Fuzzy Model for Analysing Implicit Factor Influence

Dmitry M. Nazarov, Dmitry A. Azarov, 
Yakov P. Silin, Svetlana V. Begicheva 
Ural State University of Economics 
Yekaterinburg, Russia 
slup2005@mail.ru, andr17@yandex.ru, silin@usue.ru, 
begichevas@mail.ru

Gennady N. Smorodin 
Dell EMC Academic Alliance Russia & CIS 
St. Petersburg, Russia 
gennady.smorodin@dell.com

Abstract—The article deals with the problem of analysing implicit factor influence. It later reveals a mathematical model for conducting a search for implicit factors within a system which is based on a fuzzy set theory and employs fuzzy binary relations and is put into practice as a combination of web-services posted on http://bi.usue.ru/. The model’s capability to identify implicit factors by making use of expert judgements presented in natural language is lastly demonstrated.

I. INTRODUCTION

Assessing the impact of certain factors on any system performance indicators delimits the problem area in making strategic, tactical and operating decisions. Consideration of the implicit factor effects proves to be highly relevant in contemporary studies due to their ability to transform the system and its overall performance as a result.

We treat implicit effects as non-evident impacts within the system caused by implicit factors and capable of producing a synergistic effect. Furthermore, we classify implicit factors as non-evident factors, having a significant effect on the system performance based on subtle information, previously not taken into account, which is practically useful and available for generating knowledge and making decisions [1].

The existing traditional economic and mathematical tools applying deterministic, probabilistic and integrated approaches do not yield the required results due to the existence of implicit, uncertain and indirect influences in the inner and outer environment of the systems studied. The present stage of development of economic and mathematical methods implies the use of a fuzzy control model [2–3] as an alternative to traditional approaches, which allows the user to effectively manage to take into account the uncertain, inaccurate, incomplete, non-quantitative and implicit data.

II. FUZZY SET METHOD FOR ASSESSING IMPLICIT FACTOR INFLUENCE

The basic idea behind application of fuzzy set approach consists in the fact that a certain index assumed to be interval and to be determined (fuzzified) by a span instead of a number. This point aims to reflect an actual situation of more or less exactly specified threshold values, within which a parameter can vary. As a consequence, we need to formalize the understanding of the concerned index probable values along with indication of the set of probable values and degrees of uncertainty of their adoption. Then, after calculation of the probability distribution of the overall index, we should pass on defuzzification and interpretation stages based on the system of rules and using developed output tools.

The model-building task is to find and to quantify effects of subtle, implicit factors, which are communication channel elements, on directly related parameters and through them on certain key indices. The important aspect is finding the very implicit factors. The basic model-building tools are fuzzy binary relations, their composition, and data mining algorithms. The model-building logic is illustrated below.

III. LOGIC BEHIND CONSTRUCTING A FUZZY MODEL FOR REVEALING IMPLICIT FACTOR INFLUENCE

Suppose we are given a set of \( A = \{a_1, a_2, \ldots, a_n\} \). With the specified degree of probability we can find two elements among the elements of the set, which have a relatively small mutual effect, but there is an element distinctive from the above two elements, with introduction of which the effect becomes significant.

Some definitions are introduced below. Suppose \( U \) is any set, \( U^2 \) is Cartesian square of this set \( U^2 = U \times U = \{(a, b); a, b \in U\} \). The fuzzy binary relation on the \( U \) set is the \( U^2 \) fuzzy subset. Conventional notation format of the fuzzy binary relation for the discrete and continuous sets are represented in (1) and (2), respectively.

\[
\Gamma = \sum_{i,j} \mu_{ij} \left( u_i, u_j \right) / \left( u_i, u_j \right), \quad (1)
\]

\[
\Gamma = \int_{U^2} \mu_{ij} (x, y) / (x, y). \quad (2)
\]

Matrices elements of which are values of the membership function of \( \mu_{ij} (x, y) \) fuzzy binary relation are denoted as \( J_{ij} \).

The composition of \( \Gamma_1 \) and \( \Gamma_2 \) fuzzy binary relations is specified by such a fuzzy binary relation of \( \Gamma = \Gamma_1 \circ \Gamma_2 \), that (3) is valid:

\[
\mu_{ij} \circ \Gamma \left( x, y \right) / \left( x, y \right) = \bigcup_{z \in U} \left( \mu_{ij} \left( x, z \right) / \left( x, z \right) \right) \bigcap \left( \mu_{ij} \left( z, y \right) / \left( z, y \right) \right). \quad (3)
\]
Given that the intersection of $\mu_{x}(x,z)/(x,z)$ and $\mu_{y}(z,y)/(z,y)$ single-point fuzzy sets is generally performed by the logical $\alpha$-norm, and its union is performed by the logical $\alpha$-conorm: $\bigwedge_{\alpha} a = \min\{a, b\}, \bigvee_{\alpha} a = \max\{a, b\}$, (3) takes the form of (4).

$$
\mu_{x-y}(x,y)/(x,y) = \max\left(\min\left(\mu_{x}(x,z), \mu_{y}(z,y)\right)\right)/(x,y).
$$

Equations (5) and (6) are the relation composition graph for discrete and continuous sets, respectively.

$$
\Gamma_1 \circ \Gamma_2 = \sum_{x} \mu_{(x,y)}/(x,y) = \sum_{x} \left(\max\left(\min\left(\mu_{x}(x,z), \mu_{y}(z,y)\right)\right)\right)/(x,y),
$$

if $U$ is a finite set;

$$
\Gamma_1 \circ \Gamma_2 = \int_{x} \mu_{x}/(x,y) = \int_{x} \left(\max\left(\min\left(\mu_{x}(x,z), \mu_{y}(z,y)\right)\right)\right)/(x,y),
$$

if $U$ is part of the number axis or the entire number axis.

Therefore, from (4) for a $U$ finite set we can obtain:

$$
J_{x-y} = J_{x} \circ J_{y} = \left(\max\left(\min\left(\mu_{x}(u_i), \mu_{y}(u_j)\right)\right)\right)/(x,y) = \left(\mu_{x}(u_i), \mu_{y}(u_j)\right)_{max},
$$

where $n$ is a number of $U$ set elements.

Build-up the $J_{x}$ matrix for the $A$ set.

$$
J_{x} = \begin{pmatrix}
    s_{11} & s_{12} & \ldots & s_{1n} \\
    s_{21} & s_{22} & \ldots & s_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    s_{n1} & s_{n2} & \ldots & s_{nn}
\end{pmatrix},
$$

where $s_{ij}$ is an extent of the $a_j$ index effect on the $a_j$ index.

Then, to find mediate effects we can calculate values of $J_{x-y}$ matrix using (4):

$$
J_{x} = J_{x} \cdot J_{y} = \begin{pmatrix}
    f_{11} & f_{12} & \ldots & f_{1n} \\
    f_{21} & f_{22} & \ldots & f_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    f_{n1} & f_{n2} & \ldots & f_{nn}
\end{pmatrix}.
$$

The reflexive selection of mediated factors [4] performed as part of data mining assumes existence of such a pair of $s_{ij}$ and $f_{ij}$, that $s_{ij} << f_{ij}$, and indicates the presence of an implicit effect, which becomes apparent due to an intermediate factor within the system [1].

IV. FINE-TUNING THE MODEL TO TREAT FUZZY BINARY CORRESPONDENCE

When necessary, the model may be adapted to work with fuzzy binary correspondences. The conceptual similarities and differences of this adaptation are described below.

Binary correspondence in the case of $A \times B$ set means a $\Gamma$ subset of the Cartesian product of sets $A$ and $B : \Gamma \subseteq A \times B$. The Cartesian product $A \times B$ describes the product of all the sets that feature $A$ set element in the first position, and $B$ set element in the second position. In the case when $A = B$, binary correspondences amount to the mode of binary relations $\Gamma \subseteq A^2$.

Compositions of binary correspondences and binary relations are defined similarly. Compositions of fuzzy binary correspondences $\Gamma_1 \subseteq A \times B$ and $\Gamma_2 \subseteq B \times C$ are defined by (7), provided that (8) is correct for the membership function.

$$
\Gamma = \Gamma_1 \circ \Gamma_2 \subseteq A \times C,
$$

$$
\mu_{x-y}(x,y) = \left(\mu_{x}(x,z) \cap \mu_{y}(z,y)\right)/(x,y)
$$

(8)

The intersection and the union of single-point fuzzy sets in the case of binary correspondences are made according to the abovementioned rules. In this case, (8) goes over to (9):

$$
\mu_{x-y}(x,y) = \left(\min\left(\mu_{x}(x,z), \mu_{y}(z,y)\right)\right)/(x,y)
$$

(9)

The graph of finite set correspondences’ composition follows (10); (11) defines the graph for the sets that represent an interval of the number axis or the entire number axis.

$$
\Gamma_1 \circ \Gamma_2 = \sum_{x \in A, y \in C} \left(\max\left(\min\left(\mu_{x}(x,z), \mu_{y}(z,y)\right)\right)\right)/(x,y),
$$

(10)

where $A, B, C$ are finite sets;

$$
\Gamma_1 \circ \Gamma_2 = \int_{x \in A, y \in C} \left(\max\left(\min\left(\mu_{x}(x,z), \mu_{y}(z,y)\right)\right)\right)/(x,y),
$$

(11)

if $A, B, C$ are an interval of the number axis or the entire number axis.

From (10) it follows that the matrix of the composition of $J_{x-y}$ relations, when $A, B, C$ are finite sets, is nothing but a maximin matrix product of $J_{x}$ and $J_{y}$.
where \( p \) is the amount of \( B \) set elements; \( m \) is the amount of \( A \) set elements; \( n \) is the amount of \( C \) set elements.

Building a model for estimating implicit factor effects based on fuzzy correspondences falls into two steps:

1) building sub-models \( A \) including the set of implicit factors, \( B \) incorporating the set of indirect indices, and \( C \) involving the set of key parameters;

2) integration of sub-models into a general model, its analysis and solution of the problem set.

The operation sequence of the first step may include the following procedures:

- primary specification of a number index set for each sub-model;
- making lists of number index sets.

The operation sequence of the second step includes the following:

- evaluation of interdependence between the indices in pairs: \((A,B)\), \((B,C)\), \((A,C)\);
- detecting of indirect effects that indices of the sub-model \( A \) have on indices of the sub-model \( \tilde{N} \);
- explanation of the obtained results.

Dependences are defined using \( J_{AB} \), \( J_{BC} \) and \( J_{AC} \) matrices for the set of \( A \), \( B \), and \( C \) indices:

\[
J_{AB} = \begin{pmatrix} s_{11} & s_{12} & \cdots & s_{1m} \\ s_{21} & s_{22} & \cdots & s_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ s_{n1} & s_{n2} & \cdots & s_{nm} \end{pmatrix},
\]

\[
J_{AC} = \begin{pmatrix} z_{11} & z_{12} & \cdots & z_{1k} \\ z_{21} & z_{22} & \cdots & z_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1} & z_{n2} & \cdots & z_{nk} \end{pmatrix},
\]

\[
J_{BC} = \begin{pmatrix} u_{11} & u_{12} & \cdots & u_{1k} \\ u_{21} & u_{22} & \cdots & u_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ u_{m1} & u_{m2} & \cdots & u_{mk} \end{pmatrix},
\]

where \( s_{ij} \) \((0 \leq s_{ij} \leq 1; \ i = 1, 2, \ldots, n; \ j = 1, 2, \ldots, m)\) is an extent of the \( a_i \) index effect on the \( b_j \) index, \( z_{ij} \) \((0 \leq z_{ij} \leq 1; \ i = 1, 2, \ldots, n; \ j = 1, 2, \ldots, k)\) is an extent of the \( a_i \) index effect on the \( c_j \) index, \( u_{ij} \) \((0 \leq u_{ij} \leq 1; \ i = 1, 2, \ldots, m; \ j = 1, 2, \ldots, k)\) is an extent of the \( b_j \) index effect on the \( c_j \) index.

The extent of the direct effect \( a_i \) makes on \( c_j \) is determined by the \( z_{ij} \) element of the \( J_{AC} \) matrix. Similarly, the extent of the direct effect \( b_j \) makes on \( c_j \) is determined by \( z_{ij} \). In addition to the direct effect, the \( a_i \) index affects \( c_j \) through the intermediate factor \( b_j \), which is an index of the sub-model \( A \).

The extent of the indirect effect \( a_i \) has on \( c_j \) through \( b_j \) is taken as values \( z_{ij} \), which are minimums of \( s_{ij} \) and corresponding \( u_{ij} \) and \( u_{ij} \). The equation is valid, then we showed the indirect effect through the difference.

Equation (12) specifies combined indirect effect that \( a_i \) element produces on \( c_j \):

\[
z_{ij} = \max(\min(s_{ij}, u_{ij}), \min(s_{ij}, u_{ij}), \ldots, \min(s_{ij}, u_{ij})).
\]

The matrix \( J_{AB} \) and \( J_{BC} \) product given in (13) specifies the indirect effect \( A \) sets elements on \( C \) set elements through \( B \):

\[
J'_{AC} = J_{AB} \cdot J_{BC} = \begin{pmatrix} z_{11}^* & z_{12}^* & \cdots & z_{1k}^* \\ z_{21}^* & z_{22}^* & \cdots & z_{2k}^* \\ \vdots & \vdots & \ddots & \vdots \\ z_{n1}^* & z_{n2}^* & \cdots & z_{nk}^* \end{pmatrix},
\]

where \( z_{ij}^* \) is calculated using (12).

If the extent of direct effect \( A \) makes on \( C \), determined by following the steps of hierarchy analysis, exceeds the indirect effect, then it is not worth being taken into account. If the inequality \( z_{ij}^* > z_{ij} \) is valid, then we showed the indirect and previously ignored effect that \( i \)-th implicit factor makes on \( j \)-th resulting index. Moreover, evaluation of the extent of such effect may be considered as \( z_{ij}^* - z_{ij} \) difference.

Consequently, using compositions of binary correspondences implicit, indirect relations and cross-effects between the elements of \( A \) and \( C \) sets may be found out, when correspondences for \( A \times B \) and \( B \times C \) sets are given; intermediate factors may be specified as well.
V. OPERATIONAL DEMONSTRATION OF THE MODEL

To explain the described model we’ll give an example for the $\Gamma = \{a, b, c\}$ set. Suppose $a$, $b$, and $c$ are the three characteristics of a specific system. The $J_I$ matrix specifies the mutual effects of the above features:

$$J_I = \begin{pmatrix} 0.6 & 0.4 & 0.4 \\ 0.3 & 0.8 & 0.7 \\ 0.8 & 0.9 & 0.5 \end{pmatrix}$$

To determine a mediate mutual effect of these characteristics we can calculate the composition matrix:

$$J_{I^2} = \begin{pmatrix} 0.6 & 0.4 & 0.4 \\ 0.3 & 0.8 & 0.7 \\ 0.8 & 0.9 & 0.5 \end{pmatrix} \cdot \begin{pmatrix} 0.6 & 0.4 & 0.4 \\ 0.3 & 0.8 & 0.7 \\ 0.8 & 0.9 & 0.5 \end{pmatrix} = \begin{pmatrix} 0.6 & 0.4 & 0.4 \\ 0.7 & 0.8 & 0.7 \\ 0.6 & 0.8 & 0.7 \end{pmatrix}$$

The $J_{I^2}$ matrix indicates the existence of a significant mediate effect of the $b$ factor on the $a$ factor: $\mu_I (b, a) = 0.3, \mu_{I^2} (b, a) = 0.7$. Comparison of the cross impact between the characteristics, the intermediate factor taken into account and neglected, is graphically demonstrated in Fig. 1.

To demonstrate the process of detection of the intermediate factor we can perform the analysis of evaluation of the membership function $\mu_{I^2} (b, a) = 0.7$:

$$\mu_{I^2} (b, a) = \max\{\min(0.3; 0.6), \min(0.8; 0.3), \min(0.7; 0.8)\} = 0.7$$

We make the following statement concerning the indirect effect of the $b$ factor on the $a$ factor: "The $c$ factor is determined by the $b$ factor ($\mu_I (b, c) = 0.7$), the $c$ factor makes effect on the $a$ factor ($\mu_I (c, a) = 0.8$)." As a result, this leads us to the conclusion: "The $b$ factor via the $c$ factor determines the $a$ factor ($\mu_{I^2} (c, a) = 0.7$)." Thus, the $c$ factor acts as an unobvious element of the system, which is an intermediary between the $b$ factor and the $a$ factor that are relatively weakly interconnected without considering the effect thereof. Identification of this significant factor based on data mining techniques makes it possible to perform an adequate object behavior evaluation, which otherwise would be considerably distorted.

Crucial options of the model when operated on a certain system ensure its applicability for the analysis of a wide range of data. The model using Goguen’s fuzzy implication [5] and defuzzification according to the method of the centre of gravity [6; 7] was officially approved in the research of the corporate culture [8], the military-industrial complex (MIC) of the Russian Federation [9], and demonstrated high reliability (measure of inaccuracy of significant subtle effect finding was below 3% regardless of the subject of research).

The evidence for the model superiority over popular and widely spread econometric techniques applied to analysing influences within a great number of systems becomes clear from the research into the Russian military industrial complex. The results gained show that the effects produced by the MIC on Russia’s economy can be measured against 39 main parameters. Correlation and regression analysis of the exerted impact performed as an alternative to fuzzy methods of evaluation confirms the existence of high to very high interrelation between the GDP, military expenses, volume of the state defense order and materiel exports (Table I).

<table>
<thead>
<tr>
<th>Variable</th>
<th>GDP</th>
<th>Military expenses</th>
<th>State defense order</th>
<th>Materiel exports</th>
</tr>
</thead>
<tbody>
<tr>
<td>GDP</td>
<td>1</td>
<td>0.978</td>
<td>0.661</td>
<td>0.918</td>
</tr>
<tr>
<td>Military expenses</td>
<td>0.978</td>
<td>1</td>
<td>0.791</td>
<td>0.975</td>
</tr>
<tr>
<td>State defense order</td>
<td>0.661</td>
<td>0.791</td>
<td>1</td>
<td>0.860</td>
</tr>
<tr>
<td>Materiel exports</td>
<td>0.918</td>
<td>0.975</td>
<td>0.860</td>
<td>1</td>
</tr>
</tbody>
</table>

However, it fails to provide reliable estimation basing on data collected within a period of less than 5 years. Moreover, the results obtained from the largest possible series of 18 observations cannot be considered fully accurate. Despite the fact that the OLS regression equation (14) seems quite promising due to being homoscedastic according to the White test ($F$-test = 0.32 > 0.05) and employing statistically significant coefficients ($p$-values do not exceed 0.05) while showing no signs of autocorrelation as per Durbin–Watson and Breusch–Godfrey tests ($d = 2.04$; coefficient of correlation in the errors in a regression model is -0.02 with a $t$-test equal to -0.086) there is multicollinearity between explanatory variables.

$$Y = 147.79 + 37.12 \cdot \text{Military expenses} - 0.32 \cdot \text{State defense order} - 0.85 \cdot \text{Materiel exports}$$

(14)

The efforts taken to eliminate the multicollinearity effects, albeit successful, prove to be unproductive: regression equation (15) taking into account the most highly correlated variables is even more unreliable than (14). The $Y$-intercept’s
The White test still shows no heteroscedasticity ($F$-test = 0.14 > 0.05). The empirical results gathered. The three basic research tools latest tools for preparation, conduct and analysis of the computation process through a series of open interfaces. It also means for transferring the data and controlling the an interconnected web services framework, providing the means to perform a more flexible analysis with no loss in quality of its results.

VI. TOOLKIT FOR ANALYSING IMPLICIT FACTOR INFLUENCE

The above-mentioned algorithm is automated and implemented in the form of the ‘Implicit influences Joomla component’ web-application combining three web-services, namely ‘Implicit factor search’, ‘Implicit factor assessment’, and ‘Implicit factor impact evaluation’, which allow to gain access to the functional data mining modules incorporated into the model of fuzzy binary relations and correspondences. This web-app was based on the web services architecture by W3C that supports the standards of the global information network, and operates in compliance with the LAMP technology, thereby creating a single environment for the web-services, standardizing the development tools and operating procedures. These include access control, administration, accumulation of information and knowledge, conditions for expert, specialist and programmer collaboration.

Currently, the developed tools and software have multi-layer technology architecture: the first (top) layer represents an object-oriented web application that provides a user interface to access the software components of the model; the middle layer, hidden from users, includes software which implements original models and algorithms based on data mining through fuzzy control technologies. Each algorithm is operated by a specific service. The bottom layer forms a database which is utilized to accumulate the input and output data collected.

The main rationale for the development of the ‘Implicit influences Joomla component’ web-application was to create an interconnected web services framework, providing the means for transferring the data and controlling the computation process through a series of open interfaces. It also ensures that the problem-oriented module is equipped with the latest tools for preparation, conduct and analysis of the empirical results gathered. The three basic research tools feature the following services:

- ‘Implicit factor search’ which is used to simulate the implicit factor search process and performs the reflexive selection procedure;
- ‘Implicit factor assessment’ which allows a user to assess the implicit factor basing on its logical-semantic decomposition into corresponding indicators and enabling appropriate fuzzy control modifications;
- ‘Implicit factor impact evaluation’ which is designed to give a forecast for the key indicators changing in line with the implicit factor values.

The general architecture of the ‘Implicit influences Joomla component’ application can be represented in the form of three main subsystems (Fig. 2):

- Mathematical subsystem includes main means of performing computation within models for assessing implicit factors, most notably a set of programs intended for input variable data mining as well as a set of utility programs ensuring live interaction with other subsystems.
- Graphics development support subsystem uses the GUI to provide a visual model of the subject and configure the computational domain to make the layout more convenient. It also interacts with a full range of scripts and databases used during the modelling process (Fig. 3) and constitutes a cross-platform program written in PHP with the use of embedded software libraries and having a windowed interface.
- Computational domain database subsystem contains a structured and updated relational knowledge base of subjects modelled, particularly, most notably it stores the sets of computational data, computation procedures, models and sets of indicators along with standards of data base interaction and operation.

As one of the most promising methods of finding and evaluation of implicit effects, the aforesaid algorithm in the course of implementation of the system of machine learning may be applied for the purpose of handling the problems of adaptable control with application of artificial neural networks [9], [10].

![Fig. 2. ‘Implicit influences Joomla component’ subsystems][1]

---

[1]: Image of the subsystems diagram.
VII. CONCLUSION

Thus, the problem of revealing implicit factors that exist within a system remains topical under toughening requirements to description of their current and estimated condition. The system simulation based on the study of the implicit effects between its elements represents one of the auspicious research methods to overcome control problems with the aim of ensuring the required performance. Particular attention in the course of designing a model under the state of uncertainty should be paid to fuzzy techniques offering a number of critical benefits compared to conventional methodological approaches that ensure a widespread application of a fuzzy method due to its flexibility without diminishing the reliability of the acquired results.

Fuzzy model implementation in the form of various information system modules can be very helpful in carrying out system control and management tasks. Firstly, it will enable to describe the given system’s implicit factors and examine its fuzzy response; secondly, it will give an opportunity to take into account complexity and nonlinear interrelation of external and internal environment factors on the basis of expert analysis which will consequently allow for fuzzy nature of the assessments; thirdly, it will provide means for automating the process, its multiple iteration and constructing simulation models adjusted for specific features of the test systems.

The fuzzy model described in detail in the paper is intended to work with a wide range of control systems and makes it feasible, based on the composition of fuzzy binary relations, fuzzy binary correspondences, and data mining, to spot the implicit effects within. The above algorithm gives the opportunity to detect the unobvious elements of a system, as well as to calculate their effects. The obtained results contribute towards enhancing the appropriateness of the research subject behavior assessment.

REFERENCES


