6. Osipov G.S. Optimization of single-channel queuing systems with an unlimited queue // Bulletin of Science and Practice. 2016. No. 9 (10). S. 63-71.
7. Sebeshev V.G. Features of work of statically indeterminate systems and regulation of efforts in structures. - Novosibirsk, 2009.-164 p.
8. Khachaturova K.R. Information technology as a means of development of creative abilities of primary school pupils in natural science lessons // Global scientific potential. 2015. № 9 (54). 111-113 p.
9. Masalovich, A. Neural network - the financier's weapon / A. Masalovich [Electronic resource] / Virtual computer museum. - Access mode: http://www.computermuseum.ru/histussr/neiro_net.htm (date accessed: 03/20/18).
10. Serdyukov, V. I. The use of artificial intelligence elements to improve the reliability of technical products / V. I. Serdyukov, N. A. Serdyukov, S. I. Shishkina // Bulletin of mechanical engineering. - 2017. - No. 10. - P. 29–32.
11. Serdyukov, V. I. Improving the failure-free operation of products using elements of artificial intelligence / V. I. Serdyukov, N. A. Serdyukov, S. I. Shishkina // Bulletin of the Moscow State Technical University. N.E. Bauman. Series: mechanical engineering. - 2017. - No. 1 (112). - P. 62–72.

About authors:

Samat B. Mukhanov, Master of Technical science, senior lecturer, Computer Engineering and Information Security, International Information Technology University.

Arman Ye. Alimbekov, Master student in «Data Science», tutor, Mathematical and Computer Modeling Department, International Information Technology University.

Gaukhar S. Marat, Master student in «Data Science», tutor, Mathematical and Computer Modeling Department, International Information Technology University.

Aibanu M. Uatbayeva, Master student in «Information Systems», International Information Technology University.

Assylkhan A. Aldanazar, Master student in «Computer systems and software engineering», International Information Technology University.

UDC 004.896

M.A. Mukanova, I.V. Krak, A.A. Kuandykov, A.S. Sagalova, D.A. Baibatyrov

International Information Technology University, Almaty, Kazakhstan

USING VISUAL ANALYTICS TO DEVELOP HUMAN AND MACHINE-CENTRIC MODELS: A REVIEW OF APPROACHES AND PROPOSED INFORMATION TECHNOLOGY

Abstract. *The use of a visual analytical system in machine learning is the basis for the integration of human and the use of his intellectual capabilities in the construction of models. At the same time, visual analytics is used to expand human knowledge and is used as a research tool. We investigate the forms and goals of using visual analytics workflow towards the formation of the final product. Workflow is divided into human-oriented and machine-oriented in order to build a model as an information processor and decision-making mechanism. Models are built on the basis of the end user, which can be either a machine or a human. The concepts of model building and the role of machines and humans in these processes are investigated. A practical implementation of the classification information technology in the studied concept 'using opposite model' in the machine-*oriented visual analytics workflow for using the machine model is proposed.

The basis for this model is a model formed and used by human. To classify data, human intellectual abilities are used. The boundaries of classes are determined by a human and then projected into a hyperspace of attributes with the formation of a classification model that the machine uses. Information technology allows the machine to use a model built for humans.

***Key words**: Visual Analytics, Classification, Mental Model, Formal Model, Dimensionality Reduction, Information Visualization*

1. Introduction

The visual presentation of information plays an important role in the resulting display of data analysis results. This method is the most informative and allows transferring a large amount of information for a human due to his physiological features of the perception of the outside world. The development of a simple presentation of information and analysis into structured analytical methods has allowed forming a concept of visual analysis. This concept provides analytical methods for the interactive interaction of a machine and a human. The goal is to maximize the use of the capabilities of the machine and human. A machine is able to present information in the form of data structures, summaries of results, graphic representations. A human is able to interpret the visual representation of data in the form of knowledge and analytical goals and transform the interpretation of information both in the form of intermediate solutions for the subsequent stages of analysis and in the form of final results. The evolution of information visualization is presented in the form of an interactive interaction between a machine and a human. The interactive visual presentation allows generating the resulting analytical data and aggregates them in the necessary way to summarize the data with emphasis on various aspects. They can also present analytical conclusions with presentation parameters. Analytical methods of transformation and information processing can be based on the methods of data mining and machine learning. The visual analysis enables a human to determine the importance of the data provided, determine patterns, and also discard data that does not deserve further analysis. It also allows to select and work with subsets of data, operate with parameters and determine their importance of influence. Visual analytics provides an iterative process of interaction and interpretation to determine different points of view for making optimal decisions at every step of the analytical process. The combination of the computing capabilities of a machine, the visual presentation of data as a link between a machine and a human, and the intellectual capabilities of a human allow improving the process of cognition of the surrounding reality. Since the analysis process is a dynamic process and a human interacts in some way interactively with a visual analytics system, often human-machine interaction systems are called “human-in-the-loop systems” Endert A. et al. [1], which found the greatest distribution in research.

We will consider the aspects of using visual analysis in terms of the effective use of man and precisely his intellectual capabilities for machine learning. Based on the analysis of the use of visual analytics in human-machine interaction, we define the aspects of human use in order to obtain the final product of this interaction. After defining the concepts and aspects of use, we will offer information technology that allows practically implement the technique of using a machine learning machine model formed by a human as a decision-making mechanism. The proposed approach is implemented in the form of visual analytics tools as a means of direct manipulation of data graphs to build models in the concept of human-machine interaction. The tool allows introducing a new aspect into the concept of using and building visual analytics workflows.

2. Related Work

2.1 Using the concept of visual analytics

The use of visual analytics is a dynamic process and is put into practice in visual workflow. The visual workflow consists of two types of components that are developed for the machine and for the human. These components are the epitome of combining the two machine-centric and human-centric concepts. Chen M. and Golan A. [2] provide a categorization of workflows in data analysis and visualization and expanded workflow analysis based on the information-theoretic framework for visualization Chen M. and Jänicke H. [3]. The division into six classes of workflows in data analysis and visualization, and identified four levels of typical visualization components allows defining the workflow concept for which the visual analytic tool is being developed. The specific parts of machine and human components for different workflows will be different. Four levels: disseminative, observational, analytical, model-developmental visualization Chen M. and Golan A. [3] determine the directions of two main areas of the workflow. These two directions flow from the concepts of centering: machine or human. From a workflow perspective, it's better to talk about processes and not about components. Visual analytics workflow is used to solve a specific task or series of tasks. The final consumer in using the workflow result is or machine or human. The levels of visualization of disseminative, observational, analytical are generally intended for a human as an end-user. The model-developmental level is used when forming a model for machines. The consumer of visual workflow results is central to the machine-human tandem. Therefore, two areas of workflow development should be identified: human-oriented visual analytics and machine-oriented visual analytics. This strict separation is important because it allows determining the object of consumption of the results of using visual analytics. If the main consumer is a machine, the visual workflow is used for computation of statistical indicators, supervised or unsupervised models classification, anomaly detection, prediction, features selections, comparative analysis and so on. The machine uses a human to solve the problem. A human helps a machine do its job better. The result of this interaction can significantly improve the quality of the result and determines the use of visual analytics. To automate processes, machine learning is actively used. Active participation of a human in machine learning processes allows the effective use of human intellectual abilities Amershi S. et al. [4]. Also in reference systems Jannach D. et al. [5], Pan W. [6], anomaly detection Zhao J. et al. [7], Cao N. et al. [8], McKenna S. et al. [9], use in Industry 4.0 Wu W et al. [10]. Using for diagnose the training process of tree boosting Liu S. et al. [11], diagnose and validate classifiers Krause J. et al. [12], Ren D. et al. [13], Alsallakh B. et al. [14], comparison and selection of clustering models Cashman D. et al. [15], Kwon B.C. et al. [16], selection of decision trees Mühlbacher T. et al. [17], selection and optimization of neural networks Strobelt H. et al. [18], Liu D. et al. [19], Kwon B.C. et al. [20] and so on. Visualization is used for the effective use of human intelligence. Machine learning is implemented by the machine and a human helps to do it better. Sacha D. et al. [21] proposed the ontology of VIS4ML, namely “visual analytic assisted machine learning”. VIS4ML allows to practically identify workflows in which visual analytics has been used to help improve machine learning. To determine the descriptive attributes by which the target motivation for using visual workflow can be identified, the paper defines the goals of using visual analytic for machine learning G1-G6. We can say that in this case, the human helps the machine because visual analytics assisted machine learning.

To obtain knowledge, cyclic interactions between computing processes, data transformation and visualization of content are used. van Wijk J.J. [35] Federico P. et al. [36], YALÇIN M. A. [37], Keim D. et al. [38, 39] described the visual analytics process as a process of cyclic interaction of a machine and a human in order to obtain knowledge from the transformation and visualization of data. The flow diagram of processes represents the interaction of computational methods, visual interactive interaction and data transformation, and as a result, is gain knowledge. In these studies, the consumer is the user of the results of the interaction processes. A human gains new knowledge by improving his intellectual abilities.

Visual analytics is a dynamic process that is implemented by workflow and is based on the interaction of human and machine. It uses processes to generate content and as a result of achieving the goal through visual analytics workflow.

2.2 Model concept in visual analysis

The definition ‘model’ is used in different interpretations depending on the subject areas. The concept of a model is used in various aspects of application. A human construct to help better understand surround real world systems Hestenes D. [40]. In the general case, the model can be considered as an information processor that has input data and output the expected result. Consider using the concept of a model in visual analytics. Very often, the concept of a model is determined by the context of its use. Booth P. et al. [41] describes the processes of interaction between human and machine in order to obtain a solution based on visual analytics. The survey summary presented systematization of models. Models are described for types, names, division of labor, information flow, elements and processes. The model is considered from the point of view of the process. Series of papers Booth P. et al. (Analytical Behavior Model) [41], Federico P. et al. (Conceptual Model of Knowledge-Assisted Visual Analytics) [42], Sacha D. et al. (Knowledge Generation Model) [27], Green T. et al. (Human Cognition Model) [42], Pirolli P. et al. (Notional model of sensemaking loop) [43], van Wijk, JJ (Simple model of visualization) [35], Lammarsch T. et al. (Visual Analytics Process) [44] consider the model as a process for obtaining a specific result - a solution. The models describe the processes of effective interaction between human and machine to help human in making the necessary and right decisions.

Andrienko N. et al. [45] considers the model as the result of visual analytic workflow, provides basic concepts and definitions when using Visual Analytics as Model Building. The concept of mental model is defined as a concept of a formed understanding by a human that can be described as a decision-making mechanism by a human. An important aspect is that the result of a workflow is a product for human consumption. Not a decision is formed, but a decision-making mechanism. This concept of the model is very similar to the concept of the model with the use of statistical and machine learning. A computational computer model is defined as a formal model. As visual analytics seeks to benefit from the close interaction of human and machine, the concept of a model has been analyzed and generalized. Conceptually, mental and formal models differ only in the consumer of these models - human and machine. In the future, we will use the definitions indicated by Andrienko N et al. [45].

3. Model usage generalization approaches

Considering the interaction of human and machine, it is necessary to focus attention not on the process but on the achievement of a specific result. The main role of using visual analytics workflow is to achieve a specific goal. This is presented in the form of a model, which is a decision-making mechanism. The model can be used by the consumer in the form of a human or a machine.

If a human wants to form a mental model, it is necessary to be based on initial knowledge gradually learning to form a mental model.A machine learning model must have the Interpretability Murdoch W.J. et al. [49], Zhao X. et al. [50]. Interpretability is ability of humans to understand the formal model and gain new knowledge to form a mental model. In general, Interpretability is desirable but not required. The main thing is to obtain the same results when using a formal and mental model.

Used when the analyst has an initial mental model that is adjusted in the cognitive process.

The formation of models is possible in two ways:

1. One of the models is fully formed and on the basis of it the necessary model is formed.

2. Formed parity formation of models. Models are formed gradually, mutually improving each other. However, the basic model is ahead of development and pulls up the necessary model. This option is the most used in complex problems and is characterized by iterative processes. Machine and human help each other in the formation of models.

- using opposite model. Initially, there is no basic initial model. To create the necessary model, the opposite model is used.

The formal model is the basis for the formation of human basic knowledge. There is a process of complete training and acquisition of knowledge by a human using a formal model. This is how the mental model is formed. The mental model is a local representative of the formal model, presented on the human side.

Forming a formal model using the mental is a complex process. To do this it's needed to somehow transform the mental model into a formal one. Make a projection of the understanding of human into the plane of understanding of the machine. An important factor is the lack of an initial formal model. The model is built but it is mental and the machine cannot use it. To do this, it is necessary to build a local representative of the mental model on the machine side or make a projection to the machine level or convert it to a formal model.

4. Human in visual analytics

Using visual analytics allows use of human intelligence and be useful to humans:

a) allows to effectively use the intellectual abilities of a human. Human intelligence is a resource that must be used effectively. One of the goals of visualization is to most effectively expand the capabilities of the system using a human. If we consider the end result, for example, machine learning, it can be improved by effectively integrating a human into the system using visual analytics.

b) allows a human to delve into the data. Using visual analysis, a human more and more understands different aspects of data, internal connections, structure, characteristics of signs. This helps to clarify the picture of the effective use of data capabilities. Visual analysis is carried out to solve the tasks. However, very often the analysis has a research character. In the process of analysis, new, previously unknown, patterns in the behavior of data can be revealed. New aspects of the analysis results lead to the correction of the tasks and decisions made.

c) analysis process allows to improve a human's qualifications. More deeply immersed in the data in the process of visual analytics, a human learns. Analytical work expands a human's qualification abilities in the field of data analysis, formats the aspect of data mining. In this case, we have two-way communication. For example, a human improves the model, and the process of improving the model teaches the human, while using elements of research work, increasing the level of human qualification. This process is beneficial for both the machine learning model and the human. Hall K.W. et al. [51] describe improving skills in both visualization and domain areas as results of immersion.

d) visual analytics is a dynamically expanding direction towards in-loop learning capabilities. Using visual analytics allows enhancing the skills of human analytics. A human with improved skills and acquired knowledge sees ways to improve visual analytics tools and improve visual analytics workflow in domain areas. This indicates a cyclical aspect of the evolution of human-machine interaction through visual analytics.

5. Generalization of the concept of development of visual analytics and the position of our work

In this section, we generalize the research of visual analytics and designate the provisions of the proposed information technology in the context of the proposed concepts.

The use of visual presentation has accelerated significantly over the past few years. One can observe the intensification of the integration of human-machine interaction. Many areas of visualization use concepts are close and often have much in common. The analysis of studies in the previous sections on the use of visualization in the interaction of a machine and a human is generalizing. The direction of generalization is based on the definition of the goals of using visual analytic workflow. Often the use of visual analytics is aimed at the process of learning new things, expanding knowledge, and the like. Another direction of using workflow is to get the result in the form of a solution. More promising is not getting the final product, not a solution, but a decision-making mechanism. The mechanism is implemented in the form of an information processor for decision-making and has the name 'model'. A decision-making mechanism or model can be built for both the machine and the human and is designated Andrienko N. et al. [45] as a formal and mental model. It should be noted that in the development of visual analytics, the concepts of human and machine are much similar as consumers of models. The concepts of the formal and mental model differ only in consumers - machine or human. A promising development is the construction of relationships between models and, on the basis of them, the development of new techniques for their use. This is a natural development of the processes of integration of human-machine interaction based on the developing capabilities of visual analytics. Moreover, a human is not a simple user but a necessary part of the system, benefiting both for a human and regarding the use of a human (Section 4).

A workflow designed to provide a decision-making mechanism can be oriented to a human or a machine as consumers of a product - a model. Accordingly, there are two types of work processes:

human-oriented visual analytic workflow - product mental model;

machine-oriented visual analytic workflow - product formal model.

Our work is a continuation towards the development and use of the capabilities of visual analytics. The ultimate goal of research is the development of information technology model construction. This technology should occupy a certain position in the concept of visual analytics. To this end, research was conducted in the direction of generalizing the working processes of visual analytics and systematizing the directions of their development, depending on the objectives of the analysis. The study had several stages and is displayed in the previous sections. They consisted of: conducting a generalized analysis of the processes of visual analytics and designating the directions of development of visual analytic workflow; defining goals for using workflow to implement specific model-building techniques (Section 3).

6. Information technology for model building

This section provides information technology in the context of machine-oriented visual analytic workflow.

In recent years, researchers have generalized and determined the direction of development of the interaction of the machine and human through visualization. A systematic approach is presented in papers Endert A. et al [52] integrating machine learning into visual analytics, Jiang L. et al [53] interactive machine learning, Cui W. [54] visual analytics, Andrienko N. et al [45] visual analytics as model building, Sacha D. et al [55] visual interaction with dimensionality reduction. Various aspects are considered as well as the use of visual tools. The visual tool must be used in the concept of implementing a specific workflow. Human machine interaction relies heavily on developing a visual tool. Visual tools is the link that connects the machine and the human through a dialogue of communication. Based on section 3, we define the ultimate goal as obtaining a formal model using visual analytics. We will demonstrate the use of human in constructing a formal model in the concept of “using opposite model” using the classification example.

In the machine supervised learning, we need to initially label the data. For labeling data, a human is used as an oracle, including using visualization of Bernard J. [56]. A human has a mental model that he uses to label data based on visual data grouping. Further, the data is used to obtain a formal model. A visual representation of grouped data using dimensionality reduction methods can be used for classification. Choo J. et al. [57], Yan Y. et al. [58] use data grouping visualization for classification as an interactive visual analytics system. In this case, the mental model is part of the Analytics System. A human takes an active part in the classification and is its integral part. Visual tools provide data in a convenient way for human use.

We propose a different approach. Analytics system forms a mental model at the training stage. Next, from the mental model we get the formal model, which we will use in the future. Obtaining a formal model from the mental one consists in defining the boundaries of the zones of class formation. A human visually defines the boundaries of classes indicating to the machine where and which classes and their boundaries. The boundaries of the class and determine the relationship of the data item to the class. The class boundaries that a human visually draws are projected into the n-dimensional hyperspace of attributes. The formal model actually consists of the rules for the position of the data element in the hyperspace of attributes relative to the boundaries of classes. Thus, a formal model is built for the machine based on the concept of “using opposite model”. The formal model here is the use of the mental model in a form convenient for the machine.

7. Separating n-dimensional objects by hyperplanes for classification based on data visualization

7. 1 Introduction and statement of the problem classification based on visualization

The classification task is one of the important parts of machine learning. In general, it consists in the fact that there is some finite number of objects from the studied problem, and it is known to which classes they belong, it is necessary to develop a system that will allow determining the belonging of new objects from this area of the task.

The main problem of visualization of multidimensional data relates to their presentation in two or three dimensions with minimal loss of information. Visualization is also useful for comparing various methods of reducing the dimension, which is generally quite easy to analyze. Visual presentation of data is the most informative for human perception. For data analysis and decision making based on the maximum information content. For this reason, the development of data visualization techniques is an important area. Many methods of reducing the dimension allow to reveal the hidden data structure and allow to find latent features. This can manifest itself in the spatial grouping of data, the formation of structures and clusters, as well as the degree of separability. Data separation is an important feature in classification tasks. At the same time, noise, redundancy and ambiguity of data can be reduced Cox T.F. and Cox M.A.A. [64]. It should be noted that, in general, the ultimate goal of visualization is to reduce the dimension of the feature space to a low-dimensional space that can be visually displayed.

Suppose we have a data set and we need to determine the measure of similarity between the data. This measure shows how similar or different two objects are. This can be obtained in various ways, such as calculating the correlation coefficient or the hermetic distance from the vector representation of the data. In MDS, each object in a low-dimensional space is represented by a point, and the distance between the points displays the original information about the similarity. That is, the greater the differences between objects, the farther they should be in low-dimensional space. The geometric location of the points allows to visualize the hidden data structure. This makes it easier to understand the data structure. Visually definable data clusters, agglomeration and separability of data. It is also possible to visually determine the boundaries of geometric formations based on the tasks of researching data and visualized hidden data structures. Based on this, MDS was chosen for data visualization, as a method that is based on "geometric distance", as a measure of the difference of objects. The visualization of the boundaries of agglomerations also uses geometric constructions based on distances. Thus, we can visualize the boundaries of the data sets. Information visual data analysis we can display on the data space. Thus, complementing the information content in the direction of the task of analysis.

A variety of solutions and approaches using the characteristic space of the studied data and their results are correlated. This is explained by the fact that they use a single attribute space with the extraction of its properties and features. The result of their result of their work is estimated by a number of metrics. A feature of the research presented in this paper is the development of a new technology, which is based on maximizing the use of the information component of the data set, by visual representation of the relationship between the common features. Visualization allows to show the hidden data structure, expanding their information content. This is an important element. However, it is necessary to expand the information content of the data by adding data management elements. One of the data management tools is the formation of data group boundaries on the visualized space. Thus, we not only see grouped data, but also limit the objects to our borders, which, in our opinion, based on the visual presentation, will belong to this group. This group can be a cluster, class.

Using this approach allows:

1) visualize the objects in question in the reduced space of generalized features in order to assess their distribution;

2) analyze the resulting groupings of objects in order to determine the nature of the sets and distributions;

3) most effectively assess the location and outlines of the lines that need to be set to divide the set of objects into the necessary classes;

4) using the given class restriction lines to build hyperplanes-analogues in the input-dimensional space that define the boundaries of the classes;

5) use the designated volume of space for further classification.

In general, denote $x=(x_1,x_2,...,x_n)^T$ where T - the sign of transposition, a feature vector, which characterize the object.

Since some objects belong to the same class, they are similar to a combination of characteristic features. We study the classification task for which the training sequence is determined $\Omega_x=\{x:x(1),...,x(m)\}$ where $x(j)=(x_{j1},\ldots,x_{jn})^T$, $j=\overline{1,m}$ a feature space whose elements correspond to different states of objects in the test domain. The classification problem, consider this: its need to find a hyperplane $w^T x+b=0, x\in R^n$ that when $x(j)\in\Omega_x(1)$ the inequality $w^T x+b>0$ and when $x(j)\in\Omega_x(2)$ respectively $w^T x+b<0$. Here $\Omega_x(1)\subset\Omega_x$ - a subset of feature vectors that correspond to the objects of the first class, $\Omega_x(2)\subset\Omega_x$ - a subset of signs for the second class, $\Omega_x=\Omega_x(1)\cup\Omega_x(2)$. $w=(w_1,...,w_n)^T$ - the coefficient vector, b - a some number.

The most simple in terms of application technology research is the problem of binary classification. Objects are divided into those that belong to the class and those who do not belong to the class. At the same time it is the basis for more complex tasks. We will consider the problem of two-class classification as the basic task of using information technology.

Classification tasks use labeled data. Each object is associated with an element of a finite set, a class label.

One of the main tasks in the first stage of data analysis is determining the classified data. It is necessary to evaluate the data from the point of view of separation. In this case, the visualization of data, namely the reduction of the dimensionality of signs to the space that can be visualized. The main task is to identify hidden patterns in sets of features based on a priori assumptions and the nature of these patterns. The classification task is considered in the field of classified texts. Here the number of signs can significantly exceed the number of objects of classification. This generally impairs data separability. However, it is possible to give an assessment using visualization. Since we use the training sample, the data must be separated by the classes we need. To evaluate the data for the ability to be divided into classes is a difficult task from the point of view of formalization. It is difficult to find approaches that provide such an opportunity. The most informative is the visualization of a set of interrelations of separating features from the point of view of analysis.

Therefore, it is necessary to visualize the training set to determine the location of the data in the feature space behind the division into two classes.

Visualization is the presentation of information in graphical form capable of analyzing a human. This includes the presentation of information dependencies in one, two and three-dimensional space, presented in graphical form.

Particular attention is paid to quality preservation of dependencies and information laws, which are characteristic of the original data. This grouping, distribution, form clusters, separability, etc.

The set of objects is determined feature space, which is represented in the form of a certain vector. The degree of similarity of objects is calculated on the basis of distances between the vectors of these objects. Vectors are identical, if the distance between them is zero, and the vectors are similar, if the distance between them is less than some threshold limit $\varepsilon \geq 0$ and different, if the distance is greater ε.

As a basic visual space we will use two-dimensional graphic space. This space is sufficiently informative and technically convenient in terms of the implementation of information technology.

The original is the n-dimensional feature space of objects. We need to reduce the space to two-dimensional based on the distance between the signs. For this we will use the method of multidimensional scaling, which allows for the necessary reduction.

The method of multidimensional scaling allows to place objects in a space of some small dimension (in this case, it is equal to two) in order to reproduce the observed distances between them with the smallest error. Thus, representing the visibility of the location of objects in the generalized two-dimensional space of generalized features Cox, T.F. and Cox, M.A.A. [64], Krak I.V. [65].

Further, data visualization is used as a way of displaying a multidimensional distribution of data on a two-dimensional plane, in which, the basic patterns inherent in the original distribution are qualitatively displayed. At the same time, it is necessary to minimize the loss of information content and its manifestations in the cluster structure, topological features, and dependencies between the characteristics of the location of the data in the original space. With a small amount of data, visual display allows to determine the existence of information links, which are weakly manifested when using methods in combination. In this case, informational links are difficult to determine with approaches that use a different nature of model formation Barmak O. et al. [66], Manziuk E.A. et al. [67].

The initial information is presented not in the form of a “object-feature” type table, but in the form of a square symmetric matrix D of mutual distances of objects from each other. At the intersection of *i*- row of j column in the matrix is the value of the distance from i to j object

Thus, first, each object is assigned coordinates in a multidimensional space. The task of multidimensional scaling is to construct a data set in the usual three-dimensional space or on a plane so that the distances between objects most closely correspond to the distances specified in the matrix Cox, T.F. and Cox, M.A.A. [64]. The input coordinate axes can be interpreted as some implicit factors, the values of which determine the differences between objects. If we provide each object with a pair of coordinates, then the result will be an image of the data visualization. Consider the classification method as an information technology, which is a sequence of steps.

7.2 Multidimensional scaling feature space

Initial data for the scaling is a matrix of pairwise distances between objects. Distance between i and j object is designated $\delta_{ij} = d(X_i, X_j)$. Objects are defined by multi-dimensional points $X_i = \{x_{i1}, x_{i2}, ..., x_{in}\}$. $i = 1...n$. The distance is calculated as follows:

$$d(X_i, X_j) = \left(\sum_{k=1}^{n} (x_{ik} - x_{jk})^2 \right)^{\frac{1}{2}} \quad (1)$$

The distance between the points in the space of lower dimension (reduced space) will be at a similar (1) and denoted by the formula $d(Y_i, Y_j)$.

Scaling aims to find points in space $Y_i = \{y_{i1}, y_{i2}, ..., y_{in}\}$. $i = \overline{1, n}$ so that the distance between the points in the reduced space was the closest to the distance in the multidimensional input space. Thus it is necessary to minimize the error display. Accordingly, display quality measure is determined σ - stress (stress):

$$\sigma = \sum_{i<j} w_{ij} \left(d(Y_i, Y_j) - \delta_{ij} \right)^2, \quad (2)$$

where w_{ij} - non-negative weight.

When normalizing the stress is defined as follows:

$$\sigma = \frac{\sum_{i<j} w_{ij}\left(d\left(Y_i, Y_j\right) - \delta_{ij}\right)^2}{\sum_{i<j} w_{ij}\delta_{ij}^{\ 2}}. \tag{3}$$

The normalization allows to seize the interpretation of visual quality, and reduce dependence on the number of objects and their locations.

For pairwise distance matrix formed D in a multidimensional space will carry out the following preliminary steps:

1) double centering matrix by one of the known methods;

2) based on the output dimension n define eigenvectors $e_1, e_2, \ldots, e_n$ of the obtained matrix;

3) calculate the matrix $X = E_n \Lambda_n^{0.5}$. E_n - matrix of eigenvectors $e_1, e_2, \ldots, e_n$. Λ_n - a diagonal matrix of eigenvalues.

Then, the coordinate matrix that is used to obtain a multidimensional scaling by eigenvalue decomposition of the matrix $B = XX^T$.

Note that the error function in different types of projection data are quite extensive and are based on interpretations of the multidimensional scaling and optimization algorithms. Multidimensional scaling in the information technology is used as the most comprehensive approach In various specific cases, various modifications may be used. In case of non-metric methods of multidimensional scaling, not quantitative measures of object similarity are used, but only their relative order. The minimization of the stress function σ corresponds to finding the most optimal agreement between the matrix of the initial distances and the matrix of the resulting distances.

7.3 Minimizing stress function

To minimize the stress function, the approach is to find the proximity matrix and use iterative algorithm SMACOF (Scaling by MAjorizing a COmplicated Function - scaling for majoration complex functions) to a predetermined stress value. SMACOF algorithm is based on the strategy, the use of which provides a good convergence model. The goal, in accordance with the principle of majorization, is to find a simpler and more controlled function $g(x, y)$ that majorizes the objective function $f(x)$.

At the same time for all x. $g(x, y) \geq f(x)$, here y - a fixed reference point values. The reference point is the point of tangency surface $g(y, y) = f(y)$ while minimizing point x_* satisfies inequality $f(x_*) \leq g(x_*, y) \leq g(y, y) = f(y)$ forming thereby a layered structure.

Generally majorization is an iterative procedure consisting of several steps:

- determining a reference point $y = y_0$;
- calculation x_* based on the condition $g(x_*, y) \leq g(y, y)$;
- go to the previous step of the installation $y = x_*$ if the condition has not been reached $f(y) - f(x_*) \leq \varepsilon$.

This approach successfully generalized multidimensional spaces subject to the inequality, and is used to minimize the objective function.

On the function of domination $g(x, y)$ imposed a number of conditions, which cause the advantage of its use. Required:

- minimize easier than $f(x)$;
- be at least in the initial field than the original function $f(x) \leq g(x, y)$;
- be tangent to some function $f(x)$ a foothold $f(y) = g(y, y)$.

Set $Y = \{Y_1, Y_2, \ldots, Y_m\}$ m iteratively calculate points using the transformation Guttman L.

$$Y_{k+1} = V^{+} B(Y_k) Y_k, \tag{4}$$

where k - the number of iteration;

V^{+} - pseudo inverse matrix for the matrix of weights V with elements

$$v_{ij} = -w_{ij}, i \neq j;$$
$$v_{ii} = \sum_{i=1, j \neq i}^{m} w_{ij}. \tag{5}$$

matrix $B(Y_k)$ elements comprises:

$$b_{ij} = \begin{cases} -\dfrac{w_{ij}\delta_{ij}}{d(Y_i, Y_j)}, i \neq j \,\&\, d(Y_i, Y_j) \neq 0; \\ 0, i \neq j \,\&\, d(Y_i, Y_j) = 0. \end{cases} \tag{6}$$

$$b_{ii} = -\sum_{i=1, j \neq i}^{m} b_{ij}. \tag{7}$$

With the proviso that in formula (4) scales $w_{ij} = 1$ we obtain

$$Y_{k+1} = \frac{1}{m} B(Y_k) Y_k. \tag{8}$$

Hence, for the construction of domination procedure, it's must perform the following actions:

- set the initial value of the reduced space Y_0;
- write the stress function $\sigma = \sum_{i<j} w_{ij} \left(d(Y_i, Y_j) - \delta_{ij}\right)^2$;
- find the value $Y_{k+1} = V^{+} B(Y_k) Y_k$;
- calculate the stress function $\sigma(Y_{k+1})$;
- specify iteration increment $k++$;
- verify the convergence conditions $\sigma(Y_{k-1}) - \sigma(Y_k) < \varepsilon$, Otherwise move to the stress function.

Thus, at this stage, information technology process consists of these steps:

1) formation of the matrix of pairwise distances on the basis of the input data;
2) finding the square of the distance of the distance matrix;
3) use of double centering matrix;
4) determining eigenvalues and eigenvectors of the matrix;
5) optimization algorithm maps SMACOF.

The result is a set of objects with a pair of coordinates, which can be displayed. To display the two-dimensional space in two quite generalized coordinates. Mapping objects are marked with a multidimensional data space. Since the objects are marked, it is necessary to designate the classes to the resulting plane to form a decision tree based on a linear classifier. In general, we obtain the visualization of data marked for use classification. In this case, we can define the class boundaries. Thus, the training for the information of this technology is to define the boundaries and field classes. class boundary assessed visually and is determined by taking into account the fields of the class c. class border may be spaced from the extreme object class. It's necessary, because it allows improving the generalized classification. This aspect is particularly important because it allows finding the most optimal configuration accuracy and generalized classification.

7.4 Linear discriminant function

If the linear discriminant function, the classifier $d(\bar{x})$ determined by the relation

$$d(\bar{x}) = \overline{W}^T \bar{x} + w_n, \tag{9}$$

where $\bar{x} = (x_0, x_1, ..., x_{n-1})^T$ - a feature vector that defines the image of the object to be classified; $W = (w_0, w_1, ..., w_{n-1})^T$ - weight vector classifier; w_n - the threshold value.

Belonging to one of two classes - $\Omega(1), \Omega(2)$ - defined by the rule

$$d(\bar{x}) = \sum_{i=0}^{n-1} w_i x_i + w_N \underset{>}{<} 0 \to \begin{cases} \Omega(1) \\ \Omega(2) \end{cases}. \tag{10}$$

Hence, for the formation of a linear classifier is necessary to find the coefficient vector $\overline{W}$ and threshold w_n.

Note that for practical applications, to obtain separation into two classes using only linear classifier (10) hard enough, thus it is not possible to divide the data curve or broken line. One way of solving this problem is to construct a classifier using a combination of linear classifiers, thus forming a piecewise linear set with the required degree of discretization. This approach has the advantage of reduced space visualization and enables display control data classification. Piecewise linear approach is most appropriate, since it allows to use a combination of linear separators. This allows taking advantage of linear separators and creating the necessary configuration in the visual space. When using a linear classifier in a multidimensional space is sought hyperplane, which is the criterion of separating their respective classes. Next searched vector y_i, For a new element represented by point x_i and a limit value b from the condition:

$$y_i = \begin{cases} +1, wx_i > b; \\ -1, wx_i < b. \end{cases} \tag{11}$$

Equation (11) describes at zero hyperplane. It is known that the vector w perpendicular to the desired separation line with the corresponding properties: The best possible separation line Distant from nearest thereto classes separation points. Note that the distance between these points defines the separation strip which corresponds to the condition $-1 < wx_i - b < 1$ and a band edge point no elements, the width of the separation strip is $2/|w|$. Note that when the division of the classes via the separating strips are important only boundary point, since the strip consists of parallel lines extending along the boundaries of classes. These lines do not represent a division of classes (this function assumes band division), and mark the boundaries of classes, thus limiting their lines. Hence, the problem is transformed not into finding the delimitation of, and in finding the class boundaries, which are the limits line. Thus hyperspaces divided into some limited hypervolume within classes are represented. In this classification in this case is a controlled process. If necessary, we can change the classes of the border, and this changing the way the accuracy and generalized classification.

Note that when there are multiple classes in the construction of a linear classifier, lines crossing occurs and the construction of the segments forming a piecewise-linear structure which is generally nonlinear. Using the class limit line, we get some geometric structures limits the class. An element that was in this limitation belongs to this class. Another important aspect is the formation of spatial classes' volumes. These volumes can be located at some distance. Objects that do not fall in the scope of classes do not belong to any class. Thus, there is a set of objects that do not belong to the same class. These objects may belong in a comparatively equally to different classes.

To improve the linear classification problem, an increase of the space dimension is used. The space is expanded by the mapping function to the new space. To expand the two-dimensional space in three-dimensional, the display function is represented as follows:

$$\phi(x) = \phi(x_1, x_2) = \left(x_1^2, x_2^2, \sqrt{2}x_1x_2\right). \quad (12)$$

Increasing the dimension of the space allows, due to the bending of space to find a hyperplane linear classification. Thus, the hyperplane allows linearly allocate cloven classes, which is possible when the convexity of the objective function. Further, the reverse is lowered space class separation line can suitably describe the piecewise linear way limit lines. This allows the use of multiple increasing dimensionality space under back projection to determine the hyperplanes. As a consequence, approaches using reducing space by scaling, it is possible to determine the grade boundaries imaging techniques with subsequent projection into the multidimensional space. In this case, the display function in n dimensional space will look like

$$\phi(x) = \phi(x_1, x_2, ..., x_n).$$

It should be noted that the classification boundaries may become distorted during expansion space. Education class boundaries is a flexible tool that also allows to use and field classes. Thereby expanding, class limits and compensating for errors of their determination. Flexibility visually determining class boundaries based analysis allows the data structure to determine the required amounts of classes. This makes it necessary for the accuracy of the analysis to determine the classification of objects.

7.5 Formation of a decision tree classifier

We present a method for constructing a piecewise linear classifier using decision tree algorithm based on data visualization system. Decision tree - a method for mapping rules in a hierarchical, sequential structure, wherein each object corresponds to a single node, giving solution Kruskal J. B. and Wish M. [75]. Under the rule refers to a logical structure, presented in the form of "if ... then ...". rules are defined by curves which divide a group of objects on the imaging plane of the system. Curves are specified as a piecewise-linear structure with varying degrees of necessary discretization, thereby forming, in the first approximation curves. As a result, when to add a new object, can explicitly specify the class to which it belongs. At the initial stage of training runs, which identifies areas and to what class they belong. All objects of a particular area in relation to the borders have a spatial position. This position forms the rules of the object belonging to the class. After learning when classifying a new object, a set of rules is defined. An object belongs to that class, with the rule sets of objects it matches. Rule sets define a set of attributes of an object-to-class relationship. These feature sets are used to build a decision tree.

We define a few situations from a variety of possible situations T the construction of a decision tree.

1. The set T contains elements that belong to the same class. In this case, the decision tree defines the class. If the set T contains no elements, the decision tree determines the branch and the class associated with this branch is retrieved from another set other than T, for example, an ancestor node.

2. The set T contains elements that belong to different classes; the set is divided into subsets. For this, a feature is defined that contains more than two distinct values $O_1, O_2, ..., O_n$. The set T is divided into subsets, with each subset T_i containing elements that are relevant O_i to the selected trait. The process is recursive, the final condition of which is the formation of subsets, which consist of elements of one class.

When building a tree on each internal node, it is necessary to find the condition for dividing the set on this node into subsets. As a condition, one of the attributes is accepted. The general rule is this: the attribute divides the set in such a way that the resulting subsets consist of objects that belong to the same class or are maximized by this attribute. To find the attributes, the algorithm C4.5 Quinlan J. R. [76] is used, where the attribute $Gain(\Theta)$ of the set Θ is selected by the following criterion:

$$Gain(\Theta) = Info(T) - Info_x(T) Gain(\Theta) = Info(T) - Info_x(T), \tag{14}$$

where $Info(T)$ - entropy sets T;

$$Info_x(T) = \sum_{i=1}^{n} \frac{|T_i|}{|T|} Info(T_i). \tag{15}$$

Subsets $T_1, T_2, ..., T_n$ are obtained from the original set T when checking the set Θ. An attribute is selected that gives the maximum value behind the criterion (15). At the same time, in order to reduce the number of subsets, it is necessary to minimize the number of nodes and branches. In the general case, it is recommended to minimize the number of linear elements with piecewise linear delineation. This reduces the number of class separation rules. This reduces the number of calculations, without compromising the quality of the classification.

Nonlinear (piecewise linear) classifier initially operates in a multidimensional space. To form the separating constraints in this space restrictions must be converted (line) is piecewise linear classifier reduced space hyperplanes in a multidimensional space limitations. To do this, it expands the dimension of the reduced space to the original. Thus, the boundaries projected in the original space. Borders form class membership rules.

After the expansion space and the formation of hyperplanes defined by their equations. To construct hyperplanes in n dimensional space must be correspondingly n points which have been obtained by adding extra $n-2$ a point on a line segment. Thus, we get a system of linear equations,

$$\begin{cases} wX_1 + b_1 = 0 \\ \dots \\ wX_n + b_n = 0 \end{cases} \tag{16}$$

which is generally solved by Gauss. Here w - unknown coefficients hyperplanes.

Class in the multidimensional space is defined by limiting hyperplanes. These boundaries and form the right attitude to classes of objects. To classify new data is determined by their position in the multidimensional space by determining their position relative to hyperplanes. Substituting coordinate data in the hyperplane equation determine their relative location of a plurality of $\{-1,0,1\}$. If the result is less than zero element is conditionally "right" relative to a plane when the result is greater than zero - the element is "left" plane and, respectively, if equal to zero, the element is located on the dividing plane.

Formation of non-linear classification rules produced a sequence of actions:

1) forming a piecewise linear visual class restrictions in the reduced space;
2) calculation of the reference points-rules for the class;;
3) transformation point rules in multidimensional space;
4) construction of hyperplanes in a multidimensional space based on the transformed points;
5) formation rules for the class in a multidimensional space on the basis of restrictive hyperplanes.

Piecewise linear restrictive rules define the scope of the class and allow to visually determine whether to increase or limit the class square, which is important in the border data. This allows for a good interpretability of results of classification and control of restrictive class of the field and, in fact, interactive classification system. It should be noted that the classification process takes place under the rules of the decision tree. This process is quite fast and does not require large resources.

Ensuring that the visual component of the classification is particularly important in comparison with other approaches, especially in the labeling complex boundary conditions. This ensures the presence of an additional information component via the interactive visual means of determining the class limits. This allows the tool system obtains more information and supervised classification process. The results of the system are well understood by and controlled as a result of visual presentation and interactivity restrictive rules. Area restrictions provide minimum visual boundaries that, if necessary, can be overridden. The restrictions are transformed into line multidimensional space and are presented there by hyperplanes, forming such a way restrictive area. The classification of new data occurs in a multidimensional space based on the calculated data and the relevant provisions on restricting hyperplanes. Determining the spatial position of the new element with respect to all hyperplanes, thereby determining its location in limited volumes class categories. This process is controlled because of scaled to have a reduced space and a visual representation of the classified new data elements. As a consequence, the result of classification in multidimensional space is presented and assessed visually, allowing the interpretation and analysis of compliance of new data elements with respect to the categories of classes. The multiclass conditions, if the element has the necessary information content,

For a practical demonstration of information technology software system is developed for classification used text data. The data is based on the Reuters corpus and is selected to demonstrate a method for well-separable data. Text data have a large set of attributes. This is important in terms of determining the possibility of representing and minimizing distortion while reducing the dimensionality of space. What is important is the amount of distortion, which is particularly evident with a significant reduction of space. Separating features formed hyperspace large dimension, which should be reduced to two-dimensional. This is important because a significant reduction of space is accompanied by a measure of the distance differences in the original and the new space. Based on this data were selected, which are characterized by large dimensions of signs in order to investigate the influence of the distortion measure of proximity and verification technology performance. Sample text data formed the basis and the separation capacity and the formation of class categories. The distances between classes well definable in the projected space.

The data represent the classes well and are grouped by generalized attributes. The two-dimensional representation shows that the data well represents the classes to which they relate. At the same time, the classes are fairly well spaced.

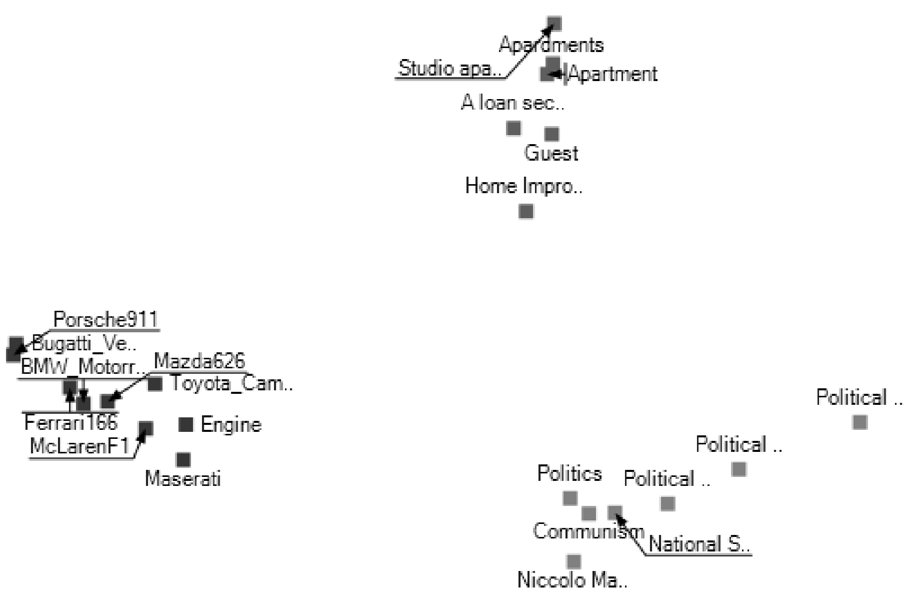


Figure 1

Three groups of objects are well grouped and the groups are located far enough in the two-dimensional space. Since the objects are from the training set, we determine that these groups represent classes.

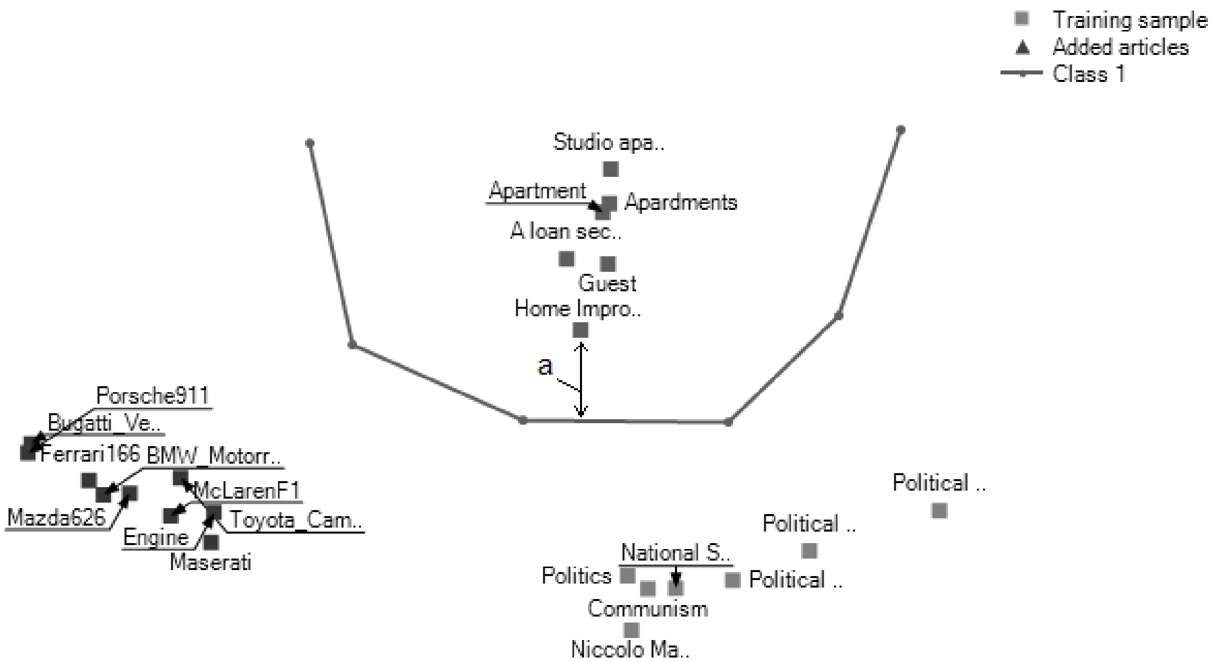
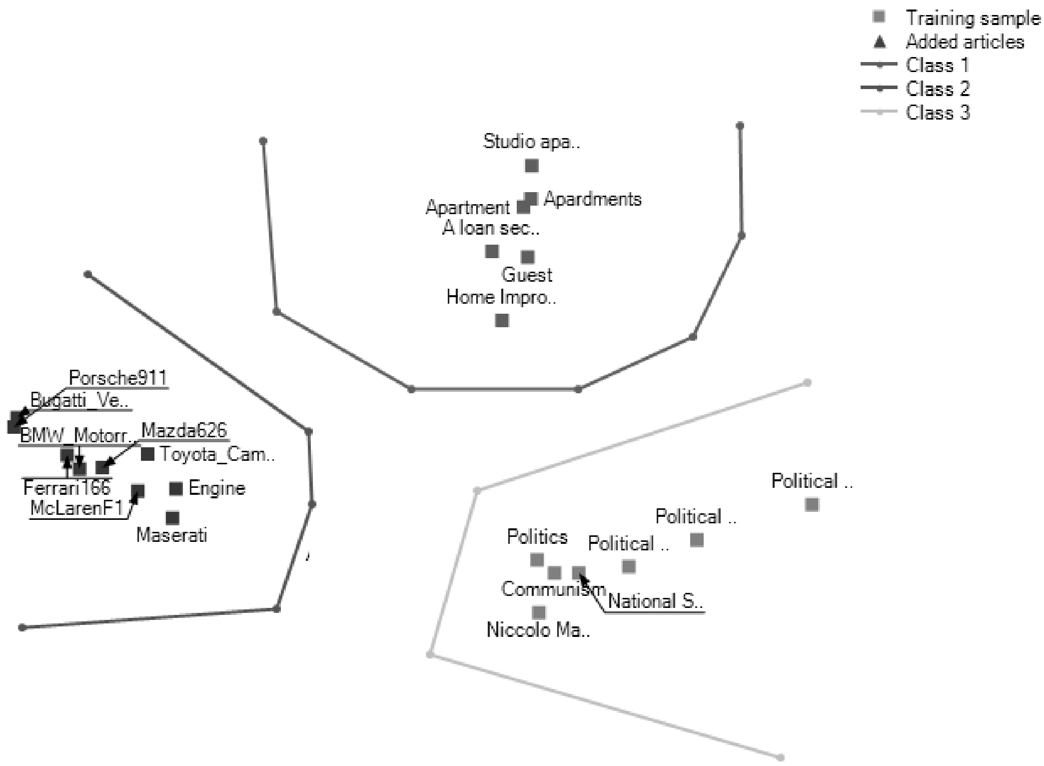


Figure 2

Next, we define a class graphically grouping boundaries. It is important not only to the Class B limits, as well as the definition of fields of class boundaries. Visually, we can define a parameter a (Fig. 2) as a class field. We can graphically become the minimum distance from the boundary to the class object. It allows to visually providing the opportunity generalized model for each class. Since the margin area can be defined quite flexibly, depending on the separability of classes, which generally can be uneven across the data area. For some classes of the border can take place quite clearly without the possibility of flexible formation.

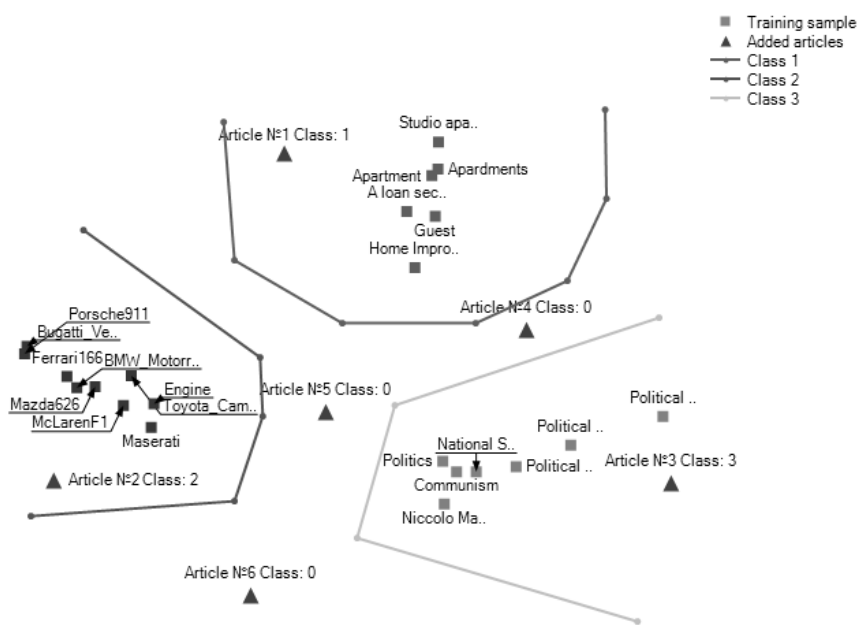


line_1	line_2	line_3	line_4	line_5	line_6
-1	-1	-1	-1	-1	-1
1	-1	-1	-1		
1	1	-1			

Figure 3

Generalized opportunity model is determined based on the separability of classes. Visual determination of the boundaries of classes allows graphically piecewise-linear method to establish the boundaries of classes and generalization performance models.

Consistently limiting class defined by linear segments for the formation of decision rules. In this instance, the decision rules are indicated in the table and determine the position of each object class with respect to borders. Since the border is a collection of connected segments, the position of the object is determined relative to each line segment. Accordingly unnecessarily advised to avoid a detailed definition of class boundaries. This will greatly speed up the learning process and the classification itself with the same accuracy. Each object of the class is characterized by the general class rules with respect to its borders. Number of rules corresponds to the number of segments of the boundary line. Fig. 3 indicates a table of rules of the three rows in a cell which are the rules of the relative position of the object class. Coordinates of the point is generalized symptoms in two-dimensional space. Rules designated position relative to grade boundaries to determine such crucial designation as the "inside" and "outside" the class. The boundaries of this class circuit of piecewise linear segments with no gaps. The position of an object is determined with respect to each segment of the loop to form a set of rules.



line_1	line_2	line_3	line_4	line_5	line_6
-1	-1	-1	-1	-1	-1
1	-1	-1	-1		
1	1	-1			

Figure 4

In the general case, it is necessary to position the indicated class of objects relative to the boundary bounding the class. The trained formal model - is the volume of hyperspace, in which there are objects of classes with the necessary tolerances of generality. When displayed on a two-dimensional space define the position of the new facilities. Data for testing are located relatively classes Article№1 Class 1, Article№2 Class 2, Article№3 Class 3. The feature of the proposed technology is the existence of band separation between the classes. We graphically define generalization performance models. We do not draw the line graph classes and limit them. Thus if an object is located between the classes does not belong to the classes. If it is necessary not to limit the classes, classes are differentiated only. In this case, there will be no objects that are between classes. The system is trained by graphic determination of class boundaries.

8. Conclusion and discussion

In this study, we determined the goals of using visual analytics workflow that focus on obtaining the final product - a model: a human-oriented visual analytic workflow builds a mental model, machine-oriented visual analytic workflow - builds a formal model. The model is considered as an information processor and a decision-making mechanism. The formal and mental models differ only in consumers, which are a machine or a human. Two concepts of model building based on model synchronization and 'using opposite model' are proposed. Using the concept of 'using opposite model', an information technology has been developed that allows the machine to obtain a model based on a mental model and demonstrates the possibility of using this concept. This allows the machine to fully use the intellectual capabilities of a human based on a model.

The use of human intellectual abilities to build machine learning models is an important area based on a number of advantages. The resulting model is fundamentally different for the construction method from other machine learning approaches naive Bayes, k-nearest neighbors, support vector machine, random forest or deep learning and so on. The main distinguishing feature is the fact that the model is formed by a human and is subsequently used by the machine. However, the difference is not only in the method of obtaining the model. The base for the formation of this model differs from the models built by the machine, which is important for use in ensembles of models. The success of using ensembles lies in the diversity of opinions made by composition algorithms. Typically, models are based on machine learning algorithms. In the proposed approach, the model is built on the basis of formatted knowledge and human experience. This allows us to propose a model with a different nature for ensembles.

However, it should be pointed out that the model formed by a human may be weak. Classical approaches can give the best result according to the quality criteria of the model. The advantage of a human-derived model is that a human can obtain new information from data. So, for example, for such an approach of ensembles as stacking, the main thing is how different a model is to each other. How much each model of new information can bring to the ensemble. In the general case, a human may make a mistake in constructing a model, but a human may also notice weak joints between the data and form a model with a different opinion. This is the main value of the proposed model building approach.

Another aspect that speaks of the need to use a human in the construction of a model is such a property of models.

Very often, the results of model decisions cannot be interpreted, for example, using neural networks. Since it is not clear what decisions are being made, the use of such approaches is limited for cybersecurity and solution against human considerations. The proposed approach allows the machine to use a model built by a human. The analytical process allows creating a decision-making model for a human and the interpretability of the result for a human is of a high level. Thus, the use of these models in the field of cybersecurity is desirable and in some cases the only possible one.

Information technology is proposed which allows maximizing the information content of the feature space of objects together. For this purpose To do this, use feature space reduction by 2D. Further reduction of dimension can reveal hidden data structure and allows to find latent features.

The main factor of information technology is to minimize the loss of information data and a visual graphical management training model for data classification. The information technology constructed using the proposed method provides a flexible data classification tool with the ability to specify the class boundaries.

Information technology has limitations. Visually presented data should be visually separable and grouped. If a human does not build a mental model, data cannot be classified. Accordingly, it is necessary to improve methods and procedures for the visual presentation of data for humans. Scatter plot is not suitable for classifying a large number of classes and data.

Information technology was developed to demonstrate the concept of ’using opposite model’. The formal model is entirely based on the mental model. The capabilities of the machine were not used for data classification. More promising is the use of the concept of model synchronization. In this case, the advantages will be used for the construction of models, both human and machine.

REFERENCES

1. Endert, A., Hossain, M. S., Ramakrishnan, N., North, C., Fiaux, P., Andrews C. The human is the loop: new directions for visual analytics. Journal of Intelligent Information Systems, 43(3), 2014, pp. 411–435.
2. Chen, M. and Golan, A. What may visualization processes optimize? IEEE Trans. on Visualization and Computer Graphics, 22(12), 2016, pp. 2619–2632,.
3. Chen, M. and Jänicke H. An information-theoretic framework for visualization. IEEE Transactions on Visualization and Computer Graphics,16(6), 2010, pp. 1206–1215,.
4. Amershi, S., Cakmak, M., Knox, W. B., Kulesza, T. Power to the people: The role of humans in interactive machine learning. AI Magazine, 35(4), 2014, pp. 105–120.

5. Jannach, D., Lerche, L., Zanker, M.: Recommending based on implicit feed-back. In: Brusilovsky, P., He, D. (eds.) Social Information Access, LNCS, 10100(14), 2018, pp 510-569.
6. Pan, W. A survey of transfer learning for collaborative recommendation with auxiliary data, Neurocomputing 177, 2016, pp. 447-453.
7. Zhao, J., Cao, N., Wen, Z., Song, Y., Lin, Y.-R. and Collins, C. #fluxflow: Visual analysis of anomalous information spreading on social media. IEEE Transactions on Visualization and Computer Graphics, 20(12), 2014, pp. 1773-1782.
8. Cao, N., Shi, C., Lin, S., Lu, J., Lin, Y.-R. and Lin, C.-Y. Targetvue: Visual analysis of anomalous user behaviors in online communication systems. IEEE transactions on visualization and computer graphics, 22(1), 2016, pp. 280-289.
9. McKenna, S., Staheli, D., Fulcher, C. and Meyer, M. Bubblenet: A cyber security dashboard for visualizing patterns. In Computer Graphics Forum, Wiley Online Library, 35, 2016, pp. 281-290.
10. Wu, W., Zheng, Y., Chen, K., Wang, X. and Cao, N. A visual analytics approach for equipment condition monitoring in smart factories of process industry, inPacific Visualization Symposium (PacificVis), 2018 IEEE.IEEE, 2018, pp. 140-149.
11. Liu, S., Xiao, J., Liu, J., Wang, X., Wu, J. and Zhu. J. Visual diagnosis of tree boosting methods. IEEE Trans. on Visualization and Computer Graphics, 24(1), 2018, pp. 163-173.
12. Krause, J., Dasgupta, A., Swartz, J., Aphinyanaphongs, Y. and Bertini, E. A workflow for visual diagnostics of binary classifiers using instance-level explanations. IEEE Conference on Visual Analytics Science and Technology (VAST), 2017, pp. 162-172.
13. Ren, D., Amershi, S., Lee, B., Suh, J. and Williams, J. D. Squares: Supporting interactive performance analysis for multiclass classifiers. IEEE Transactions on Visualization and Computer Graphics, 23(1), 2017, pp. 61–70.
14. Alsallakh, B., Hanbury, A., Hauser, H., Miksch, S. and Rauber, A. Visual methods for analyzing probabilistic classification data. IEEE Trans-actions on Visualization and Computer Graphics, 20(12), 2014, pp. 1703–1712.
15. Cashman, D., Humayoun, S. R., Heimerl, F., Park, K., Das, S., Thompson, J., Saket, B., Mosca, A., Stasko, J., Endert, A. et al. A user-based visual analytics workflow for exploratory model analysis. In Computer Graphics Forum, 38, 2019, pp. 185–199.
16. Kwon, B. C., Eysenbach, B., Verma, J., Ng, K., deFilippi, C., Stewart, W. F. and Perer, A. Clustervision: Visual supervision of unsupervised clustering. IEEE Trans. on Visualization and Computer Graphics, 24(1), 2018, pp. 142–151.
17. Mühlbacher, T., Linhardt, L., Möller, T. and Piringer, H. TreePOD: Sensitivity-aware selection of pareto-optimal decision trees. IEEE Trans. on Visualization and Computer Graphics, 24(1), 2018, pp.174–183,
18. Strobelt, H., Gehrmann, S., Pfister, H. and Rush, A. M. LSTMVis: A tool for visual analysis of hidden state dynamics in recurrent neural networks. IEEE Trans. on Visualization and Computer Graphics, 24(1), 2018, pp. 667–676.
19. Liu, D., Cui, W., Jin, K., Guo, Y. and Qu, H. DeepTracker: Visualizing the Training Process of Convolutional Neural Networks. ACM Transactions on Intelligent Systems and Technology (TIST), 10(1), Article 6, 2018.
20. Kwon, B.C., Choi, M.J., Kim, J.T., Choi, E, Kim, Y.B., Kwon, S., Sun, J., Choo, J RetainVis: Visual Analytics with Interpretable and Interactive Recurrent Neural Networks on Electronic Medical Records, Visualization and Computer Graphics IEEE Transactions on, 25(1), 2019, pp. 299-309.
21. Sacha, D., Kraus, M., Keim, D. A., Chen, M. Vis4ml: An ontology for visual analytics assisted machine learning. IEEE Transactions on Visualization and Computer Graphics 25(1), 2019, pp. 385–395.

22. Card, S.K., Mackinlay, J D. and Shneiderman, B. Using vision to think. Readings in information visualization: using vision to think, 1999, pp. 579-581.
23. Fayyad, U., Piatetsky-Shapiro, G., Smyth, P. Knowledge discovery and data mining: Towards a unifying framework. In Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD’96, AAAI Press, 1996 ,pp. 82-88.
24. Yi, J. S. Kang, Y.-A., Stasko, J. T. and Jacko, J. A. Toward a deeper understanding of the role of interaction in information visualization. IEEETrans. on Visualization and Computer Graphics, 13(6), 2007, pp. 1224-1231.
25. Rind, A., Aigner, W., Wagner, M., Miksch, S., Lammarsch, T. Task cube: A three-dimensional conceptual space of user tasks in visualization design and evaluation. Information Visualization, 15(4), 2016, pp.288-300.
26. Thomas, J. and Cook, K. A. editors. Illuminating the Path: The R&D Agenda for Visual Analytics. IEEE Press, 2005.
27. Sacha, D., Stoffel, A., Stoffel, F., Kwon Bum Chul, K., Ellis, G., Keim, D.A. Knowledge generation model for visual analytics. IEEE Transactions on Visualization and Computer Graphics 20(12), 2014, pp. 1604–1613.
28. de Oliveira, M. C. F., Levkowitz, H. From visual data exploration to visual data mining: a survey. IEEE Transactions on Visualization and Computer Graphics 9(3), 2003, pp. 378–394.
29. Federico, P., Wagner, M., Rind, A., Amoramorós, A., Miksch, S., Aigner, W. The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics. In Proc. IEEE Conference on Visual Analytics Science and Technology (VAST), 2017, pp. 92-103.
30. Chi, E.H. A taxonomy of visualization techniques using the data state reference model. In IEEE Symposium on Information Visualization (InfoVis), 2000, pp. 69–75.
31. Sedlmair, M., Heinzl, C., Bruckner, S., Piringer, H., Möller, T. Visual parameter space analysis: A conceptual framework. IEEE Transactions on Visualization and Computer Graphics 20(12), 2014, pp. 2161–2170.
32. Micallef, L., Schulz, HJ., Angelini, M., Aupetit, M., Chang, R., Kohlhammer, J. The Human User in Progressive Visual Analytics EuroVis'19 Short Papers, 2019, pp.19-23.
33. Windhager, F, Federico, P, Schreder, G, Glinka, K, Dörk, M, Miksch, S, Mayr, E. Visualization of cultural heritage collection data: State of the art and future challenges IEEE transactions on visualization and computer graphics 25(6), 2019 , pp. 2311-2330.
34. Kehrer, J., Hauser, H. Visualization and visual analysis of multifaceted scientific data. A survey.IEEE Transactionson Visualization and Computer Graphics, 19(3), 2013, pp. 495–513.
35. van Wijk, J.J. The value of visualization. In: IEEE Visualization, 2005 p. 11.
36. Federico, P., Wagner, M., Rind, A., Amor-Amorós, A,. Miksch, S,. Aigner, W. The role of explicit knowledge: A conceptual model of knowledge-assisted visual analytics IEEE Conference on Visual Analytics Science and Technology (VAST), 2017, pp. 92-103.
37. Yalçin, M. A., Elmqvist, N., Bederson, B. B. Cognitive stages in visual data exploration. In BELIV, ACM, 2016, pp. 86-95.
38. Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., Melançon, G. Visual analytics: Definition, process, and challenges. In Information Visualization: Human- Centered Issues and Perspectives, Kerren A., Stasko J. T., Fekete J.-D., North C., (Eds.). Springer, Berlin, 2008, pp. 154-175.
39. Keim, D., Kohlhammer, J., Ellis, G., Mansmann, F. (Eds.) Mastering The Information Age – Solving Problems with Visual Analytics. Eurographics Association, Goslar, 2010.
40. Hestenes, D. Modeling methodology for physics teachers. In E. Redish & J. Rigden (Eds.), The changing role of the physics department in modern universities. Part II American Institute of Physics., 1997, pp. 935–957.

41. Booth, P., Gibbins, N. and Galanis, S. Towards a Theory of Analytical Behaviour: A Model of Decision-Making in Visual Analytics. In: Proceedings of the 52nd Hawaii International Conference on System Sciences. Honolulu: University of Hawaii at Manoa, 2019, pp. 1607-1616.
42. Green, T., Ribarsky, W., Fisher, B. Building and Applying a Human Cognition Model for Visual Analytics. Information Visualization 8(1), 2009, pp. 1–13.
43. Pirolli, P. and Card, S. The Sensemaking Process and Leverage Points for Analyst Technology as IdentifiedThrough Cognitive Task Analysis. in Proceedings of international conference on intelligence analysis, 2005, pp. 2–4.
44. Lammarsch, T., Aigner, W., Bertone, A., Miksch, S. and Rind, A. Towardsa concept how the structure of time can support the visual analytics process. In S. Miksch and G. Santucci, editors,Proc. Int. Workshop VisualAnalytics (EuroVA) in conjunction with EuroVis, 2011, pp. 9-12.
45. Andrienko, N., Lammarsch, T., Andrienko, G., Fuchs, G., Keim, D., Miksch, S. and Rind, A. Viewing visual analytics as model building. Computer Graphics Forum, 37(6), 2018, pp. 275-299.
46. Lee, T., Johnson, J., & Cheng, S. (2016). An Interactive Machine Learning Framework. ArXiv, abs/1610.05463.
47. Holzinger, A.; Plass, M.; Kickmeier-Rust, M.; Holzinger, K.; Cri¸san, G.C.; Pintea, C.M.; Palade, V. Interactive machine learning: Experimental evidence for the human in the algorithmic loop. Appl. Intell. 49(7), 2019, pp. 2401-2414.
48. Teso, S., Kersting, K. Explanatory interactive machine learning. AIES '19 Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 2019, pp. 239-245.
49. Murdoch, W.J., Singh, C., Kumbier, K., Abbasi-Asl, R., Yu, B. (2019) Interpretable machine learning: definitions, methods, and applications arXiv:1901.04592
50. Zhao, X., Yan, X., Van Hentenryck, P. (2019) Modeling Heterogeneity in Mode-Switching Behavior Under a Mobility-on-Demand Transit System: An Interpretable Machine Learning Approach arXiv:1902.02904
51. Hall, K.W., Bradley, A.J., Hinrichs, U., Huron, S., Wood, J., Collins, C., Carpendale, Sh. (2019) Design by Immersion: A Transdisciplinary Approach to Problem-Driven Visualizations. CoRR abs/1908.00559
52. Endert, A., Ribarsky, W., Turkay, C., Wong, B.L.W., Nabney, I., Díaz Blanco, I. & Rossi, F. 'The state of the art in integrating machine learning into visual analytics, Computer Graphics Forum, 36(8), 2017, pp. 458-486.
53. Jiang, L., Liu, S. and Chen, C. Recent research advances on interactive machine learning. Journal of Visualization, 22(2), 2019, pp. 401-417.
54. Cui, W.,Visual Analytics: A Comprehensive Overview. in IEEE Access, 7, 2019, pp. 81555-81573.
55. Sacha, D., Zhang, L., Sedlmair, M., Lee, J.A., Peltonen, J., Weiskopf, D., North, S.C., & Keim, D.A. Visual Interaction with Dimensionality Reduction: A Structured Literature Analysis. IEEE Transactions on Visualization and Computer Graphics, 23, 2017, pp. 241-250.
56. Bernard, J., Hutter, M., Zeppelzauer, M., Fellner, D. and Sedlmair, M., Comparing Visual-Interactive Labeling with Active Learning: An Experimental Study. in IEEE Transactions on Visualization and Computer Graphics, 24(1), 2018, pp. 298-308.
57. Choo, J., Lee, H., Kihm, J., Park, H. ivisclassier: An interactive visual analytics system for classication based on supervised dimension reduction. In Visual Analytics Science and Technology (VAST), 2010 IEEE Symposium on, IEEE, 2010, pp. 27-34.
58. Yan, Y., Tao, Y., Jin, S., Xu, J. and Lin, H. An Interactive Visual Analytics System for Incremental Classification Based on Semi-supervised Topic Modeling, 2019 IEEE Pacific Visualization Symposium (PacificVis), Bangkok, Thailand, 2019, pp. 148-157.

59. Ivakhnenko, A.G. Samoorganizuyushchiyesya sistemy raspoznavaniya i avtomaticheskogo upravleniya (Self-organizing pattern recognition system and automatic control). - K .: Technique,1969, 395p.
60. 60.Vapnik, V.N. Statistical Learning Theory. – New York: Wiley, 1998, 736p.
61. Kirichenko, M.F., Krak, Yu.V., Polishchuk, A.A. Pseudo inverse and projective matrices in problems of synthesis of functional transformers // Kibernetika i Sistemnyj Analiz., 40 (3), 2004, pp. 116-129.
62. Ruan, Y., Fox, G. A robust and scalable solution for interpolative multidimensional scaling with weighting, Proceedings of the 2013 IEEE 9th International Conference on e-Science ser. ESCIENCE '13, 2013, pp. 61-69.
63. Huang, S., Ward, M. O., Rundensteiner, E. A. Exploration of dimensionality reduction for text visualization, CMV '05: Proceedings of the Coordinated and Multiple Views in Exploratory Visualization, 2005, pp. 63-74.
64. Cox, T.F., Cox, M.A.A. Multidimensional scaling, 2nd ed. Chapman and Hall / CRC. – 2001. – 328p.
65. Krak, I.V., Kudin, G.I., Kulyas, A.I. Multidimensional Scaling by Means of Pseudoinverse Operations // Cybernetics and Systems Analysis, 55(1), 2019, pp. 22-29
66. Barmak, O., Krak, Y., Manziuk, E. Characteristics for choice of models in the ensembles. Proceedings of the 11th International Conference of Programming UkrPROG 2018 Kyiv, Ukraine, May 22-24, 2139, 2018, pp 171-179.
67. Manziuk, E. A., Barmak, O.V., Krak, Iu.V., Kasianiuk, V.S. Definition of information core for documents classification. Journal of Automation and Information Sciences, 50(4), 2018, pp. 25-34.

М.А. Муканова, И.В. Крак, А.А. Куандыков, А.С. Сагалова, Д.А. Байбатыров

Использование визуальной аналитики для разработки человеческих и машино-центрических моделей: обзор подходов и предлагаемых информационных технологий

Аннотация. Использование визуальной аналитической системы в машинном обучении – основа интеграции человека и применения его интеллектуальных возможностей при построении моделей. В то же время визуальная аналитика используется для расширения человеческих знаний и используется как инструмент исследования. В статье используются формы и цели применения рабочего процесса визуальной аналитики для формирования конечного продукта. Рабочий процесс делится на ориентированный на человека и ориентированный на машину, чтобы построить модель в качестве процессора информации и механизма принятия решений. Модели строятся на основе конечного пользователя, которым может быть машина или человек. Исследуются концепции построения моделей и роль машин и людей в этих процессах. Предлагается практическая реализация классификационной информационной технологии в изучаемой концепции «использование противоположной модели» в машинно-ориентированном рабочем процессе визуальной аналитики с применением машинной модели. В основе этой модели лежит модель, созданная и используемая человеком. Для классификации данных используются интеллектуальные способности человека. Границы классов определяются человеком, а затем проецируются в гиперпространство атрибутов с формированием модели классификации, которую запускает машина. Информационные технологии позволяют машине использовать модель, созданную для людей.

Ключевые слова: визуальная аналитика, классификация, ментальная модель, формальная модель, уменьшение размерности, визуализация информации

М.А. Муканова, И.В. Крак, А.А. Куандыков, А.С. Сагалова, Д.А. Байбатыров

Адам және машина-орталық үлгілерді дамыту үшін визуальдық талдауларды пайдалану: ұсынылған ақпараттық технологияларға және әдістерге шолу

Аңдатпа. Машиналық оқытуда визуалды аналитикалық жүйені қолдану адамның интеграциясы мен оның интеллектуалды мүмкіндіктерін модель құруда пайдалану үшін негіз болып табылады. Сонымен бірге визуалды аналитика адамның білімін кеңейту үшін және зерттеу құралы ретінде қолданылады. Біз түпнұсқалық өнімді қалыптастыру бағытында визуалды аналитикалық жұмыс процесін қолданудың нысандары мен мақсаттарын зерттейміз. Ақпараттық процессор және шешім қабылдау механизмі ретінде модель құру үшін жұмыс процесі адамға және машинаға бағытталған болып бөлінеді. Модельдер машина немесе адам болуы мүмкін соңғы пайдаланушының негізінде құрастырылған. Модель құру тұжырымдамалары және машиналар мен адамдардың осы процестердегі рөлі зерттелген. Зерттелген «қарама-карсы модельді қолдану» тұжырымдамасында ақпараттық технологияларды классификациялаудың машиналық модельді қолдану үшін машиналық бағытталған визуалды аналитикалық жұмыс процесінде тәжірибелік енгізу ұсынылады. Бұл модельдің негізін адам қалыптастырған және қолданатын модель құрайды. Мәліметтерді жіктеу үшін адамның интеллектуалды қабілеттері есепке алынады. Сыныптардың шекараларын адам анықтайды, содан кейін машина қолданатын жіктеу моделін қалыптастыра отырып, атрибуттардың гипер кеңістігіне шығарылады. Ақпараттық технологиялар машинада адамдарға жасалған модельді пайдалануға мүмкіндік береді.

Түйінді сөздер: визуалды аналитика, жіктеу, ақыл-ой моделі, формальды модель, өлшемді азайту, ақпаратты визуализациялау

About authors

Mukanova Makpal Abusadykova - 2-year, PhD

Kuandykov Abu Abdykadyrovich - Full Professor, IITU, Almaty, Kazakhstan,

Yu.V. Krak - Doctor of Physical and Mathematical Sciences, Professor, Ukraine,

Sagalova Assel Sanatkyzy - 2-year, PhD,

Baibatyrov Doszhan Asylkanovich - 2-year, PhD

УДК 004.041

Myrzakanov D.U.*, Kuandykov A.A.

International Information Technology University, Almaty, Kazakhstan,

ИСТОЧНИК ИНФОРМАЦИИ И БОЛЬШИЕ ДАННЫЕ

Аннотация. *В этой статье освещены основные понятия больших данных и источники информации в логистике. Логистика является неотъемлемой частью транспортировки и складских помещений вплоть до интернет-магазинов. Краткая информация о логистике описывается в современных процессах логистики. Большие данные показывает область применения и популярность данного направления. Правильное использование источников информации отражает корректное понимание как разбираться в любой области, анализируя технологии, подходы, методы, алгоритмы и технологии разработки.*

Ключевые слова*: логистика, виды информации,* big data *(большие данные).*

Введение

Источник информации – объект, определяющий происхождение информации, а также объект, определяющий происхождение информации; единый элемент подмножества опреде-