Fuzzy sets visualization for decision-making processes in social studies

O Berestneva¹, O Marukhina¹2, S Romanchukov¹

¹Engineering School of Computer Science and Robotics, National Research Tomsk Polytechnic University, Lenin Avenue, 30, Tomsk, 634050, Russia
²Corresponding author: Marukhina@tpu.ru

Abstract. This paper is devoted to visualization of multidimensional fuzzy series and describes analysis methodology based on images that can be built after the initial time series transformation. It is shown that the most rigid structure of a fuzzy number is the number of components with the highest value of membership function. Built in accordance with the proposed method a series of visual images reflect the tendency of the fuzzy time series. The application of fuzzy time series visualization facilitates the identification of high-quality bonds in the consideration of dynamical semi-system (including social and economic). The proposed approach allows a qualitative level to solve the problem of grouping, structuring and forecasting in such systems, but there are doubts in numerical accuracy and quality.

1. Introduction
Social system modeling requires a number of basic assumptions, relying on statistics, expert estimates, theoretical material from previous studies, etc. A number of decisions should be made and any algorithms that support decision-making process tend to be useful or at least interesting. At the same time the main feature of decision-making systems is the increasing mathematization of such systems. The reason for attracting mathematical methods is the possibility to use unambiguous procedures that can be performed on computers. Moreover, most of the tasks solved on a computer have an explicit algorithmic nature, which requires the solution and correct computational methods to exist (also, to be unique and reliable in a finite time).

The consideration of decision-making methods in modeling social and economic systems suggests that in many cases these circumstances cannot be fulfilled. The existence of a solution can be proved only in individual cases, when moving from the practical formulation of the problem to its mathematical analogue, one has to ignore many factors influencing the decision, the analysis of the practical formulation of the problem leads to conclusions about the possibility of the existence of a set of “solutions”. In such cases, the decision maker finds himself in circumstances where the choice of decision turns into the process of choosing among a variety of alternatives, often based on conflicting conditions. It should be borne in mind that in social and economic systems there are many factors that change over time. Tracking the status of such systems is associated with the problem of time series (the set of statistical data itself can serve as an example of multidimensional time series based on random variables). These series have a high degree of uncertainty and require the use of appropriate methods for their processing. Such methods do not provide high accuracy of the result. In this case, we can point to neural network and fuzzy models, as well as artificial intelligence models [1-4].
Modeling the behavior of social and economic systems is based on a model of a fuzzy dynamic process, called the fuzzy time series. Multidimensionality fuzzy time series is a property that ensures the complexity of the presentation of information while maintaining independence between the recorded indicators. This paper discusses the simplest case when the multidimensional fuzzy time series is represented by indicators that are measurable and are expressed by real numbers:

\[ W = \{ V(t_1), V(t_2), \ldots, V(t_i), \ldots, V(t_k) \}. \]  

Here \( W \) stands for multidimensional fuzzy time series, the final set of performance vectors; \( t_1, t_2, \ldots, t_i, \ldots, t_k \) – timestamps, when \( V(t_i) \) can be observed and registered:

\[ V(t_i) = \{ q_1(t_i), q_2(t_i), \ldots, q_n(t_i) \}. \]

The vector of indicators is measured and recorded at the moment \( t_i \). Fuzziness is hidden in the variables \( q_j(t_i) \). The analysis of the multidimensional fuzzy time series is complicated by the fact that obtaining complete quantitative data is not possible or existing data are not sufficient, and that is a classical situation for social studies.

2. Problem statement

For a given multidimensional fuzzy time series a visual image is required to simplify manipulations and analysis of the multidimensional fuzzy time series. As it follows from (1) and (2), the multidimensional fuzzy time series \( W \) is most easily represented as a spreadsheet or a list of lists. In this case, \( V(t) \) is a row vector consisting of various indicators measured and recorded at time \( t_i \):

\[ V(t) = \{ q_1(t), q_2(t), \ldots, q_n(t) \} = \{ q_1, q_2, \ldots, q_n \}, \]

where \( q_i \in R \), \( V(t_i) \) can be considered a point in space \( R_n \) for times \( t_i \).

From (1) it follows that the multidimensional fuzzy time series represents a sequence of points \( R_n \).

\[ W = \{ V(t_1), V(t_2), \ldots, V(t_i), \ldots, V(t_k) \} = \{ V_1, V_2, \ldots, V_i, \ldots, V_k \}. \]

It is a finite sequence, and it can be considered a segment of multidimensional fuzzy time series. Consider the set of polynomials orthonormal on \([0,1]\):

\[ P = \{ P_0(\tau), P_1(\tau), \ldots, P_{n-1}(\tau) \}^T = \text{colomn} \{ P_0(\tau), \ldots, P_{n-1}(\tau) \} \]

\( F(\tau) \) graphical image

The main feature of these polynomials will be the fact that they can be represented as graphs of functions of the argument \( \tau \in [0,1] \). Then a function \( F_i(\tau) = (V(t_i), P) = \sum_{l=1}^{s} q_{i,l} P_l(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_i) \). Sequences \( W \) matches with the sequence \( F_i(\tau) \) can be assigned for each line \( V(t_
This vector can be observed only as a set of values in a spreadsheet, and the function \( F_i(\tau) \) can be represented as a graph (Fig. 1).

3. Fuzzy time series visualization

The visual properties of orthonormal polynomials are most pronounced on the interval \([0,1]\), and for this reason it serves as the basis for the representation of images \( F_i(\tau) \) and prototypes \( V(t_i) \) \([5, \, 6]\). There are two spaces: the space of the originals \( V(t_i) \) and the space of the images. These two spaces are interconnected by the isometric condition. Indeed, the Euclidean norm in the space of the originals \( R_n \) equals to

\[
\| V(t_i) \|_{R_n} = \sum_{k=1}^{n} q_{ik}^2
\]

and the norm of the function \( F_i(\tau) \) in the image space \( L_n(\tau) \) can be calculated by the formula

\[
\| F_i(\tau) \|_{L_n} = \sqrt{\int_0^1 \left( \sum_{k=1}^{n} q_{ik} P_k(\tau) \right)^2 d\tau} = \sqrt{\sum_{k=1}^{n} q_{ik}^2}
\]

The last equality is achieved in view of the orthonormal properties of the polynomials \( P(\tau) \). Thus, the following mathematical equality is fulfilled \( \| V(t_i) \|_{R_n} = \| F_i(\tau) \|_{L_n} \), as a proof of this two spaces isometry.

It allows defining \( r \) as \( r(A(\tau), B(\tau)) = \| A(\tau) - B(\tau) \| \) and observing the differences between the two functions - images \( A(\tau) \) and \( B(\tau) \). The created correspondence between \( A(t_i) \rightarrow A(\tau) \) and \( B(t_j) \rightarrow B(\tau) \) allows observing the difference between the images on the computer screen.

In this case, the analytical ability of the decision maker is included in the scheme for studying images, and hence the originals. The human visual system is the fastest image analysis system and allows drawing conclusions about the properties of sets of originals. Thus, there is a separation of functions between the decision maker and the computer. On a computer, the spreadsheet is routinely processed and images are prepared, and the decision maker performs more subtle, analytical data processing.

If we treat \( A(t_i) \) and \( B(t_j) \) as two points of multidimensional space, then segment \( AB \) can be represented as \( AB = (1 - \lambda) A(t_i) + \lambda B(t_j) \), \( \lambda \in \[0, \, 1\] \). The parameter \( \lambda \) is called a sequential parameter, i.e. parameter responsible for the sequence of images.

Based on three variables (composition parameter \( \tau \); variable for the value of images \( F \); sequential parameter \( \lambda \)), it is possible to form the visual space \( \{ \tau, F(\tau, \lambda), \lambda \} \), which is used to represent the multidimensional fuzzy time series.

As a parameter \( \lambda \), it is possible to use the metric function \( r \). In this case, we get a clear picture of the multidimensional fuzzy time series on the computer screen. The sequential parameter \( \lambda \) can play the role of time \( t \). Note that a transition has been made from the sequence of originals to the sequence of visual images, which have all the informational features of fuzziness. They are stored in the original spreadsheet and have no effect on the image diagram.

From a practical point of view, all the originals belong to a certain universe, and each indicator is associated with a certain property, and this property can generate a fuzzy predicate that can be characterized by a truth value. It is considered that a fuzzy predicate can take a continuum of truth values \([0,1]\). And, as always, the number 0 corresponds to the concept of "zero", and the number 1 – the concept of "truth." The more the indicator corresponds to the property in question; the closer to 1 should be the truth value of a fuzzy predicate \([7]\).

Membership functions can be used as parameters that define the sequential parameter \( \lambda \). Therefore, the visual images \( F(\tau) \) will be ordered according to the values of the membership functions. For all images \( F(\tau) \), one membership function is fixed with different values for each of the images in accordance with the data.
The visualization allows exploring the question of the impact of a particular membership function on the sequence of images and their mutual arrangement. This information can be obtained in the form of lists, but the visual picture covers the entire ensemble of images and allows selecting the most "outstanding" images in it. The position of these images will be grouped closer to the ends of the sequential segment, and the entire ensemble of images can be characterized by image condensation points. The type of the membership function will influence the sequence of images, for this reason the membership function can be set by an expert, focusing on such properties that can be measured in a certain known quantitative scale. Due to the fact that the fuzzy time series does not require precise definition of membership functions, sometimes it is enough to fix the most characteristic values [8].

Visual grouping of data allows the decision maker presenting the structure of the fuzzy time series and, most importantly, roughly determining the development trend of this series. However, when forming visual images, it is necessary to take into account that the indicator $q_i(t)$ is presented as a list of fuzzy values. It means that

$$q_i(t) = \{q_1(t), \mu_1(t) \}, \{q_2(t), \mu_2(t) \}, \ldots, \{q_m(t), \mu_m(t) \},$$

$q_i(t)$ – is the value of the index for which the membership function value is defined and associates it with a real number from the interval $[0,1]$. Thus, the indicator $q_i(t)$ is represented by a fuzzy set.

Obviously, with such values abundance, visualization or transition from originals to images is a problem. To solve this problem, we offer several approaches, including the following: the most reliable fuzzy time series is distinguished from the data presented in the fuzzy time series $W$. This can be achieved if we move from $W$ to $W_0$, which has not a set like $q_i(t)$, but one fuzzy value, and this value is $q_0(t)$ - the most reliable, and therefore has the maximum value of the membership function in the set $q_i(t)$. In this way,

$$q_0^0(t) = \max_{\mu_{ki}} \{q_{ki} \}.$$ 

This gives us:

$$W_0 = \{v^0(t_1), v^0(t_2), \ldots, v^0(t_k) \},$$

$$v^0 = \{g_1^0(t_1), g_2^0(t_1), \ldots, g_k^0(t_1) \}.$$ 

Fuzzy time series $W_0$ is called the zero base time series, and for it we can build $\{F^0_i(t) \}_{i=1}^k$ based on the principles given above. On this basis, we will have a visual image for $W_0$, which we denote by $F_0$. Consideration of the image of $F_0$ suggests that this is a fuzzy image and its properties are a consequence of the $W_0$ properties. On the other hand, feedback is also possible, i.e. the properties of $F_0$ can be judged on the $W_0$ properties.

With sufficiently consistent development of time series, its changes are not relatively large and the resulting image $F_0$ has the properties of $W_0$. We can consider $W_0$ to be a rigid "framework" $W$, and, accordingly, $F_0$ will be a more stable construction than $F$. In visualization this will manifest itself in by changing $W$, the zero base time series $W_0$ will not change unless changes in $W$ touch the components $W_0$. This circumstance will appear on the image $F_0$.

If all the $W_0$ components are excluded from $W$, we can obtain $W_1$ – the first baseline time series and match it with $\{F_1^1_i(t) \} = F_1$. Transformation $F_1$ will be less resistant to changes in $W$ and will show a more variable picture. When analyzing the image $F_1$ we can set wider variation limits for images $\{F_i^1(t) \}$. The additive nature of transitions is preserved due to the linearity of the transformation (5) $F_0, F_1, \ldots, F_{n-1}$.

At the current stage we are trying to implement this approach to build a decision support component helping us to decide which linearly dependent factors should be excluded from consideration (or saved) while building the regional socio-economic development model. Initial statistical data can be threatened as a nonlinear fuzzy time series [9, 10].
4. Sections, subsections and subsubsections
It is possible to solve a number of problems leading to a necessity to discover the analytical form of the statement on the basis of visualization. These are tasks such as segmentation, clustering, forecasting - such transformation is being is being used right now in our project first of all, we expect numerical results, which appears with linguistic accompaniment, reflecting the qualitative aspects of a subject area.

Let us especially note that the stated approach is not connected with the existing mathematical model of the system for which the fuzzy time series is being analyzed. Hence it can result in the low accuracy of forecasting, the lack of criteria for the quality of a fuzzy description, which significantly limits fuzzy modeling of time series. This circumstance determines the new approach to the analysis of series, which are characterized by a high degree of uncertainty. This is relevant when considering the dynamics of weakly structured systems when identification of the model class is impossible and therefore obtaining high-precision models is difficult.

Acknowledgments
The work was supported by the RFBR, and conducted by joint efforts of two research teams under project № 18-07-00543 and project № 18-37-00344.

5. References
[1] Mofidi R 2007 Identification of severe acute pancreatitis using an artificial neural network Surgery 141(1) 59-66
[2] Polat K, Akdemir B and Gbne S 2008 Computer aided diagnosis of ECG data on the least square support vector machine Digit Signal Process 18(1) 25-32
[3] Zaripova G R, Bogdanova Yu A, Kataev V A and Khanov V O 2016 Advanced models of expert decision support systems to predict the operative risk during the surgery Tavricheskiy Medical and Biological Bulletin 19(4) 140–145
[4] Gubler E V 1978 Computational techniques of analysis and recognition of pathological conditions (Leningrad: Nauka)
[5] Berestneva O G, Volovodenko V A, Gerget O M, Sharopin K A and Osadchaya I A 2013 Multidimensional Medical Data Visualization Methods Based on Generalized Graphic Images World Applied Sciences Journal 24 18-23.
[6] Volovodenko V A, Berestneva O G, Sharopin K A. 2011 Visual interpretation of quantitative characteristics of Biosystems Fundamental medicine: from scalpel toward genome, proteome and lipidome (Russia, Kazan, KSU), pp. 126-129.
[7] Pospelov D A 1986 Fuzzy sets (Russia, Moscow: Science).
[8] Leonenkov A V 2005 Fuzzy simulation in MATLAB and fuzzy TECH (Russia, Saint-Petersburg, : BHV-Petersburg).
[9] Marukhina O V, Rogozhkina I D, Bobrova M V, Kozlova N V and Berestneva E V 2016 Informative predict system of educational activity of the first-year students Advances in Computer Science Research 51 116-119
[10] Mastitsky S E, Shitikov V A Statistic analysis and visualization of data by R http://r-analytics.blogspot.ru/