Revealing of Entities Interconnections in System Dynamics Modelling Process by Applying Multimodal Data Analysis Paradigm

Olga Kalyonova, Ivan Perl
ITMO University
Saint-Petersburg, Russia
ovkalyonova, ivan.perl@corp.ifmo.ru

***Abstract*—Nowadays System Dynamics is one of the most popular modelling approaches because it provides instruments for simulation and understanding of complicated processes in a wide ranch of application domains. In consequence of this, modellers, who are working in this area, are dealing with big number of entities both, common to the modelling approach and specific for their particular application area. By its nature, system dynamics approaches allow to investigate overall behaviour of the system or process. In the same time many of relations between particular model nodes remains unclear. Described multimodal approach allows to dive deeper in to the insights of each particular model participant and helps to attempt to reveal inner effects make to each other.**

I. INTRODUCTION

Nowadays, in the era of a large amount of information and data, growing demand to not only have the data available as such, but also to be able to organise and present them. There is large number of heterogeneous sources that store the required data, however, because of the complexity of organisation, linking and aggregation are becoming a challenging tasks. All such data sources and the data itself can in general be described by applying a multimodal data paradigm.

A. Problems

There are a lot of different data sources providing lots of data about same objects.

Terminological approach to definition of multimodal data is poorly structured and expressed. The concept of multimodal data found in many scientific domains, and its often uses a different names. Root cause of this is that in each discipline special terms are used to describe a specific application. However, from the mathematics point of view all these terms can be converted to multimodal data. For example, organoleptic [8] - science that describe the understanding process by sensory organs: vision, hearing, smell, etc.

From the other side, multimodal data are also engaged in other scientific domains: modal logic [7], [4] and set theory [9]. In this paper, proposed a generalised approach for the concept of multimodal data.

As noted above, the phenomenon of multimodal data has a wide distribution and, on this basis, there are several different instrumental approaches for its practical use, including: semantic web and ontology [10], object-oriented approach [6], cognitive modelling [5], [6], topic maps [11], data matrix and truth tables [4].

At this moment methods of multimodal data presentation allowing extraction of new relationships in the data set which does not exist yet. This will increase the connectivity of the original data sets and will allow better organisation of search among large sets of data and will allow better patterns identification. Thereby increasing the efficiency of work with data sets.

To show application of proposed approach and support assumption about its efficiency, in term of this article, we will address several issues from system dynamics modelling area, including following:

1) A long way from making a decision on the necessity of studying and/or improving the system or process to real improvement.
2) The result of a simulation is a series of numeric values and not informative by themselves.
3) The simulation results do not make clear the structure of the system: do not help to detailize or simplify it.
4) A large amount of data requires special approaches and methods of analysis and knowledge from the analyst and the expert of the subject area.
5) Refine, optimisation and improvement of both model and real system or process depending to a large extent on the knowledge of an engineer and conversation skills of a subject area expert.

B. Goals

As mentioned above, currently there is no method of multimodal representation of data that allows detection relationships in the data set. Identification of new relationships between data entities is an important task when working with any data. This is the main task that data scientists should solve. Internal relationships inside the data set help to identify patterns, but to do this without representations of the multimodal data is almost impossible. It is obvious that patterns better identified when a large coherence of data available. Thus, we come to the fact that it is necessary to arrange linked access to multimodal data and to provide the opportunity to discover new correlations in the source dataset.

To achieve this goal, it is necessary to understand the definition of multimodal data, and to form its common and correct definition. Based on the definitions it will be possible to develop a method of representing multimodal data for coherent access.

II. INFORMATIONAL MODEL

A human's memory is arranged in such a way that thought about one object is produced thoughts about other objects. This phenomenon is described by associative chains and networks. Based on this we can say that the results of the methods of presentation should be a set of multimodal objects associated with each other.

The connection between multimodal objects can be both direct and indirect. Direct connections between multimodal objects may be called direct associations and direct links between simple representations. While indirect connection occurs through associative connections between the essential concepts related objects. Direct links can be specified from the outside, while using the developed method there is the possibility of identifying indirect relationships, even if direct links between objects are not known in advance.

The developed method can calculate the coherence power, and both direct and indirect relationships of multimodal objects. The coherence power w cannot be more than the specified number of modalities m.

In the structure of multimodal data, it is possible to identify an abstract object and its entity presentation in each modality from 1 to m.

It should be noted that not all abstract objects in the input data array can have representations in any given modality, in other words, the essential presentation of the abstract object of the given modality can be ignored. In some cases, it can, in consequence, to extrapolate the values of the nearest neighbours. The set of all abstract objects can be represented as a graph $G(V^n, E^k)$, where V^n - is, the vertex set of the graph or of the set of abstract entities [1], whose power is equal to n the number of abstract objects. E^k is a set of edges between abstract objects, whose power is equal to k the number of link between abstract objects.

The connections between the elements of the section of the essential presentation related to the same abstract object can be excluded as they can be obtained by linear combinations of relations of refund (See Fig.1). Data matrix the essential presentation can be described as D_m.

The relationship between the AO in the graph structure:

$$r_v = \{r_{v_1}, r_{v_2}, , r_{v_k}\} \tag{1}$$

The relationships between elements of the section (data relating to one AO):

$$r_s = \{r_{s_1}, r_{s_2}, , r_{s_{m(m1)}}\} \tag{2}$$

The refund relationships are provided (between the data and the modality of the AO to which they relate):

$$r_{dv} = \{r_{dv_1}, r_{dv_2}, , r_{dv_{mn}}\} \tag{3}$$

The relationships between data of one modality

$$r_{dm} = \{r_{dm_1}, r_{dm_2}, , r_{dm_{mn(n1)}}\} \tag{4}$$

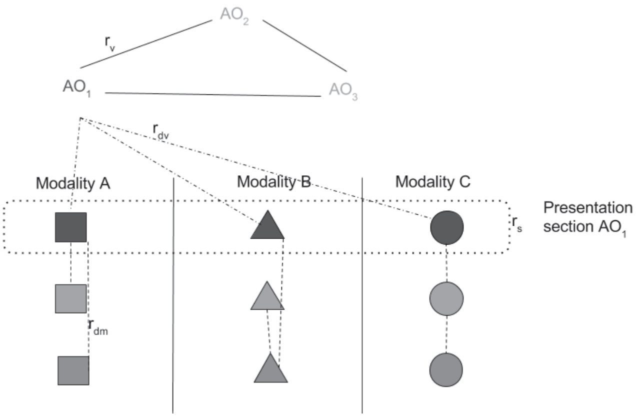


Fig. 1. The structure of multimodal data

III. METHOD DESCRIPTION

A. Input data set

In the beginning, we have disparate data sets or one data set, but the stored data are fundamentally different from each other. Take such sources or types as the modalities, the number of which we denote to m. Let us denote data of one modality X_i tuple of values of one modality, where $i = 1..m$. We denote respectively $X^m = \{X_1, X_2, ..., X_i, .., X_m\}$. In the same time, it is known that the data in the modalities describe a certain number of abstract objects (singular terms). We denote this set as $V^n = \{V_1, V_2, .., V_n\}$. Then you can record, $X_i = \{x_{i_1}, x_{i_2}, .., x_{i_n}\}$

B. Classification of modality data

The second stage of a method of representing multimodal data is to classify the tuples in the input data of the essential presentation of X^m on modalities.

The classification at this stage is necessary for the detection of relationship entity data submissions on the modalities among themselves.

In the future, we will need to know the extent of such relationship. In this regard, it is proposed to build on the original data graphs of hierarchical modalities [2]. The advantage of this task is given to machine methods of classification, when possible. However, there are situations when you need an expert evaluation. In this case, it is suggested to use another method of classification. We denote the obtained at this stage, the hierarchical graphs of modalities such as Y^m. Then, a classifier is a function to convert $X^m \rightarrow Y^m$ and $\{X_1, X_2, .., X_m\} \rightarrow \{Y_1, Y_2, .., Y_m\}$.

In some cases, the values of entity representations in the modality may not be unique and set to a linguistic variable. For example, such a modality is m, and the set of values of a linguistic variable $\{a, b, c\}$, then a hierarchical graph of the modality m can be specified as follows (See Fig. 2).

C. Calculation of height of a common parent

At this stage, based on a hierarchical graph modalities Y^m and tuples of input values is consistent with a pairwise search

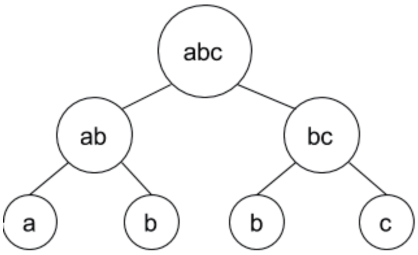


Fig. 2. Hierarchical graph for modality where value from linguistic variable

for the nearest common ancestor of all Y_i where $i = 1..m$ max-min strategy. Under the height, you can understand the path to a common ancestor. For example, searches for a common ancestor for y_{i_a} and y_{i_b} of modalities i, i.e. the graph in a hierarchical modality Y_i. Their common ancestor is $y_{i_{ab}}$. Following scenarios were identified for calculation of a common ancestor height:

1) Common ancestor exists and is the same distance from y_{i_a} and y_{i_b}. Then the height of the common ancestor $h_i(a, b) = d_i(a, ab) = d_i(b, ab)$.
2) The height of the common ancestor from y_{i_a} and y_{i_b} is not equal. Then: $h_i(a, b) = maxd_i(a, ab), d_i(b, ab)$.
3) One of the specified nodes is the limit for the other. Suppose $ab = a$, then $h_i(a, b) = d_i(a, b)$.
4) If the set of values of the essential presentation is not unique, it may happen that the desired high-speed performance equal, i.e. $y_{i_a} = y_{i_b}$ then: $h_i(a, b) = 0$.
5) Does not exist or more than h_{term}. The search for a common ancestor can be limited with the parameter h_{term}. If the path length has become greater than or equal to a terminal, and a common ancestor has never been found, search terminates, and the distance $h_i(a, b) = \infty$.

It may be a situation when you want to remember or choose only those with the height of the common ancestor in the modalities does not exceed the incremental value of h_{max}. Then in the results table, H_m is entered, only those, which $h_m \leq h_{max}$.

D. The calculation of forces and coherence power

At this stage, based on H^m and V^n are consistent with the calculation of the characteristics of the supposed link between AO (singular terms from the set V^n) the coherence power w.

Under the coherence power we will understand the characteristic of link between AO, which is equal to:

$$w^h = \sum_{i=1}^{m} \begin{cases} 0 \frac{h}{h_i(a,b)} < 1 \\ 1 \frac{h}{h_i(a,b)} \geq 1 \end{cases} \tag{5}$$

Values of w^h that satisfy the threshold condition of $w_{min} \leq w_h(a, b)$, is added to the table the coherence power W, and, thus, shows the desired link.

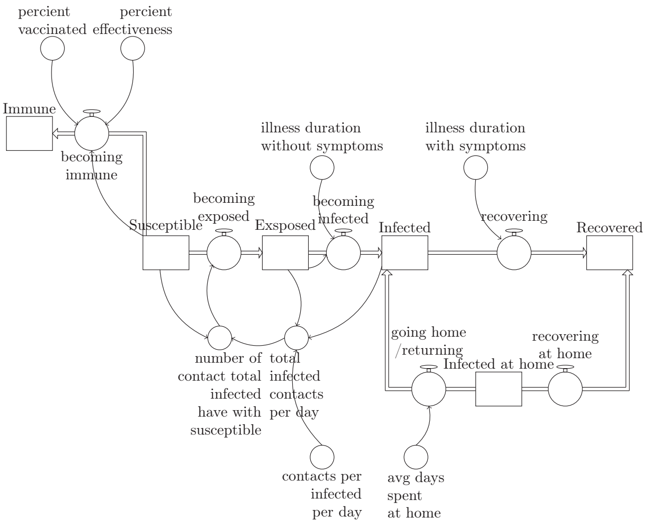


Fig. 3. The system dynamics model of the outbreak of H1N1 influenza

Under w_{min} will understand the minimum value of coherence power which is necessary for establishing link between V_a and V_b. The link is added to the set of R_v.

Then we get a graph of abstract objects G whose vertex are the objects of the V_n set and the edges are the objects of R_v set.

IV. FEATURES MODELLING DATA IN SYSTEM DYNAMICS

In the structure of the system dynamics modelling data we can distinguish several various levels of multimodal data.

1) In the simplest case, as abstract objects are selected stocks of system dynamics models without accounting for flows. As modalities can be selected formulas, graphs, titles, etc.
2) On the second level, the structure of multimodal data, the system dynamics becomes more complicated. As essential presentations of abstract objects, stocks, there may be sets of values, obtained by the execution of system dynamics models. You need to pay attention to the fact that in this case, all abstract objects (stocks) have communication of the essential presentations in the external in relation to this modality - time. Thus, time, not being the modality of the production dataset, and binds them with a basic knowledge in one or several modalities all the essential presentation.
3) On the third level as multimodal data can be selected not only stocks, but the flows described in the structure of system dynamics models. In this case, we deal with multimodal data of the second order, because flows are connections between the stocks and at the same time, flows are already multimodal data. The modality of stocks and flows in the general case differ.

As an example, consider the case of the first level. As a model of system dynamics was chosen model of the outbreak

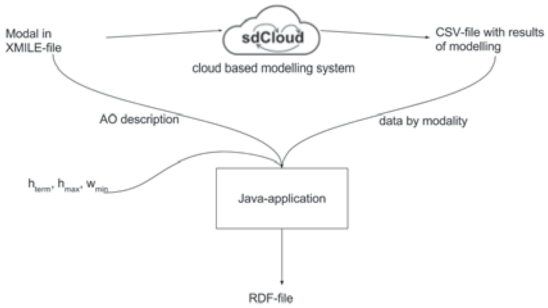


Fig. 4. Architecture of the program software solution

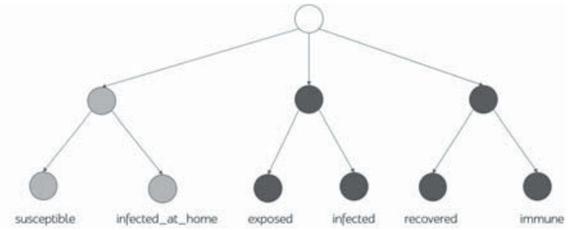


Fig. 5. The hierarchical graph of modality of changing value in stocks

of H1N1 influenza (See Fig.3). According to the rules of building system dynamics models [12], on the diagram, the rectangles indicate the stocks of model, and the thick arrows flows.

To obtain not only model defenition itself, but also model execution results, models execution engine was required. There is a certain amount of tool providing such capabilities, including: pySD, SDEverywere, sdCloud and others. For that particular work cloud-based platform sdCloud was used, because it allows to perform models execution in the cloud without setting up any kind of system dynamics environment locally and also allows to use various models executors like PySD or SDEverywhere.

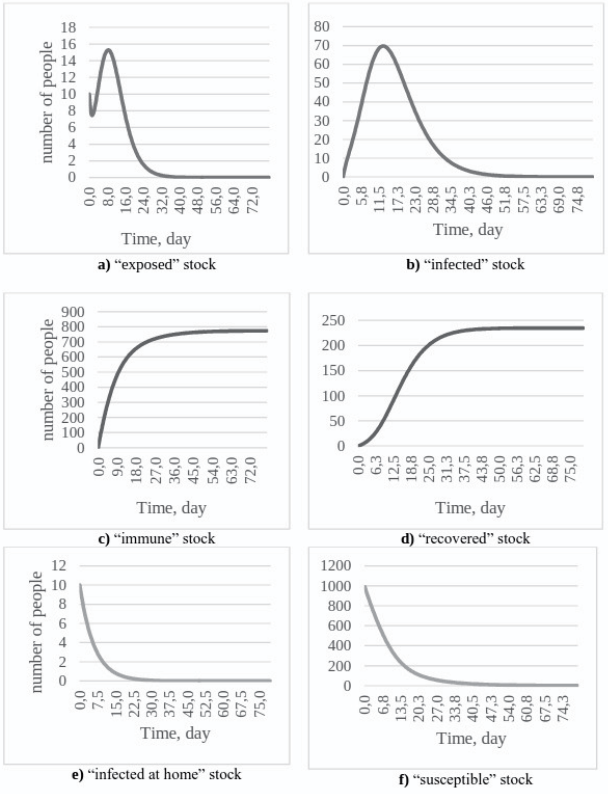
As a data source for set V^n was used a model file in system dynamics XMILE format the same file has been uploaded to the cloud-based simulation system sdCloud (See Fig. 4) [13]. Modelling results were obtained in the CSV format with the values of stocks in each time point of the simulation. Based on the modelled values of models stocks set of chars were build for each of them. (See Fig. 6).

A. *Method implementation example*

To show how poposed method of multimodal data defenition and representation can be applied, let's walk through an example based on a this system dynamics model.

According to the method definition, we need to perform following number of actions:

1) First, we need to run the model using one of the available model execution tools (i.e. sdCloud) with two different modelling steps: h and h/2.



a) "exposed" stock

b) "infected" stock

c) "immune" stock

d) "recovered" stock

e) "infected at home" stock

f) "susceptible" stock

Fig. 6. Function of value changing in stocks

2) The next step is to parse model file (in our particular case in XMILE format) using an application, implementing proposed meth. At this point, application will extract information about model stocks. Stocks names or ids will be used as labels for the future multimodal objects. At this point essential representation of the labelled objects is not identified yet.
3) After processing model file and identification of future multimodal objects, we can start filling them with actual modalities data. To do so, we will upload the result of first model execution with modelling step h to the application (modelling results are represented with CSV file). In the header of the results table located names of stock to which particular column of values relates. Based on matching column names with multimodal entities labels extracted from the model file, these two entities are mapped to each other. Now we have a number of multimodal objects, labelled with stock names with added modality containing the list of the stock values during model execution.
4) Processing number series as an essential representation of some entity is a common, but not a trivial thing. To make is easier for the particular example for series approximation spline fit method was used. Spline fit configuration can also be treated as modality and added to the particular stock.
5) Based on number series and characteristics of approximating polynomial charts for each particular stock are created. These charts are shown on a Fig. 6. Let's

assume these charts as a key modality of the stocks.

6) Taking into account charts built for number series and approximating polynomials we are generating a hierarchical graph for essential representations for the key modality. This graph will represent kinship of functions for changing values in stock and it is shown on a figure 5.
7) In the same way as for the first modelling results, we are processing results for the second model execution with modelling step h/2. Since model structure remains the same, we are getting new number series for the same stocks and mapping them in the same way as it was done in 3.
8) For the new numeric series, we will compute the modelling displacement. For that purpose, we will subtract from the new modelling, with the smaller modelling step, results of the initial execution with bigger modelling step and after that we will get the module of the resulting values, since we are interested in an absolute value of the displacement. This approach is called Runges method and frequently applied for this kind of operations [3]. Number of elements in a number series containing displacement values will twice less than in the second model execution results and equal to the number of the results in the first model execution experiment.
9) Now we will perform the same number of manipulations on number series containing displacement values as it was described on step 5 and we will user resulting charts of for displacement as a second modality for each stock.
10) Following the same approach as it is described on step 6, we will create a hierarchical graph for the second key modality, which will describe kindship of the functions of changing displacement between stocks.
11) Having all this data about stocks as multimodal objects, we will create all possible paired combinations of stocks and put them in the first column of resulting table. To make it easier readers, stock names were shortened.
12) For each combination of multimodal objects (stocks), we will compute the height of the hierarchical graph based on the first key modality and will define a height of their common parent element. For example, charts depicting values for stocks 'exposed' and 'infected' are similar, hence they are using different numbers sets. Then, according to the hierarchical graph for the first modality, height of their common parent node will be equal to one, where common representation of such looking function located. In the same time, charts, representing values of stocks 'exposed' and 'immune' look completely different and their common parent node height will be 2, where this common parent saying just that both multimodal object have some charts. Taking into account symmetry of the hierarchical graph for the first modality in current example, there is not cases where height of a common parent element for the given pair of essential representations of multimodal entities varies, but if such things happen, from the two options the maximal should be taken into account.

TABLE I. CALCULATED VALUE OF PARENT HEIGHT AND COHERENCE POWER

Stock pair	Function of value	Function of error	w1	w2	w3
imm - exp	2	1	1	2	
imm - susc	2	2	0	2	
imm - infec	2	3	0	1	2
imm - ininh	2	1	1	2	
imm - recov	1	2	1	2	
exp - susc	2	3	0	1	2
exp - infec	1	1	2		
exp - ininh	2	1	1	2	
exp - recov	2	2	0	2	
susc - infec	2	2	0	2	
susc - ininh	1	1	2		
susc - recov	2	1	1	2	
infec - ininh	2	1	1	2	
exp - infec	2	2	0	2	

13) Values, obtained on the previous step are placed in the second column in the results table.
14) Now we will repeat all the manipulations performed in steps 12 and 13 for the second modality based on displacement data and are put to the column number 3 for the same pairs of multimodal data.
15) In column number 4 of resulting table we will put a relationship power between multimodal objects for which value of common parent node height not exceeds 1. In this particular example we are not taken into account cases when height of the common parent element in the graph equals to zero, because values of all essential object representations are unique and there are no duplications. If there were equal values in some modalities, height of their common parent will be equal to 0. Value of the relationships power is calculating as number of '1' in columns 2 and 3 for the given row. In the current example, computation of relationships power is based on an assumption that both modalities are equal for the analyst. If importance of modalities differs weight coefficients can be applied to them. This approach is widely used in the decision theory.
16) In a column 5 we will put number of '2' in columns 2 and 3 for each row, which is added to the value obtained on the previous step in column 4. Alternatively, this can be calculated as a number of '1' and '2' in columns 2 and 3.
17) We can continue computation of relations power, by increasing a height of a common parent for the items, until it became equal to the number of available modalities. Besides this, maximum power of height of common parent can be limited by analyst.

The strongest links in the table are found by minimising the elevation values common ancestor h in w_h in the column header and maximising the value of w_h in the rows. In the described example, the strongest connections are the following:

1) Exposed Infected
2) Susceptible Infected at home

We can see that first strong link is well mapped to the actual model structure, where stocks "Infected" and "Exposed" are directly connected with the flow. Another pair: "Susceptible" and "Infected at home" are located relatively far from each other, but in the same time, described method shows the strong correlation in their behaviour, what may cause closer attention from the analytics to find out why this particular stocks behave this way and maybe, understanding of this will lead to the model improvement where less obvious link between these two nodes can be taken into account and this nodes will get closer relationships in an enhanced model structure.

Thus, using the described method of representing multimodal data based on the described information model we obtained all the possible links between stocks, chosen as multimodal entities in terms of modalities obtained as a result of transformation of the simulation data. All links have their own weight, and, consequently, the value for the analyst.

In different situations for different models and subject areas, meaning of similarity of the behaviour of flows may varay. The task of clipping, too weak relations remains with the analyst that he can do that by setting the required parameters.

Special attention analytic should give to the fact that rash the incorrect method parameters when working with large models or other multimodal data can lead to significant loads on the equipment used, as well as a significant time-consumption, because the maximum number of connections is $n(n-1)$.

V. CONCLUSION

Most of the data now can somehow be represented as multimodal data. This is especially important when data for multimodal data represented in different data sources. Identify relationships and dependencies in data, an important task for data analysts and the domain expert. The developed method allows to represent multimodal data to find and calculate such relationships. The identified communications can significantly simplify the task of analysing multimodal data.

This paper describes new approach for multimodal data presentation applying to modelling data of system dynamics.

System Dynamics is widely used for analysis of different processes and complexity of target domains can be high. Complicated applications along with detailed models may lead modelling to the point where model will allow prediction of system behavior well, but system internal processes and dependencies will completely buried undre large amount of available details. Method, proposed in this article allows to analyze model along with its modelling process to extract more relevant information about inner dependencies and relationships within analyzed system.

Application of multimodal data paradigm to system dynamics modelling area was not described before, which emphasises the novelty of the material under study and the results presented.

REFERENCES

[1] Kalyonova O.V. *The method of multimodal data presentation for data of colour and colour combination* // Anthology of scientific works of young scientists of ITMO University. Volume2. -2016. pp. 253 255.

[2] Kalyonova O.V. *Applying of the method of multimodal data presentation* // The book of abstracts of the Congress of young scientists / V Russian Congress of young scientists. Digital edition. -2016

[3] Gavrilyuk, I. and Hermann, M. and Makarov, V. and Kutniv, M.V. *Exact and Truncated Difference Schemes for Boundary Value ODEs* Basel : Birkhauser, 2011.

[4] Blackburn, Patrick, Rijke, Maarten de, V. *Modal Logic* (Cambridge Tracts in Theoretical Computer Science) Published by Cambridge University Press, 2017

[5] Jeff Hawkins, Subutai Ahmad *Why Neurons Have Thousands of Synapses, a Theory of Sequence Memory in Neocortex* // Frontiers in Neural Circuits, 10 (2016) 113, http://journal.frontiersin.org/article/10.3389/fncir.2016.00023

[6] Xinyu Li, Yanyi Zhang, Jianyu Zhang, Yueyang Chen, Shuhong Chen, Yue Gu, Moliang Zhou and Ivan Marsic, Richard A. Farneth and Randall S. Burd *Progress Estimation and Phase Detection for Sequential Processes* // IMWUT/Ubicomp 2017

[7] Christian Mahesh Hansen and Roger Antonsen and Martin Giese and Arild Waaler *Incremental variable splitting* // Journal of Symbolic Computation vol. 9, 2012 pp.1046 - 1065

[8] D. Valentia, G. Denarob, F. Giarratanac, A. Giuffridac, S. Mazzolab, G. Basiloneb, S. Aronicab, A. Bonannob, B. Spagnolo *Modeling of Sensory Characteristics Based on the Growth of Food Spoilage Bacteria*

[9] Marek Balcerzak, Paolo Leonetti *On the Relationship between Ideal Cluster Points and Ideal Limit Points*, 2017

[10] Robert B. Allen *Rich Semantic Models and Knowledgebases for Highly-Structured Scientific Communication*, 2017

[11] Stefano Gurciullo, Slava Mikhaylov *Topology Analysis of International Networks Based on Debates in the United Nations*, 2017

[12] J. W. Forrester, *The Beginning of System Dynamics*, Cambridge: MIT, 1989.

[13] Mulyukin, T. Kossovich and I. Perl, *Effective Execution of Systems Dynamics Models*, in 19th Conference of Open Innovations Association FRUCT, University of Jyvaskyla, 2016, pp. 358-364.

[14] *sdCloud - Bringing System Dynamics Into Cloud* URL: http://sdcloud.io/

[15] J. Houghton, M. Siegel, *Advanced data analytics for system dynamics models using PySD*, 33RD INTERNATIONAL CONFERENCE OF THE SYSTEM DYNAMICS SOCIETY, vol. 2, 2015, pp. 1436-1463.

[16] GitHub website, Source code of PySD by James Houghton, Web: https://github.com/JamesPHoughton/pysd/

[17] GitHub website, Source code of SDEverywhere by Todd Fincannon, Web: https://github.com/ToddFincannon/SDEverywhere

[18] Perl I. A., Ward R., *sdCloud: Cloud-based computation environment for System Dynamics models*, Proceedings of the system dynamics conference, 2016.