APPLICATION OF PRINCIPAL COMPONENT ANALYSIS FOR OPTIMIZATION OF A SYSTEM OPERATION ASSESSMENT CRITERIA SET

Łukasz Muślewski

University of Technology and Life Sciences
Machine Maintenance Department
Prof. S. Kaliskiego Street 7, 85-789 Bydgoszcz, Poland
tel.: +48 523408723
e-mail: lukasz.muslewski@utp.edu.pl

Leszek Knopik

University of Technology and Life Sciences
Faculty of Management
Fordońska Street 430, 85-790 Bydgoszcz, Poland
tel.: +48 523408228
e-mail: knopikl@utp.edu.pl

Abstract

The analysis of results of experimental tests and a literature survey of the issue reveal that the subject matter connected with determination of the optimal number of a given transportation system operation assessment criteria has a direct influence on the result of the considered assessment. The analysis was made within the research on operation quality of a selected transportation system. In order to optimize the analysed system assessment criteria, a theoretical description has been made and an example of canonical correlations application has been presented. Analysis of canonical correlations involves determination of linear combination parameters of the studied sets so that the obtained correlation coefficient will have maximal value. In the successive step, the next pair of linear combination are not correlated with the combinations determined in the first step. The determined correlations can measure the power of relation between two sets of variables and are useful in the process of choosing significant criteria for a given research object operation quality assessment.

Keywords: *quality, criteria, transport system, analysis, canonical correlations , optimization*

1. The research object

All the considerations refer to operation quality of complex transport systems, especially those ones that carry out passenger and freight transport tasks by water, land or air. The main goal of such systems operation is to provide transport services with the use of technical objects, in a given environment, quantity, time and under the influence of given environmental factors. Thus, providing the object with required operation quality and its assessment in terms of safety, efficiency reliability, availability, including the economic factor are of key importance for the operation process. The studied transport systems belong to a group of sociotechnical systems of the type H-M-E (human- machine- environment) in which their operation quality assessment is made depending on changes of values of features describing actions of operators, technical objects controlled by them, and the impact of the environment [2].

On the basis of identification and an analysis of real transport systems it was established that at particular levels of their decomposition, there can be distinguished the following subsystems:

– logistic, including actions connected with the system management, information flow and

processing are performed as well as maintenance serviceability of transport means used in the system and this subsystem consists of:
  - decision making subsystem,
  - traffic continuity maintenance subsystem,
  - information subsystem,
- executive, whose main goal is, to provide transport services,
- environment – a synergy subsystem [3].

2. The system operation quality

This section contains a description of rules, on the basis of which the system operation quality assessment method has been formulated with special emphasis on municipal systems of public transport.

On the basis of literature and the author's own research it has been defined that: the system operation quality is a set of its features expressed by means of their numerical values in a given time t, defining the fulfilment degree of the set requirements [9].

It was assumed that the evaluator establishes a set of criteria for assessment of system K operation quality. Next, the research object is identified and on this basis, with reference to the established criteria, a set of features – X, describing the system in terms of its operation quality, is determined.

Assessment made in this way depends on determination of its criteria, that is, requirements set by outside observers (users, decision makers, operators, maintenance workers), with the assumption that it makes sense when:

$$K_1(t) \cup K_2(t) \cup \ldots \cup K_{n-1}(t) \in True, \quad (1)$$

where:
$K_i(t)$ – logical variable,
0 – if the i-th criterion has not been accepted,
1 – if it has been accepted.

The assessment process involves monitoring whether and to what degree particular features fulfill the established K criteria. Evaluation is performed on the basis of the features values measured in time t (measurable features) or states in which they are in a given time t (immeasurable features), through assigning appropriate identifiers to them. In connection with this, the level of the system operation quality in given time t determines a set of values of significant features $\{X_i\}$ i=1,2,…,p, accepted for its description, from an established point of view.

3 Model of quality assessment of transport operation systems functioning

The system model signifies such a system to be devised or implemented which reflecting or reconstructing the research object is capable of replacing it in such a way that upon being examined it provides new information on this object [3, 4]. It is assumed that. A model should aim at distinguishing significant, variable features of the examined phenomena and processes, neglecting others. Division into significant and insignificant variables depends largely on the researchers, their knowledge, possibilities of calculation and measurement and the accepted by them methods, tools and research techniques.

Defining the fulfilment degree of set requirements-criteria provides the basis for evaluation of a given transport system operation quality. Condition justifying acceptance of given criterion is dependent on whether its fulfilment degree can be checked by at least one of describing it (significant, variable, measurable, non-correlated) features. Thus, the general, criteria based assessment model is described by dependence 2:

$$K_1(t) =< X_1(t), X_2(t), \ldots, X_{K_1}(t) >,$$
$$K_2(t) =< X_{K_1+1}(t), X_{K_1+2}(t), \ldots, X_{K_2}(t) >,$$
$$\ldots$$
$$K_i(t) =< X_{K_i+1}(t), X_{K_i+2}(t), \ldots, X_{K_{i+1}}(t) >, \tag{2}$$
$$\ldots$$
$$K_{n-1}(t) =< X_{K_{n-1}}(t), X_{K_{n-1}+1}(t), \ldots, X_n(t) >.$$

Thus, for a random i-th criterion the condition of non-emptiness needs to be satisfied – condition of existence of a set of criteria described by dependence 3:

$$\Lambda_{i\in\{1,2,\ldots,n-1\}} K_{i+1} - K_i \geq 1. \tag{3}$$

Identification of criteria set K_i, is the main problem to be addressed at the beginning of the research (as it determines correctness of the assessment process), therefore, further in this study, an example of canonical correlations application is presented as one of the tools supporting the choice of the most significant criteria, being input data to be used for building a resultant model for a given system operation quality assessment [5, 6].

4. Canonical correlations

The dependence between two sets of variables X = $\{X_1, X_2, \ldots, X_p\}$ and Y = $\{Y_1, Y_2, \ldots, Y_q\}$ is frequently investigated in empirical tests. Multi factor regression from variables X_1, $X_2, \ldots, X_3$ is discussed separately for each variable Y_k due to loss of information about the relations that occur in set $\{Y_1, Y_2, \ldots, Y_q\}$. The analysis of canonical correlations involves determination of coefficients of linear combinations of sets X and Y so as to make the correlation coefficient maximal. In the next step, another pair of linear combinations with maximal correlation coefficient is determined, on condition that the determined linear combinations are not correlated with the combinations determined in the first step. The determined correlations measure power of the relation between two sets of variables.

4.1 Testing the significance of canonical correlations

The test of canonical correlations significance is a sequential procedure. The first verified hypothesis is a statistical hypothesis assuming that all correlations are insignificant. The alternative hypothesis assumes that, at least one correlation, is significant [1,7,8].

We assume that random variables X i Y have normal distributions $N_p(m_1, \sum_{11})$ i $N_p(m_2, \sum_{22})$. Let:

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}, \tag{4}$$

where:

$\sum_{11}$ – is correlation matrix of random variable X,
$\sum_{22}$ – correlation matrix of random variable Y,
$\sum_{12}$ – correlation matrix between variables X i Y,
$\sum_{21}$ – matrix transposed to matrix $\sum_{12}$, $\sum_{12} = \sum_{21}^T$.

Matrix $\sum_{11}$ is a matrix of p order, $\sum_{22}$ matrix of q order, matrix $\sum_{12}$ is of [p x q] dimension. Squares of canonical correlation $\lambda^2_1 \geq \lambda^2_2 \geq \ldots \geq \lambda^2_q$ are determined from determination equation of the following form:

$$\det(\Sigma_{22}^{-1} \Sigma_{21} \Sigma_{11}^{-1} \Sigma_{12} - \lambda^2 E) = 0, \tag{5}$$

where E - is a unit matrix of q order.

Canonical correlations significance tests are performed by means of Bartlett test. Let:

$$A_0 = \prod_{i=1}^{q} (1 - \lambda_i^2), \tag{6}$$

For big tests, verification of the zero hypothesis, which assumes that all canonical correlations are equal to 0, can be performed by means of statistics T. Where:

$$T = -(n - 1 - (p + q + 1)/2)\ln(\Lambda_0). \tag{7}$$

With the assumption of the hypothesis rightness, random variable T has approximately distribution χ^2 with freedom degrees pq.

After rejection of the zero hypothesis, significance q – k of the smallest canonical correlations is tested.

In a general case, when the hypothesis is rejected for k canonical correlation, verification of the fact that remaining q-k are equal, is checked by means of statistics.

$$A_0 = \prod_{i=k}^{q}(1 - \lambda_i^2) \tag{8}$$

Testing statistics:

$$T = -(n - 1 - (p + q + 1)/2)\ln(\Lambda_k), \tag{9}$$

for the assumption of the hypothesis rightness, has distribution χ^2 with (p - k)(q – k) freedom degrees. The testing procedure is stopped when there is no zero hypothesis rejection.

4.2. Analysis of experimental data

Statistical analysis was applied to set n=150 of vectors of data concerning service quality of a public transportation system. Each of the analysed vectors has 16 components Ki, which account for: safety, efficiency, availability, operational readiness, ergonomics, environment friendliness, usability, atmospheric factors, damageability, availability, esthetics, informativeness, timeliness, reliability, time of transport service accomplishment, external factors, damageability, reliability and cost-efficiency [5, 6].

The following set is divided into two separate sets: X = {1, 2, 7, 8, 10, 11, 12, 13, 16} and Y = {3, 4, 5, 6, 9, 14, 15}. In set Y, there are q informativeness criteria, which directly apply to technical objects, used in the analysed research object.

In connection with the above, the size of sets X and Y are respectively: p = 9 i q = 7 for the analysed set n= 150.

Table 1, shows tests results of canonical correlations significance with their coefficients values in the order from the smallest to the largest one.

Tab. 1. Testing significance of canonical correlations

Number of canonical correlation	Coefficient of canonical correlation R	Calculated values of statistics χ^2	df	p- value
1	0.7868	291.13	63	0.00000
2	0.6505	155.54	48	0.00000
3	0.4707	78.25	35	0.00004
4	0.3867	43.06	24	0.00983
5	0.2734	20.30	15	0.16074
6	0.2310	9.39	8	0.31088
7	0.1090	1.67	3	0.64181

On the basis of the calculations presented in Tab. 1. It can be said that canonical correlations with numbers 1, 2, 3 and 4 vary considerably. It proves that the four first canonical variables are statistically significant. This confirms advisability of application of the method of canonical correlations analysis of this data. On the basis of calculations, 7 pairs of canonical variables, representing relations of two sets of variables X and Y, have been obtained.

Graphic interpretation of tests results is presented in Fig. 1

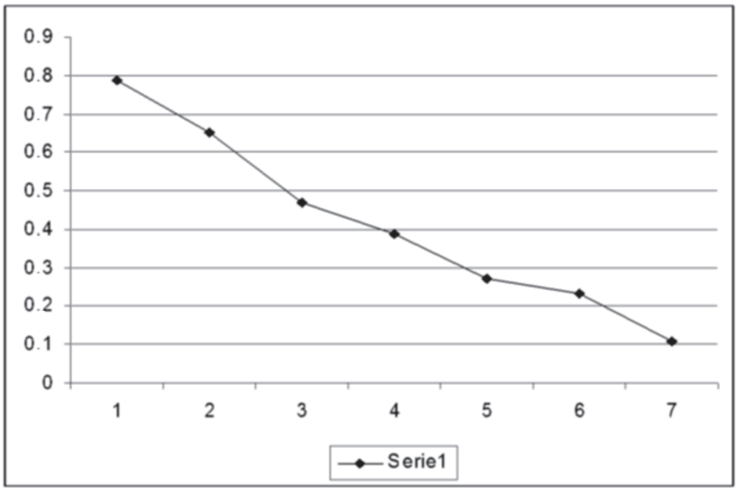

Fig 1. Chart of successive canonical correlations

For the first pair, the first canonical variable, the highest absolute weight values are attributed to the following variables:

- external factors,
- accomplishment time,
- efficiency,
- timeliness.

For the second pair, the first canonical variable, the highest absolute weight values are attributed to the following variables:

- ergonomics,
- esthetics,
- operational readiness.

This provides the basis for concluding that the external factors and ergonomics have the largest influence on formation of the first canonical correlation between the sets of data.

5. Conclusions

In this work, the set of variables has been divided into two separate sets. The first set X contains variables connected with safety, efficiency, informativeness, availability, external factors, timeliness as well as cost-efficiency.

The second set Y contains variables connected with assessment of the system technical objects operation, including: operational readiness, ergonomics, environment friendliness, ergonomics, esthetics, damageability and reliability.

This study aims at proving that the identified subsets are linked by a strong relation. The analysis of this dependence has been made by means of the method of canonical correlations. The analysis of results of tests performed with the use of actual data from a real transportation system, confirms suitability of this method to be used in practice and justification of its application in the presented assessment process.

References

[1] Comrey, A .L., *A First Course in Factor Analysis*, New York Academic Press, New York 1973.

[2] Muślewkski, Ł., *Control Method for Transport System Operational Quality,* Journal of KONES Powertrain and Transport, Vol. 16, No. 3, Warsaw 2009.

[3] Muślewski, Ł., *Evaluation Method of Transport Systems Operation Quality,* Polish Journal of Environmental Studies, Vol. 18, No. 2A, Hard Olsztyn, Olsztyn 2009.
[4] Muślewski, Ł., *Study and assessment of transport system operation efficiency*, Journal of KONES Powertrain and Transport, Vol. 17, No. 4, Warsaw 2010.
[5] Muślewski, Ł., *Identification and analysis of transport systems operation quality assessment criteria*, Journal of KONES Powertrain and Transport, Vol. 17, No. 3, Warsaw 2010.
[6] Muślewski, Ł., Knopik, L., A*pplication of principal component analysis for optimization of a system operation assessment criteria set,*Journal of KONES Powertrain and Transport, Vol. 19, No. 4, Warsaw 2012.
[7] Tabachnik, B. G., Fidell, L., *Computer – Assisted Research Design and Analysis*, Allyn & Bacon, Boston 2001.
[8] Tabachnik, B. G., Fidell, L., *Using Multivariate Statistics,* Haper & Row, New York 1996.
[9] Woropay, M., Muślewski, Ł., *Jakość w ujęciu systemowym*, ITeE, Radom 2005.