
DECISION SUPPORT AND EXPERT SYSTEMS

THE MUCHNIK'S DIAGRAM 'SUCCESS-SUPPORT' FOR SELECTION OF MULTIPARAMETRIC OBJECTS

Anna Pashkovskaya, Mikhail Alexandrov, Yuriy Pervishin

Abstract: *The man-machine graphic procedure of multiparametric object selection is presented. The principal approach is based on reducing multidimensional problem to 2D problem, which is resolved by means of visualization. The technology includes 4 steps: (1) an expert builds a criterion ('Success') as a composition of object parameters; (2) parameters having a positive correlation with Success ('positive parameters') are selected; (3) these parameters are substituted by one generalized parameter ('Support'); (4) new coordinates of objects in 2D space of Success-Support are determined that allows to build the diagram. To calculate Support we use the first principal component of the correlation matrix related to positive parameters. The best objects are those having the largest values of Success and Support simultaneously. We demonstrate the proposed technology with the data of 784 Russian companies of mobile communication.*

Keywords: *decision making support, Success-Support diagram.*

ACM Classification Keywords: *I.2 Artificial Intelligence.*

Introduction

In this paper we consider multiparametric object selection with respect to the problem of investments. However the proposed method can be used for other applications where multiparametric objects are considered. The principal properties of object selection for investments consist in high dimensionality of the problem and in high level of responsibility of decision-makers. The former defines the necessity to use automatic methods of object selection [EC, 2008; Gotze, 2008]. The latter defines the necessity to use manual or semi-automatic methods when expert is involved in the process of solution. However this is possible if the problem dimension is not high and if visualization is convenient for experts.

These requests are resolved by means of building the diagram 'Success-Support' proposed several years ago by I. Muchnik from the Rotgers University (USA). Here: 'Success' is a scalar variable being a composition of parameters an expert selects to express his/her preferences. This variable can be considered as an integral criterion of object quality. 'Support' is a scalar variable that expresses the influence of all parameters positively related to 'Success'. This variable can be considered as an additional criterion of object quality. The principal element here is using the technique of principal components to form the criterion Support. It is supposed that the large values of both criteria correspond to the best objects. Therefore, if we build a diagram of objects using the axes Success and Support then the best objects are concentrated on the upper-right corner.

This method was firstly realized and tested in one Russian company of mobile communication. The second version of the program was developed in B.Sc. thesis [Pashkovskaya, 2013]. In this work the stability of results was explored. In this paper we improve the last version by means of additional procedures of preprocessing and post processing.

The paper is built by the following way. In section 2 we describe the algorithm in details. In section 3 we build the diagram and study stability of results. Section 4 contains the conclusions.

Algorithm

Software platform

The source information is supposed to be presented in Excel sheets. Just for this reason we developed our program on the platform Excel-VBA [Walkenbach, 2010]. We use Excel sheets only for data storage but all calculations and all interfaces are programmed in VBA. The program contains 5 separated modules presented in Figure 1:

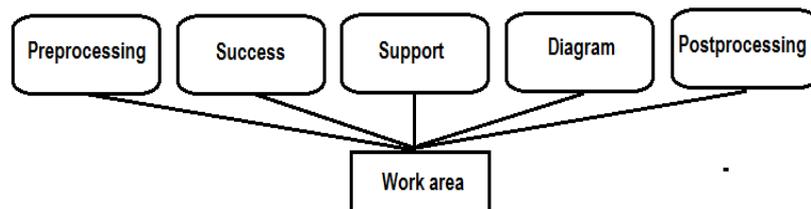


Figure 1. Program architecture

Each module is related to one Excel sheet. The interchange of information between these modules is realized by means of one work sheet. Such a way allows to modify modules independently one from the others and to provide easy intermediate control of module functionality.

Source information and module 'Preprocessing'

In our experiments we use data from the Data Base (DB) SPARK. The data include information about annual activity of 784 Russian companies of mobile communication. Each company is presented by its 51 parameters. Therefore we have a matrix 784x51. Here is the example of parameters presented in SPARK: *Fixed assets, Gross profit, Net profit, Capital and reserves, Cash, Stock, Nominal capital, etc.*

The first a user does on the stage of preprocessing is the selection of parameters, which could be used for building Success and Support. It is a manual procedure: the user only sets indicators located near the titles of parameters. Then the user tests these parameters: whether they are suitable for further processing. For this the user implements 4 procedures of preprocessing: testing completeness, revealing outliers, testing variability, and implements normalization.

- **Testing completeness**

Completeness is measured by portion of existing data for a given parameter. A user sets a threshold for this portion. If this condition is satisfied then the parameter is marked by '1', otherwise by '0'. Here 1 means that this parameter can be used without any correction (when all cells contain any data) and '0' means that this parameter is not ready for further processing.

In order to recover omitted data it is necessary first of all to complete all other procedures: to reveal outliers, to test variability, and to normalize data. After that one can recover data. The simplest solution is:

- to calculate the average value for each selected parameter having marked as the good one;
- to calculate ratios of these average values between themselves;
- to use these ratios in order to recover given parameters taking into account the other parameters of the same object.

- **Revealing outliers**

Speaking 'outliers' we mean parameter values, which are outside the acceptable interval. The acceptable interval is described by means of the formula $[m-k\sigma, m+k\sigma]$, where m is an average value of a given parameter and σ is its standard deviation. It is well known that for the normal law approximately 95% of parameter values are in the interval $[m-2\sigma, m+2\sigma]$, and for the any arbitrary law approximately 99% of parameter values are in the interval $[m-10\sigma, m+10\sigma]$. The latter is the consequence of the famous Chebyshev's inequality. A user sets the critical value k and then the corresponding interval is calculated. If a given parameter has no values outside the mentioned interval then this parameter is marked by '1' otherwise it is marked by '0'.

- **Testing variability**

The variability is measured by means of variation coefficient $\theta = \sigma / m$, where σ and m are defined above. A user sets the threshold for the minimum value of θ . If θ exceeds this threshold then this parameter is marked by '1', otherwise by '0'. The latter means that this parameter is a constant or almost constant and therefore it is not interesting for analysis.

- **Normalization**

In this procedure parameter values are transformed according one of the rules: a) the minimum and maximum values of a given parameter are calculated and then all values reduce to the interval $[0,1]$ or $[-1,1]$; b) a user himself/herself assigns two boundary values a and b and then all values reduce to the interval $[a,b]$; c) the average value m and standard deviation σ of a given parameter are calculated and then all values p are transformed as $(p-m)/\sigma$.

- **Final operations**

User should implement the procedures described above according the following order: completeness=>outliers=>variability=>normalization. Then the user can resolve the problem of incompleteness if it exists. The typical values of the threshold for completeness are $T_c=10\%-20\%$, the typical intervals of acceptable values for testing outliers are $T_o=[m-5\sigma, m+5\sigma]$ or $[m-10\sigma, m+10\sigma]$, the typical values of the threshold for variability are $T_v=0,1-0,2$. The indicators '1' and '0' can help to automate the stage of preprocessing. It is well-known that that preprocessing is the most important stage of calculations, which defines the success of all following calculations. That is why we give much consideration to this question.

Module 'Success'

Parameters selected on the stage of preprocessing are the basis for formation of the criterion Success. It should say that user is free in formation of this criterion. Namely, he/she can:

- use all parameters or any part of them;
- build compositions of parameters and use them in Success.

An example of the composition is the well-known profitability of sales. It is a ratio of net profit to sales volume. Both net profit and sales volume are included to the list of 51 primary parameters mentioned above.

In one of our experiments we build Success as a weighted sum of 3 primary parameters: *Capital and reserves*, *Cash*, and *Current assets*. It can be written in the form $\text{Success} = \sum w_i p_i$, $i=1,2,3$. Here: w_i are weights, and p_i are parameters mentioned above. User assigns the weights w_i according his/her preferences. In particularly, in our first experiment we used: $w_1=0.5$, $w_2=0.3$, and $w_3=0.2$. The program allows to change these weights and immediately recalculate Success for all objects. New values of Success are reflected on the diagram Success-Support.

We would like to say once more that the lineal combination of parameters is not the alone possibility for the formation of Success. The criterion Success can be equal to one of the primary parameters, or to any secondary parameter as the profitability of sales, or to lineal combination of the secondary parameters, etc.

Module 'Support'

All primary parameters selected on the stage of preprocessing are the candidates to be elements of the criterion Support. The principal Muchnik's idea consists in substitution for all this parameters by one generalized parameter that could support the criterion Success. The following steps realize this idea:

- 1) Coefficients of correlation between each parameter and Success are calculated. Parameters having the significant positive correlation with Success are and selected. These parameters are named 'supporting parameters'. Note: we include to the list of 'supporting parameters' the Success itself as an additional parameter if Success contains more then one primary parameter.
- 2) Correlation matrix of the supporting parameters is built. The first principal component (the main factor) of this matrix is calculated. Therefore we have a variable reflecting a generalized relation between Success and all supporting parameters. This variable is a vector in the space of successful parameters. It is just the axis of the criterion Support.
- 3) Projection of the supporting parameters of objects on the vector Support gives the value of Support for this object. Therefore the projections for all objects are calculated.

The threshold for selection of parameters with the significant positive correlation is assigned by a user. The typical values of this threshold are $Tr=0,3-0,5$. Obviously, when we change this threshold we change the list of supporting parameters. With this we change the criterion Support. The library of VBA does not contain the procedure for calculation of the first principal component. So, we have to program this procedure using the well-known iteration method [Hoffman, 2001]. Figure 2 shows the part of interface related to the module Support of the program.

	1	2	3	4	5	6	7
1	Support						
2							
3		Delete		Correlation			
4							
5		Main Factor		Support			
6							
7							
8		Fix assets	Gross profit	Stock	Net profit	Success	
9	Correlation	0,317083841	0,49277738	0,276943781	0,489848626	1	
10	Suitability	1	1	0	1	1	
11							
12		Threshold	0,3				
13							

Figure 2. Interface of module Support

Module 'Diagram'

As a result of the previous stages we have the presentation of all objects in 2D space Success-Support. The visualization of these objects is the Muchnik's diagram. Usually axis X is associated with Support, and axis Y is associated with Success. To facilitate the diagram exploration we divide data of each criterion on 5 intervals marked by (A, B, C, D, E). 'A' denotes the best interval and 'E' denotes the worst interval. Therefore the best objects are collected on the right-upper corner. The part of the diagram is presented on Figure 3.

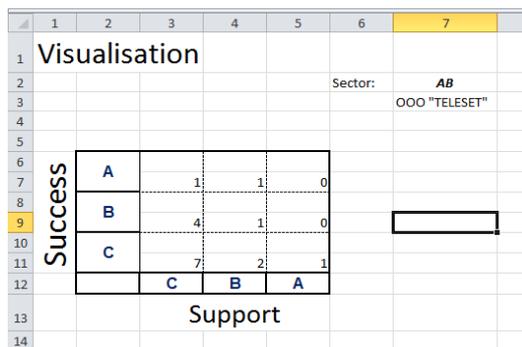


Figure 3. Diagram Success-Support, the part of interface

Interface of the module Diagram allows to show objects related to various cells of the diagram. For example, a user can point the cell 'AB' and obtain the list of objects reflected in this cell. Here: A is the value of Success and B is the value of Support. One can see just this case in Fig.3: the cell AB contains only one object and this object is presented at cell (3,7).

Besides the diagram Success-Support the module 'Diagram' builds two additional charts: the histogram of object distribution on axis Success and the histogram of object distribution on axis Support.

Module 'Postprocessing'

Here, a user studies two questions:

- Whether the best or the worth objects on the diagram Success-Support are also the best or worth for the selected parameters.
- What parameters do not affect the selection of companies on Success-Support diagram.

For this the user indicates: a) an area on the diagram, for example, {AA}, or {AA, AB, AC}, or {AA,AB,BA,BB}, etc.; b) any primary parameter from the list of parameters selected on the stage of preprocessing. Then the program builds two histograms: 1) the histogram of all objects distributed on the axis of a given parameter; 2) the histogram of the objects from the marked area distributed on the same axis with the same scale.

The example is presented on Figure 4. Here a user considers the best objects from the area {AA,AB,BA,BB} and the parameter Capital and reserves. One can see that a) the best objects on the diagram are located in the middle part of the values of parameter Capital and reserves; b) these objects are located as one compact group.

Experiments

Plan of experiments

The object of consideration is a set of 784 Russian companies of mobile communication. Each company is presented by its 51 parameters. The following 9 parameters were selected for our experiments: *Fixed assets*, *Gross profit*, *Capital and reserves*, *Cash*, *Stock*, *Current assets*, *Non-current assets*, *Nominal capital*, *Net profit*. All these parameters passed all procedures of preprocessing without any correction. In the experiment description we use the term 'results'. Here it means the list of companies belonging to a given area from the diagram Success-Support.

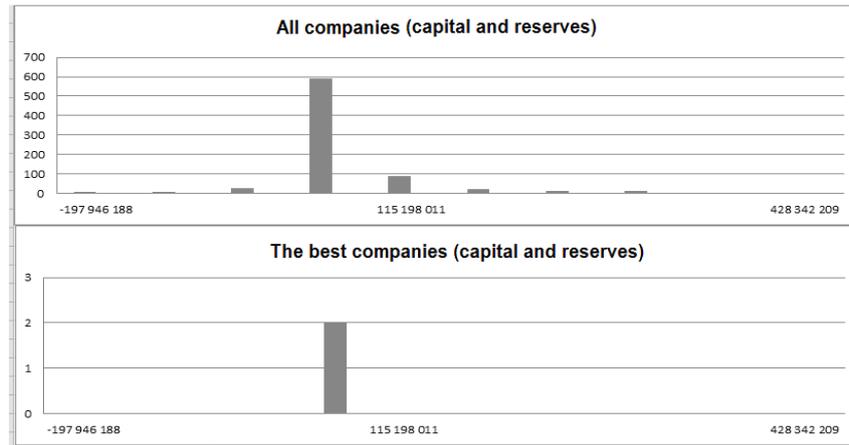


Figure 4. Distribution of all objects and the best objects with respect to parameter capital&reserves

In the first series of experiments we study the sensibility of results to the contents of the criterion Success. For this we form 3 groups of parameters with 3 parameters in each group. The weighted sum of 3 parameters forms the criterion Success.

In the second series of experiment we study the sensibility of results to the number of supporting parameters. For this we implement calculations with the different number of these parameters. We eliminate parameters having the lowest correlation with the Success.

We suppose that the most financially attractive companies should demonstrate stable rankings with different contents of both Success and Support criteria. Just for this reason we study the sensibility to the mentioned criteria.

Variation of the criterion Success

Table 1 summarizes the results of the first series of experiments with the different contents of the criterion Success. The contents of this criterion are reflected in the first column. The criterion Support is based on remaining parameters for each of the four experiments. Here each cell includes the names of companies from the diagram Success-Support.

Table 1. The best companies for different contents of the criterion Success

	AA	AB	BA	BB
Non-current assets, Gross profit, Capital and reserves		OAO "NKS"		ZAO "Samara Telecom", OOO "Prestizh-internet"
Gross profit, Cash, Stocks			OAO "Dagsviaz", OAO "NKS", OOO "Teleset"	ZAO "Samara Telecom", OOO "Prestizh-internet"
Cash, Stocks, Fixed assets	OOO "Teleset"			OOO "Prestizh-internet", ZAO "IT-center"
Stocks, Current assets, Fixed assets	OAO "Dagsviaz"	OOO "Prestizh-internet"	ZAO "Centel"	ZAO "Bashsel"

Variation of the criterion Support

In this series of experiments the criterion Success is fixed. It contains *Capital and reserves*, *Current assets*, and *Fixed assets*. To support this criterion we consider parameters Capital and reserves, Gross profit, Fix assets, and Cash. Table 2 shows the correlation between the criterion Success and these parameters. Table 3 summaries the results of experiments with different elements of the criterion Support. These elements are reflected in the first column. Here each cell includes the names of companies from the diagram Success-Support. The letters refer to the criterion Success.

Table 2. Correlation between the criterion Success and elements of the criterion Support

	Capital & reserves	Gross profit	Fix assets	Cash
Correlation	0,34	0,26	0,57	0,27

Table 3. The best companies for different elements of the criterion Support

	A	B	C
Gross profit, Cash, Capital and reserves, Non-current assets	OAO "Dagsvyaz"	OOO "Prestizh-Internet"	OOO "Infolada"
Cash, Capital and reserves, Non-current assets	OAO "Dagsvyaz"	OOO "Prestizh-Internet"	ZAO "Novgorod telecom"
Capital and reserves, Non-current assets	OAO "Dagsvyaz"	OOO "Prestizh-Internet"	
Non-current assets			OOO "Prestizh-Internet" OAO «Dagsvyaz»

The Table 1 and Table 3 taking together show that the companies OOO "Prestizh-Internet" and OAO "Dagsvyaz" keep their rankings with different contents of the criteria Success and Support. So, these companies can be considered as the most preferable ones for financial investments. This conclusion completely corresponds to the opinion of experts, which selected the most attractive companies using the other man-machine methods. Speaking more exactly these companies were in their list of preferences.

Conclusions

This article describes the original method of object selection based on visual presentations. The method allows to reconcile a subjective opinion of experts concerning the most attractive objects and the formal object description in multidimensional space of parameters. The key element of the consideration is the technique of principal components.

The method was demonstrated on the real example related to selection of mobile communication companies in Russia. The results were evaluated by experts as the very promising ones.

In future we suppose to develop the new version of the program in the environment SciLab or MatLab.

Acknowledgement

The authors are grateful to Mr. Ilya Mafer for his interest to our work and valuable advice.

Bibliography

- [EU, 2008] D. Hubner (Ed.) Guide to cost-benefit analysis of investment projects // European commission, July 2008; http://ec.europa.eu/regional_policy/sources/docgener/guides/cost/guide2008_en.pdf
- [Gotze, 2008] U. Gotze, D. Northcott, P. Schuster, Investment Appraisal Methods and Models // Springer, Berlin, 2008
- [Hoffman, 2001] J. Hoffman, Numerical Methods for Engineers and Scientists // Taylor & Francis Publ., 2001
- [Pashkovskaya, 2013] A. Pashkovskaya. Muchnik's diagram 'Success-Support' and its application to ranking modern mobile communication companies. // B.Sc. thesis, M., RANEPa, 2013
- [Walkenbach, 2010] J. Walkenbach, Power Programming with VBA // Wiley Publishing, inc, 2010

Authors' Information



Anna Pashkovskaya – *M.Sc student, Lomonosov Moscow State University; GSP-1, Leninskie Gory, Moscow, 119991, Russian Federation*

e-mail: APashkovskaja@gmail.com

Major Fields of Scientific Research: decision-making support, world economy



Mikhail Alexandrov - *Professor of the Russian Presidential Academy of national economy and public administration; Prosp. Vernadskogo 82, bld. 1, Moscow, 119571, Russia; fLexSem Research Group, Autonomous University of Barcelona, 08193 Bellaterra (Barcelona), Spain;*

e-mail: MAlexandrov@mail.ru

Major Fields of Scientific Research: data mining, text mining, mathematical modelling



Yuriy Pervishin – *Professor of the Russian Presidential Academy of national economy and public administration; Prosp. Vernadskogo 82, bld. 1, Moscow, 119571, Russia;*

e-mail: pervishin@ranepa.ru

Major Fields of Scientific Research: macroeconomy, mathematical modelling