R^n$ are defined by experts is used to make indices.

Definition 3. *The index $q_i \in \mathbf{R}^1$ is a scalar, and corresponds to feature set $\mathbf{a}_{i\cdot}$ of an i-th object υ_i.*

On exploring the object set Y, the vector $\mathbf{q} = \langle q_1,...,q_m\rangle^T, \mathbf{q} \in \mathbf{R}^m$ is considered as an index of object set, described by the matrix $A \in \mathbf{R}^{m\times n}$.

Principal components method. To find the first principal component [2,3] of standardized (2) and centered matrix $\widetilde{A}$ the first eigenvector $\Theta_1 = [\theta_{11},...,\theta_{1m}]^T$ of its covariance matrix must be found. The index is calculated as $\mathbf{q}_1 = \widetilde{A}\Theta_1$.

Singular vectors method. The index $\mathbf{q}_2$ calculation procedure using the singular value decomposition is the following. The source matrix A can be represented as $A = U\Lambda V^T$, where U and V are real orthogonal matrices. Λ is diagonal matrix subject to $\lambda_1 \geq ... \geq \lambda_r \geq \lambda_{r+1} = \lambda_n = 0$, where r is rank (A).

The first singular vector is used to make an index $\mathbf{q}_2 = U_1\lambda_1$; this vector corresponds the maximal singular number λ_1.

The Pareto slicing [4]. The descriptions $\mathbf{a}_{i.}$ of the object set Y were represented as $T = \bigcup_{\zeta=1}^{l} S_\zeta$ subject to $S_\zeta \cap S_\eta = \varnothing$ if $\zeta \neq \eta$, where the Pareto set of ζ-th slice is $S_\zeta = \{\mathbf{a}_{i.} : i \in \{1,...,m\}\}$ and l is the number of slices for the set $\{\mathbf{a}_{i.}\}$.

A vector $\mathbf{a}_{\xi.} = \langle a_{\xi 1},...,a_{\xi n}\rangle$ is non-dominated if there is no vector $\mathbf{a}_{i.}$, so that $a_{ij} > a_{\xi j}, i = 1,...,m$ and $j = 1,...,n$.

For all $\zeta = 1,...,l$ the set S_ζ was defined as the set of non-dominated vectors, which are not in the set $S_{\zeta-1}$. Each vector $\mathbf{a}_{\xi.}, \xi = 1,...,m$ corresponds to the index ζ of the set S_ζ, subject to $\mathbf{a}_{\xi.} \in S_\zeta$. The obtained set of slice indices $\Xi = \{\zeta_\xi\}_{\xi=1}^{m}$ must meet the condition (1), so the index $\mathbf{q}_3 = \{\max(\Xi) - \zeta_\xi\}_{\xi=1}^{m}$.

4. Concordance procedure

Often experts that have their own opinions in particular applications want to assign indices or expert estimations. There are three types of the estimations. The first is when the expert can compare objects or say which object is better and which object is worse. The second is that the expert can assign the estimations in the convenient scale that is to say more precise than previous case. That is so called ranking. The third type is when the expert can set the estimations with high precision degree. Such estimations are called the linear-scaled.

Expert estimations can prove or disprove the indices that were results of "non-expert" methods. There is a reason to compare these two results such as objects' expert estimations and indices, and correct or specify them. Such correction or specification composes expert estimations on concordance technique.

Like the objects estimation procedure, the expert can assign feature significance weights. Using the weights and measured data we can compute the objects' indices. And vice versus: using the objects' estimations and data we can compute weights. The proposed technique to solve the problem is to concord expert estimations, indices, and weights.

Each object υ_i corresponds to an expert estimation q_{0i}, also, each feature ψ_j corresponds to expert estimation w_{0j}, in the other words, there are given vectors $\mathbf{q}_0 = \langle q_{01},...,q_{0m}\rangle^T$ and $w_0 = \langle w_{01},...,w_{0n}\rangle^T$.

A triplet $(\mathbf{q}_0, \mathbf{w}_0, A)$ is represented as a table, where each item of the vector $\mathbf{q}_0$ corresponds to a row and each item of the vector $\mathbf{w}_0$ corresponds to a column of the matrix A:

	$\mathbf{w}_0^T$
$\mathbf{q}_0$	A

In general case the expert estimation vector $\mathbf{q}_0$ and vector of object' feature values weighted sum $A\mathbf{w}_0$ are different, $\mathbf{q}_0 \neq A\mathbf{w}_0$, also, $\mathbf{w}_0 \neq A^+\mathbf{q}_0$, where A is the linear mapping operator, represented by the table, and let A^+ be mapping operator, pseudoinverse [5] for A exists.

Definition 4. *Concorded values of index and feature weights are values* $\hat{\mathbf{q}}$ *and* $\hat{\mathbf{w}}$*, so that the next condition is fare:*

$$\begin{cases} \hat{\mathbf{q}} = A\hat{\mathbf{w}}, \\ \hat{\mathbf{w}} = A^+\hat{\mathbf{q}}. \end{cases} \qquad (3)$$

Definition 5. Concordance operator Φ of expert estimations is operator, which maps the initial triplet $(\mathbf{q}_0, \mathbf{w}_0, A)$ to the concorded triplet $(\hat{\mathbf{q}}, \hat{\mathbf{w}}, A)$, where vectors $\hat{\mathbf{q}}$, $\hat{\mathbf{w}}$ meet the condidion (3): $\Phi : (\mathbf{q}_0, \mathbf{w}_0, A) \to (\hat{\mathbf{q}}, \hat{\mathbf{w}}, A)$.

The concordance operator was defined in the next way. Linear mapping operator matrix A was given. It maps the feature weights' space W, $\mathbf{w}_0 \in W$, to the object indexes' space Q, $\mathbf{q}_0 \in Q$, $A : W \to Q$ and pseudoinverse operator A^+ which maps object indexes' space to feature weights' space $A^+ : Q \to W$.

There exists a singular value decomposition $A = U\Lambda V^T$ of a matrix A, and the next statement is fare. *The matix* $A^+ = V\Lambda^{-1}U^T$ *is pseudoinverse for the matrix* A.

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

For each source values (they are expert stated) of index vector $\mathbf{q}_0$ let $\mathbf{w}_1 = A^+\mathbf{q}_0$ and $\mathbf{q}_1 = A^+\mathbf{w}_0$. So, the segments $[\mathbf{q}_1, \mathbf{q}_0]$ and $[\mathbf{w}_1, \mathbf{w}_0]$ are given. Euclidian distances $\|\mathbf{q}_1 - \mathbf{q}_0\|$ and $\|\mathbf{w}_1 - \mathbf{w}_0\|$ were used as measure of expert estimations inconcordance. One easily can find the concorded estimations, that lie on the segments. The convex linear combinations of the vectors $\mathbf{q}_1, \mathbf{q}_0$ and $\mathbf{w}_1, \mathbf{w}_0$ were represented as two sets

$$\{\mathbf{w}_\alpha : \mathbf{w}_\alpha = (1-\alpha)\mathbf{w}_0 + \alpha\mathbf{w}_1\} \in [\mathbf{w}_1, \mathbf{w}_0],$$
$$\{\mathbf{q}_\beta : \mathbf{q}_\beta = (1-\beta)\mathbf{q}_0 + \beta\mathbf{q}_1\} \in [\mathbf{q}_1, \mathbf{q}_0],$$

where $\alpha, \beta \in [0,1]$. The next statement: *for any values* $\alpha, \beta \in [0,1]$, *the vector values* $\mathbf{w}_\alpha, \mathbf{q}_\beta$ *meet concordance conditions (3) and* $\alpha = 1 - \beta$ is fare.

In that way the concorded expert estimation can be calculated using the next equation

$$\begin{aligned} \mathbf{w}_\alpha &= (1-\alpha)\mathbf{w}_0 + \alpha A^+ \mathbf{q}_0, \\ \mathbf{q}_\alpha &= \alpha \mathbf{q}_0 + (1-\alpha) A \mathbf{w}_0, \end{aligned} \qquad (4)$$

where $\alpha \in [0,1]$ is the object indices expert estimations versus feature weights expert estimations importance parameter. On value $\alpha = 0$ object indices expert estimations will be ignored and feature weights expert estimations will be considered; on value $\alpha = 1$ feature weights expert estimations will be ignored and object indices expert estimations will be considered.

The triplet $(\mathbf{q}_\alpha, \mathbf{w}_\alpha, A)$, *which is obtained by the concordance procedure (4) meets concordance requirements (3).*

That is, using the equation (4), one can choose the parameter α, which defines concorded values $\mathbf{q}_\alpha = A\mathbf{w}_\alpha$. On fixed value α one can easily calculate the error: Euclidean distance between source vectors and obtained vectors in integral indicators' space and weight's space are equal to

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

The condition of minimal distance between the initial and concorded expert estimations in both spaces Q and W was chosen as a criterion of parameter α choice. Since dimensions the spaces are equal to m and n correspondingly, one can standardize squared distances and find concorded vector values $\mathbf{q}_\alpha$ and $\mathbf{w}_\alpha$, so that they meet the next condition: $\frac{\varepsilon^2}{m-1} = \frac{\delta^2}{n-1}$.

Usually experts have to choose the α parameter value considering that it depends on expert object estimations versus features estimations preferences. The obtained results can be represented and proposed to experts to discuss in the following way:

Initial values		$\mathbf{w}_0^T$
	Final values	$\mathbf{w}_\alpha^T$
$\mathbf{q}_0$	$\mathbf{q}_\alpha$	A

Often during the discussion the value α was changed and in that case of the concordance procedure, described above, repeated and then, the newly obtained results were proposed to the next discussion.

5. Conclusions

Many index construction algorithms supposed that the feature weights are estimated with precision, which enough to obtain results, adequate from experts' point of view. In practice, there is a difficult problem to assign the feature weights. The expert estimations concordance approach was proposed in this paper. Experts assign the initial object indexes' estimation and the features weights' estimation and then make the estimations non-contradicted using the concordance procedure.

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 of different economical, sociological, and ecological problems. There are some of them:

1. Russian nature protected areas management effectiveness evaluation. Annually, all the state nature protected areas make report on main activities: reservation, science, educational. There are 101 nature-protected areas in Russia and report contents 139 features. To check report adequacy the experts were involved. The experts point to more important features and estimates integral indicators. The adequate, for the expert opinion, indices were computed.
2. Integral indicator for quality of life in Russian regions. The State Statistics Committee collects data for 76 Russian regions to make hierarchic integral indicator. The indicator included wealth rate, people quality, public health service quality and ecology potential for each region. The integral indicator is based on more than 80 features.
3. Human Development Index in Russia was developed according to UNO development program. Using an expert who assigned the human development index order for each of 73 regions, the index model was reconstructed and feature weights were founded.
4. Kyoto-index is indented to evaluate ecological footprints of power plants. The index was based on an expert model and was used for model assessment of Ohio power plants, USA. Kyoto-index also is a measure of any industrial system ecological footprint.

Literature

1. Aivazian S. A., Mkhitariyan V. S. Applied statistics and essential econometrics. Moscow: UNITI, 1998. P. 111.
2. Rao, C. R. Linear statistical inference and its applications. John Wiley & Sons, 1965. P.530-533.

3. Aivazian S. A., et. al. Applied Statistics. Classification and space reduction. Moscow: Finansy i statistica, 1989. P. 334.
4. Shakin V. V. Pareto slicing of finite sets. Multivariate statistical analysis application in the economics and quality estimation. V-th scientific conference CIF. Abstracts. Moscow: CEMI RAS, 1993. P.96-97.
5. Golub, G., Van Loan, C. Matrix computation. John Hopkins Series in the Mathematical Sciences 1999. P. 223.

Vsevolod V. Shakin, Dorodnitsyn Computer Centre of the Russian Academy of Sciences.
Address: Vavilov st. 40 Moscow 119991 Russia
Tel: +7(095)1354163
Fax: +7(095)1356159
E-mail: shakin@ccas.ru

Vadim Strijov, Dorodnitsyn Computer Centre of the Russian Academy of Sciences.
Address: Vavilov st. 40 Moscow 119991 Russia
Tel: +7(095)1354163
Fax: +7(095)1356159
E-mail: strijov@ccas.ru

Indeksų konstravimas: ekspertinė – statistinė technologija

Vadim Strijov, Vsevolod Šakin
Rusijos Mokslų akademijos Dorodnicyno skaičiavimo centras

(gauta 2003 m. lapkričio mėn.; atiduota spaudai 2003 m. gruodžio mėn.)

Indeksai gali būti kuriami įvairiais būdais. Tačiau, parinkus algoritmus ir gavus tam tikras išvadas, kyla klausimas, koks apskaičiuoto indekso adekvatumas. Galima suabejoti, kiek yra teisingi ekspertų vertinimai. Aptarsime tris indeksų konstravimo būdus. Pirmuoju atveju indeksas sudaromas remiantis duomenimis be eksperterinio įvertinimo. Antruoju sukuriamas indeksas, kuriame ekspertai įvertina nagrinėjamų požymių reikšmingumo svorius. Konstruojant indeksą abu metodai naudoja požymių tiesinius darinius. Skirtingai negu minėti būdai, ekspertinių vertinimų suderinimo technika kaip matą naudoja duomenis, objekto kokybės ekspertinius vertinimus ir požymių reikšmingumo svorius. Ekspertai pašalina matavimo rezultatų ir ekspertinių įvertinimų prieštaravimus, įvertindami „svorių versus objekto“ reikšmingumą pasirinktam duomenų modeliui. Gautas rezultatas – tikslūs indeksai. Be to , turime pagrįstus ekspertų vertinimus. Taip pat gauname svorius, kurie gali būti panaudoti kuriant „neekspertinius“ metodus.

Sukurta technologija panaudota sprendžiant įvairias ekonomines, socialines ir ekologines problemas: Rusijos gamtos apsaugos efektyvumui įvertinti ir integruotam gyvenimo kokybės Rusijos regionuose, žmogaus vystymo indeksui Rusijoje bei JAV jėgainių Kioto indeksui sudaryti.