# Introduction to Multimodal Data Analysis Approach for Creation of Library Catalogues of Heterogeneous Objects

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

***Abstract*—Nowadays, the amount of both the data themselves and their sources is constantly growing. However, such diversity often leads to heterogeneity and fragmentation of data, which makes complete analysis difficult. At the same time, the multimodal data approach is able to address these issues and not only to find new links in different data sets, but also to calculate the coherence power of their proximity. The approach described in the article is extremely important for the organisation and preliminary analysis of data, but it also can be used as an independent method for various data analysis activities.**

## I. Introduction

Designing a system which is dealing with data representation or processing, whose sources can be heterogeneous and, moreover, representing significantly different objects may appear a challenging task. One of the most important aspect will became a selection of a proper method for data organisation and for entities relationships identification and definition. For wide range of applications to address these set of issues methods of multimodal data processing can be used.

### A. Problems

Under multimodal object in this article we will understand an object or process, which is described with various characteristics of different modalities. Under modality we will on assume one of more attributes which describe a specific characteristic of an object. Set of characteristics of the same modality will be called essential presentation. It is not mandatory that each multimodal object exists in all modalities.

By their nature, heterogeneous objects are described with number of attributes, which are not compatible with each other, for example, colour, weight, price, and these characteristics can be treated as modalities of this particular object. When object has multiple modalities describing it, it can be called a multimodal object.

### B. Goals

The goal of this work is to define a method and definitions required for performing of analysis of heterogeneous data sets with non-trivial relationships between its entities. To achieve this following topic should be covered:

1) Determine a unified approach for defining multimodal objects based on real heterogeneous objects. It means that we need to provide a mechanism explaining how to represent any heterogeneous objects with its multimodal twin suitable for further processing and analysis.
2) Define an approach for representing relationships between multimodal objects, extracted from real objects under investigation.
3) Define an approach for processing multimodal data to find relationships and correlations between different multimodal objects and their modalities.

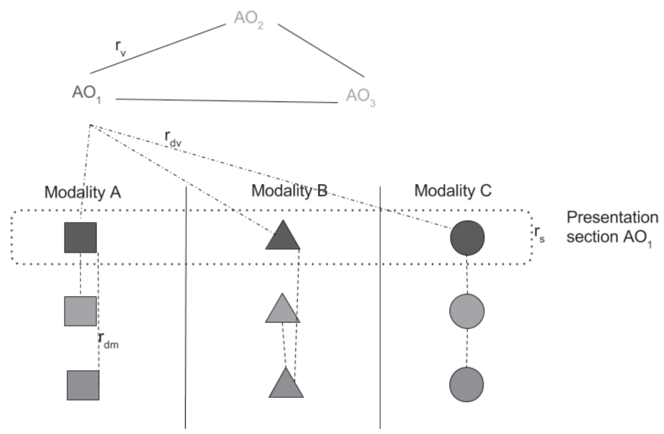


Fig. 1. The structure of multimodal data

## II. Basic multimodal data structure

At the top level, we have a set of multimodal objects. Each of the multimodal objects has a set of essential representations in different modalities. The number of modalities of multimodal objects will be denoted by $m$. Multimodal objects are connected between themselves by direct and indirect links. Direct links are given from the outside and are known in advance, and the mediated ones are calculated on the basis of the developed method of multimodal data representation, which is described in more detail in [1], [2], [3], and a summary is given in section III.

A simple structure of multimodal objects is shown in the Fig.1.

## III. Basic method of the multimodal data representation

This method consists of four main stages.

ISSN 2305-7254

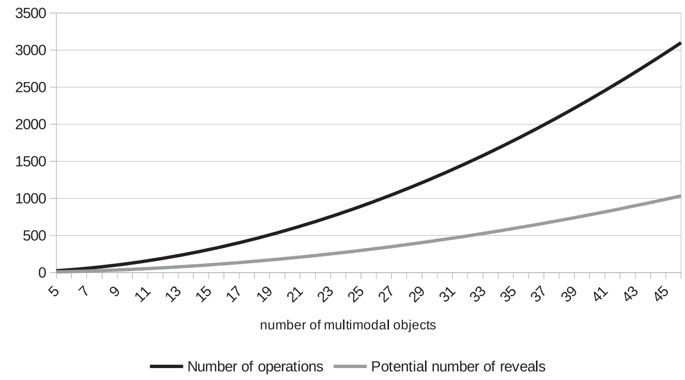


Fig. 2. Number of operation and potential reveals for full learning

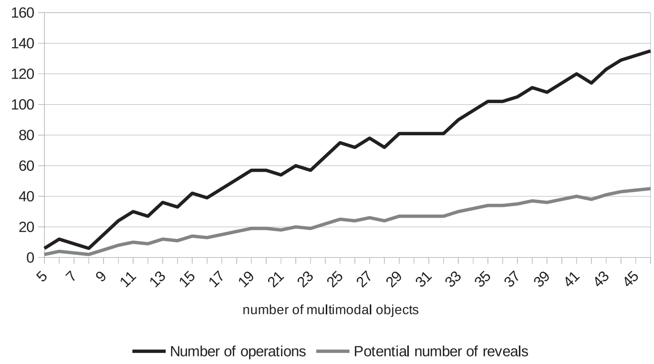


Fig. 3. Number of operation and potential reveals for incremental learning

*a) The input data sets:* At this stage, we have many input data divided by modality. The size of the modality determines the number of $n$ multimodal objects.

*b) Classification of data by modalities:* In the next step, the resulting data sets are passed to the classifiers according to the modalities. Note that the classification methods depend on modality and are not principal from the point of view of the whole method. However, the output format is strictly defined and must be a hierarchical graph of values in a given modality.

Classification is necessary to determine the degree of relationship between the value representations of different multimodal objects, which is performed at the next stage.

*c) Calculation of height of a common parent:* The result of the step 2 for each modality of graph is a hierarchical modality that defines reciprocity of various essential representation of multimodal objects. Denote by $h(a, b)$ the height of the common parent for multimodal objects $a$ and $b$. This value will be calculated according to the maximum strategy as the distance from each of the selected nodes ($a$ and $b$) to the parent node.

Beside it, additional parameters $h_{max}$ and $h_{term}$ can be entered. The parameter $h_{max}$ determines a maximum retained height value of common parent, and $h_{term}$ - the maximum height of the common parent to which the search is performed.

*d) Calculation of coherence power:* During this stage, the coherence power between all pairs of multimodal objects is calculated based on the heights of the common parents for all modalities. In the simplest case, this value is calculated as follows:

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

and $w^h$ is the coherence power, taking into account that the height of the common parent in each modality for the studied objects will not be higher than the preset value $h$.

This method can work in two main modes: full and incremental learning. The first mode assumes that the system will be trained once and will not change. The second mode, on the contrary, allows you to work with periodically added data without a complete re-training of the system. In this case, the number of checks increases depending on the number of multimodal objects as shown in the Fig.2 for full training and on the Fig.3 for incremental learning.

TABLE I. MODALITIES OF MULTIMODAL DATA KINDS

| Books | Articles | SD Models |
|---|---|---|
| Bibliographic description | Bibliographic description | Name |
| Tags | Tags | Tags |
| UDC | UDC | Description |
| Book file | Article file | Model file |
| URL of publisher | URL of publisher | Model structure |
| | | Model pattern |
| | | Modelling result |

## IV. THE SPECIFICITY OF HETEROGENEOUS DATA

A distinctive feature of heterogeneous data is the fragmented sources of such data and / or the fragmented components of the objects obtained from the data, including incomplete data. The reasons for this heterogeneity can be quite a lot; however, the problem as a whole is reduced to the heterogeneity of methods for analysing such objects.

One example of a heterogeneous data can be data obtained from various sources, including Internet resources. In this project, an attempt was made for the development of virtual knowledge library for system dynamics. In this library can be used for storage of the three main types of entities:

1) Books;
2) Articles;
3) Models of system dynamics.

A necessary condition is the ability to relate these entities to each other. Obviously, all three types of entities have a different set of parameters describing them (See Tab.I).

In addition to the entities types' heterogeneity, the modality of the sources of the library entities also takes place.

Due to the fact that in the projected library the data is collected automatically with the help of software agents, according to a predetermined but constantly expanding list of sources, the method for processing heterogeneous data must

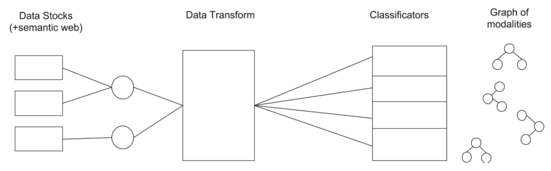


Fig. 4. Pipeline of application process of representation of multimodal data

be adaptive and be able to work with only a part of the characteristics.

Thus, the data aggregation scheme for the system dynamics knowledge library looks as shown in Fig.4.

We now describe in more detail the process of sequential collection and presentation of multimodal data, which by definition are heterogeneous objects of the main kinds of entities of the system dynamics library.

1) First of all, it is necessary to find and determine the list of data sources used. In the projected library, it is supposed to use the available Internet resources as data sources as well, labelled according to the rules of the Semantic Web. This is justified by the need to constantly update and update the library. In connection with the obviously necessary increase in the number of entities and data sources, updating the data with the help of a human operator servicing a virtual library becomes an unnecessarily time-consuming task.
2) To solve this problem, it is suggested to use special software-agents that check and index the specified data resources with a certain periodicity. If such an agent finds an entity that is appropriate according to its rules, then information about this is passed on to the process.
3) The next step is to clean and pre-process the data received from the software agents. For this, an analysis and additional checks of the received data are carried out.
4) All received essential representations of the added element (and in fact, the multimodal object) are transferred to the classifiers in accordance with the modalities. It is assumed that the hierarchical modality graphs have already been constructed for the previous elements in the system. Thus, when adding new data to the multimodal data representation system, an incremental approach is used. Then, when calculating the predicted necessary for this operation, we will use the formulas of the incremental growth of the system, rather than full training, because otherwise the system would be either static, or it would constantly perform the same work, getting the knowledge that had already been obtained earlier. Particular attention should also be paid to the fact that the classifiers being developed are primarily focused on the processed modality. Accordingly, the classification methods must be chosen based on the types of data being processed.
5) At the next stage, the heights of common parents are calculated and written between the essential representations already existing in the system and added. This procedure is performed for all modalities in which the value of the essential representation of the added element exists. Strategies for calculating heights can vary and be limited by the configuration parameters.
6) Based on the table compiled in the previous stage, a table of pairs of multimodal objects (library entities) is compiled or added with other entities of the same type (articles / books / models). In this table, the coherence power of two objects is calculated as the number of modalities in which the value of the heights of the common ancestors is not more than the given one. In the simplest case, when weights are not introduced for modalities, the number of modalities is the value of the maximum coherence power.
7) After this, the coherence powers are calculated and the search for similar elements for entities of other types, i.e. outside the catalogue.

TABLE II. MULTIMODAL DATA ON STRUCTURES: SIMPLIFIED CASES

| Book | |
|---|---|
| Bibliographic description | String - simple essential representation |
| Tags | Name from the classifier. Tags, a string of words, ordinary essential representations. |
| UDC | Value from the classifier without cross-references |
| File of book | Basic Text Formats |
| URL of publisher | URL or publisher name |
| **Article** | |
| Similarly | |
| **System Dynamic Model** | |
| Name | String |
| Description | String |
| Tags | similarly |
| Model file | Basic SD model formats: XMILE, mdl |
| Model structure | Parameter describing the structure of the model |
| Model pattern | Template number |
| Modelling result | Graph of changes in model values |

Table II shows examples of simple structures of multimodal data for sets of heterogeneous data of the system dynamics knowledge library.

## V. CONFIGURING THE WEIGHT FUNCTIONS

An important task for improving the results of the method is to configure and calibrate the main parameters of the method. Parameters $h$ and $w$ can be assigned to such parameters first of all.

The issue of configuring classifiers is beyond the scope of this article, because the process of classification depends more on the types and values of the data being processed. The question of choosing the values of the maximum height of the common parent for each of the modalities will also not be considered separately in this article, since it is more dependent on the resulting modal classification of the modality, and it is specific for each individual modality.

Most importantly, the result of the method will be influenced by the parameter $w$, since it depends on it the value of the bonds and the number of connections with such values, which is extremely important for analysing the obtained connections.

It is necessary to distinguish different coefficients for different types of connections between heterogeneous objects of the library, which are represented in the form of multimodal objects. Earlier in Section IV 6 such links were singled out: 3 internal ones - between objects of the same type and 3 external ones - between catalogues with different types of objects. The strategy for calculating the coherence power $w$ can be specified as follows:

1) As the number of modalities in which the values of the essential representations of the objects under consideration are close. (All the modal coefficients are, respectively, 1).
2) As the coefficients weighed by the analyst. Each modality has its own coefficient, the numerical value of the weight, the sum of which determines the maximum value of $w$, and to calculate the concrete connection it is necessary to sum only those coefficients in which the required proximity of the common parent is reached.
3) With the help of given functions, depending on the heights of common parent: $w = a_1(h_1)+a_2(h_2)+\cdots+a_m(h_m)$

The functions $w$ by types of links:

Book-book

1) $w_{max} = 5 = 1+1+1+1+1$
2) Let $w_{max} = 0.5+0.7+1+0.4+0.5 = 3$;
   $a_1 = 0.5; a_2 = 0.7; a_3 = 1; a_4 = 0.4; a_5 = 0.5$.
3) Let $a_1(h_1) = \frac{0.5}{h_1} \cdots$

Book-article: similarly

Article-article: similarly

Book-model (for identical modalities: name and tags):

1) $w_{max} = 2 = 1+1$
2) Let $w_{max} = 1.7 = 1+0.7$
3) Let $w = \frac{1}{h_1} + \frac{0.7}{h_2}$

Article-model: similarly

Model-Model:

1) $w_{max} = 7$
2) $w_{max} = 1+0.4+0.7+0.6+0.7+0.3+1 = 4.7$
3) $w = \frac{1}{h_1} + \frac{0.4}{h_2} + \frac{0.7}{h_3} + \frac{0.6}{h_4} + \frac{0.7}{h_5} + \frac{0.3}{h_6} + \frac{1}{h_7}$

For the test sample, a set of 90 heterogeneous objects was prepared. The number of heterogeneous objects of each of the types described above was 30. As a result, 2048 links were obtained, provided that the height $h$ is not more than 3, i.e. for $w^h = w^3$.

The graph in Fig.5 shows the obtained number of reveals by the coherence power in fractions of the maximum possible value for each function $w_{max}$. This graph was obtained for the entity type "model", consisting of 7 modalities. The graphs $w_1$, $w_2$ and $w_3$ show, respectively, the formulas by which the coherence powers were calculated. These formulas were derived from the empirical coefficients described above.

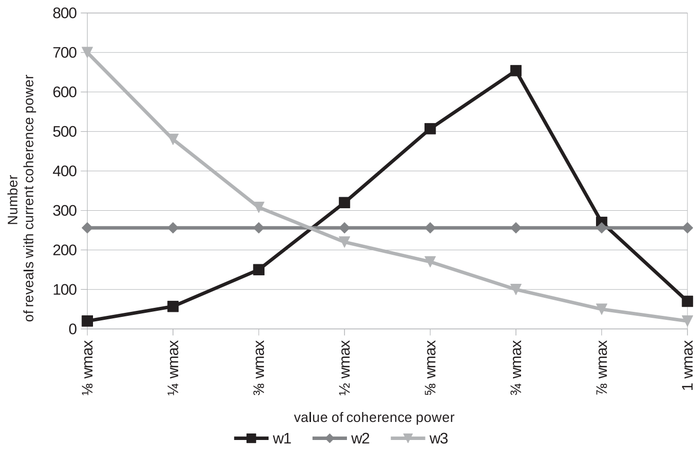


Fig. 5. Number of reveals by the coherence power in fractions of the maximum possible value for each function $w_{max}$

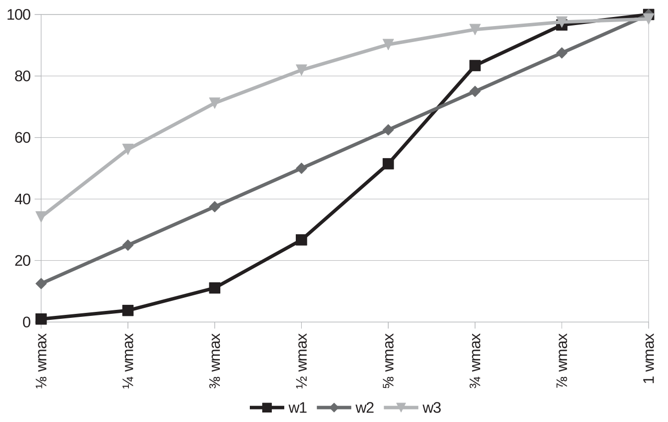


Fig. 6. Probability distribution functions from the value of the coherence power

The graph in Fig.6 shows the probability distribution functions from the value of the coherence power. These values are important to take into account when further processing of the results of the method, for example, when building a cores frame in the graph of multimodal objects.

The choice of the shape of the graph and, accordingly, the coefficients for the modalities is determined primarily by the tasks of further processing the results of the method. For example, in connection with the fact that based on the results of the method, a matrix of similarities between objects can be obtained and a matrix of distances between them can be obtained (based on the calculation of the coefficient inverse to $w$ - the greater the coherence power, the smaller the distance), then Cluster analysis should be applied. If the cluster analysis methods used assume a small number of clustering centres, then the objective function should be such that the number of reveals with the maximum coherence power is minimised. If the problem of uniform distribution is put, then a method of calculating $w$ using weighted coefficients should be chosen.

Most often in practice, there are Poisson distributions of one degree or another, depending on the parameter $h$ for $w^h$.

## VI. Future direction

In practice, we are often dealing not with one object, but with systems or collections of heterogeneous objects. When ordinary sets of multimodal objects are not enough for describing heterogeneous objects combinations, more complicated structures of multimodal data can be used. For that purpose following complicated multimodal data structure can be used:

1) Multimodal data with a links to external modalities
2) Multimodal data of higher level
3) Recursive multimodal data

Describing of this structures is beyond the scope of this paper and it will be given in the following articles.

Growing complexity of multimodal data structures increases the number of data processed. However, it becomes possible to represent complex heterogeneous data in such a way as to find new links in sets and the actual relevance of the found links between objects of different types (library entity types).

## VII. Conclusion

In this article was shown an approach for defining a multimodal object to represent heterogeneous real entities. To show real application of the proposed approach it was applied to analysis of system dynamics model and model execution results. In the described example shown definition of links between multimodal data entities as multimodal objects too. As a next step, also shown an escalation of multimodal data levels allowing to find correlations between all involved multimodal object disregarding were they real entities in the beginning or they represent links between them.

Going further, method was applied to the wider range of entities like articles and books related to system dynamics were included into analysis. Results of this extended mapping and analysis were applied for designing and implementation of a system dynamics library designed and builds by sdCloud project. Goal of this project is to create a library allowing to index and provide a linked access to wide range of information objects related to the system dynamics modelling like articles, books, models, model execution results and so on, and multimodal data approach was chosen as a most powerful and promising approach for indexing and linking data entities. Key focus of this research is increasing the effectiveness of the method via improving computation of the values of the weighting functions of the coherence power $w$.

## References

[1] Kalyonova O., Perl I. *Revealing of Entities Interconnections in System Dynamics Modelling Process by Applying Multimodal Data Analysis Paradigm* // Proceedings of the 21st Conference of Open Innovations Association FRUCT - 2017, pp. 156-161

[2] 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.

[3] 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

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

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

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

[7] A. 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.

[8] sdCloud - Bringing System Dynamics Into Cloud, Web: http://sdcloud.io/

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

[10] W. Attaoui, E. Sabir *Multi-Criteria Virtual Machine Placement in Cloud Computing Environments: A literature Review* ArXiv e-prints, 1802.05113, 2018.

[11] Nvea B. da Silva, Marcos O. Prates, Flvio B. Gonalves *Bayesian linear regression models with flexible error distributions*

[12] I.A. Burenkov, A.K. Sharma, T. Gerrits, G. Harder, T.J. Bartley, C. Silberhorn, E.A. Goldschmidt, S.V. Polyakov*Full statistical mode reconstruction of a light field via a photon-number resolved measurement* Phys. Rev. A 95, 053806 Published 2 May 2017

[13] Alibeigi M., Nili Ahmadabadi M.,Nadjar Araabi B. *Incremental learning of high-level concepts by imitation* ArXiv e-prints, 1704.04408, 2017

[14] Xitong Yang, Palghat Ramesh, Radha Chitta, Sriganesh Madhvanath, Edgar A. Bernal, Jiebo Luo *Deep Multimodal Representation Learning from Temporal Data* eprint arXiv:1704.03152, 2017