Semiotic Approach in the Development of Interactive Visual Analytics Systems[*]

Alena Zakharova[1][0000-0003-4221-7710], Evgenia Vekhter[2][0000-0003-0604-0399]
and Aleksey Shklyar[2][0000-0003-4442-7420]

[1] Bryansk State Technical University, Bryansk, Russia
zaa@tu-bryansk.ru
[2] Tomsk Polytechnic University, Tomsk, Russia
vehter@tpu.ru, shklyarav@tpu.ru

Abstract. Increasing information saturation in all spheres of life entails accumulation of large amounts of data that must be perceived, processed and decisions must be made on their basis. Therefore, the issues of interaction with data and of visual analytics require solutions at a new level. The authors investigate communication between the user and the data, highlight a number of negative trends that prevent effective use of visualization in solving practical problems, formulate aspects of subjective influence on the decision-making procedure, namely those leading to significant decrease in the efficiency of visual analytics systems.

The paper proposes an approach to the problem of reasonable use of existing and potential visualization capabilities for solving data analysis problems and making control decisions. A significant factor hindering the development of visual analytics is a lack of a model for coordinated use of computational and subjective resources corresponding to the technical level of computer visualization capable of ensuring close communication between the researcher and the available data. The paper describes an approach based on the concept of visual communication, the properties of which are determined on the basis of a number of key concepts of semiotics and linguistics.

Keywords: Visualization, Semiotics, Visual Communication, Data Analysis.

1 Introduction

The continuously increasing saturation of information environment of our time is a factor influencing the effectiveness of use of new data and knowledge both in everyday life and in scientific research. The availability of up-to-date digital data as well as the possibility of their (practically) unlimited preservation, at a certain stage, made

vital the task of preliminary analysis, data formalization, aimed at extracting new knowledge and reducing the amount of information involved in decision-making procedures for solving control problems in real scientific and production processes.

In practice, many existing approaches to the formalized data presentation provoke one of the possible negative situations. On the one hand, a simplified representation of data used to increase the speed of decision making leads to distancing of the decision maker from the initial data and makes the resulting solution dependent on subjective experience and knowledge, as well as on the correctness of the formalized representation. In the opposite case, tools are used to ensure the availability of both the initial data and the results of their subsequent analysis. For situations involving generation of many alternative decision hypotheses and their subsequent verification, the use of such tools can significantly increase the amount of analyzed data and, as a result, complicate the final decision.

Additional circumstances, that often make a significant contribution to the efficiency of data analysis and making control decisions based on it, are simultaneous participation of more than one subject in this process and the necessity for selecting the most attractive decision hypothesis on the basis of joint analysis of data related to different subject areas and accumulated knowledge. In this case, the conflict of opinions becomes an obstacle, overcoming of which depends on circumstances that go beyond the boundaries of the current problem, for example, on coordination of participants' competencies or the priority of their opinions.

2   Interaction with data

Technologies for solving applied problems associated with the use of machine learning capabilities as a basis for making a final decision are deliberately left out of the discussion [1]. This is due to the need for development of technologies applicable to the analysis of non-standard data, including for prompt forecasting of the state of analyzed systems or for using new data that were not initially taken into account in the formal model of these systems. Thus, obtaining the required answer in data analysis tasks is considered as a result of subjective interaction of a conditional user (group) with a certain amount of data having a heterogeneous level of formalization [2].

Upon close examination, communication between the user and the data appears to be a poorly researched process, despite the fact that it is mandatory for analytical tools of any type. The increasing complexity of analysis problems and tools proposed for their solution leads to the emergence of significant requirements for methods of such communication including their implementation in the form of HCI (human-computer interface) tools. Proliferation of computer systems for data processing and analysis is gradually forming an expanding community of users who employ a specialized communication system to interact with information environment [3].

A distinctive feature of the development of such a system is borrowing ways of presenting information that are characteristic of other types of human activity. The variety of acceptable representation methods is based on the use of all possible senso-

ry types and their combinations. This enables transfer of various types of data that satisfies constraints of a specific task including achieving a desired transfer rate, interpretation time, the need to employ user experience and any others. Visualization tools are still the most common variant of technical solutions in the development of data analysis tools, because visual perception is rightly believed to have a number of necessary advantages. However, further progress in this direction depends on the answers to several questions.

3　Visualization problems

Review of existing and emerging data visualization tools, especially those used in science, social process management or business analytics [4; 5], allow recognizing a number of negative tendencies that impede the effective use of visualization in solving practical problems.

1. Application of many modern technical developments in the field of computer visualization is more caused by the need for their commercial promotion than the achievable effectiveness.
2. The developed HCI tools use conceptual solutions proposed many years ago and probably do not correspond to the increased complexity level of information communication tasks [6].
3. Peculiarities of communication between the user and the visualized data are described by regularities reliability of which needs to be clarified due to the introduction of technical means that implement new and significantly different conditions of visual perception.
4. The set of perception peculiarities (perceptual portrait) of a potential user of visualization tools, the level of his preliminary awareness and practical cognitive models (understanding of which is necessary to select the most adequate means of data presentation) are constantly changing.
5. Reactions of the user of visualization tools, control of his emotional state and the directed use of perceptual psychology are unemployed reserves of visual analytics tools.
6. Active private use of computer and information technologies transforms the conceptual apparatus and reliability criteria belonging to the user, which can affect interpretation of visualized data [7; 8].
7. Developers of visualization systems provide the user with an opportunity to choose a unique way of presenting data according to a set of preferences relevant to him; this leads to the need for this user to receive specialized training.

Thus, the development of visual data representation tools intended for the user to obtain a solution to a certain practical analysis problem requires a formalized description of communication between the user and the visualized data (hereinafter, visual communication) [9]. Availability of such a description will allow to determine the features of visual representation that maximize its potential through reasonable and purposeful use.

4   Visual communication

Special attention should be paid to the subjective side of visual communication with the analyzed data, since it determines the features of the achieved result, which distinguish the use of visual analytics from the use of machine learning tools. There are several aspects of subjective influence on the final decision-making procedure, ignoring or mistaken use of which leads to a significant decrease in the effectiveness of visual analytics systems:

- Uncertainty in the results of interpretation experienced by the user largely depends on the consistency of his own data representation (internal image) and the proposed visualization metaphor. Therefore, the use of expressive means known or understood by the user earlier should help to overcome one of the communication barriers.
- The procedure for verifying the information obtained as a result of visual communication, taking into account the peculiarities of visual perception, occurs as a multiple repetition of the search for an answer to the formulated question in an accessible visual data image, with the consideration of new information accumulated by the user at each step of communication. A need arises for a systematized and operational, from the user's point of view, access to the changing volume of unverified data.
- Doubt. The repeated procedure of comprehension (interpretation), depending on the culture of thinking inherent in the user, can last as long as desired until the image formed under the influence of new information is accepted as an element of the user's knowledge system.

There arises a problem of coordinating several ways of representing information: visualization tools used for the initial data, temporary fixation of unverified versions as well as images that are remembered and used in the future by the user as his own knowledge.

5   Semiotic Approach

Based on definitions of a number of terms used in linguistics, developers of visualization systems face several independent tasks that currently do not have substantiated and complete solutions (Figure 1).

Visualization sigmatics can indicate problems related to the selection or development of visually perceptible objects designed to convey the required meaning to the user. Visualization semantics, obviously, should determine principles of coordinated use of available visualization technologies to form a consistent system for displaying initial data and interpretation results. Visualization syntactics should focus on solving problems of effective combination of visual objects integrated in data images. Finally, visualization pragmatics corresponds to the study of peculiarities of understanding by an observer of a visual image or their sequence, including such complex situations as

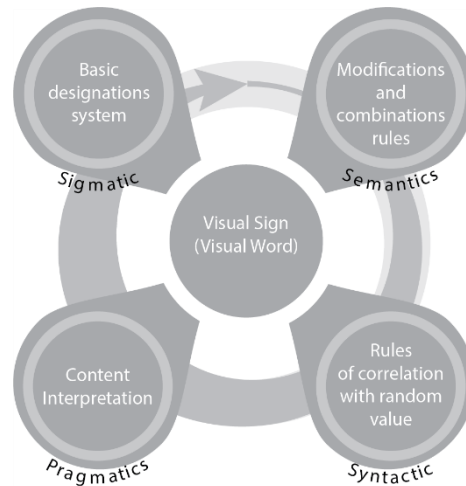visualization of abstract data or ideas that do not have perceptually-based representations known to the user.

Fig. 1. Compliance with linguistic concepts.

6   Pragmatics of visualization

A certain paradox is obvious in the development of visualization tools and their practical application associated with insufficient attention paid by developers to the pragmatics of visualization. Modern technical and software solutions used by computer visualization systems allow achieving a high level of complexity and detail of images reproduced in various ways. However, assessment of information content and its interpretability necessary for reasonable application of such systems cannot be obtained on the basis of the widespread concept of visual communication as a process similar to the functioning of communication systems.

As a consequence of peculiarities of user perception and thinking implemented at the time of visual communication (Figure 2), ambiguity of data visual interpretation arises which predetermines several approaches to the use of visualization tools that differ in tasks solved, visualization requirements and end user characteristics [10]. It is understood that the achievement of the goal of visual communication can occur in different ways, but in each case it is necessary to agree on all the above aspects. Definition of the preferred option corresponds to the use of visualization tools matching the data interpretation task pursued, but the choice is based on a number of criteria external to the data itself. For example, the choice of one out of many acceptable visualization techniques may occur with a preference based not so much on the features of the source data (format, volume, origin), but on the options familiar to the end user. This is consistent with the pragmatics of visualization and, as a result, makes visual communication more comfortable and effective (Figure 3).
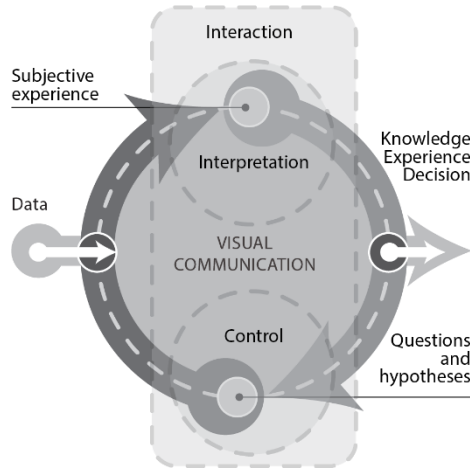
Fig. 2. Cyclical nature of visual communication.

The established hypothetical correspondence between the concepts of linguistics and some processes inherent in visualization functioning suggests the existence of similarities between speech and visual communication. An important consequence of this resemblance is preparedness of the user's thinking for the appearance in the samples of synthesized computer visualization of expressive techniques similar to the usual verbal communication: emotional statements, rhetorical questions, metaphors, ambiguous images, etc. Moreover, visualization, based on the modern understanding of its technical capabilities, can be considered as a communication option, similar to language communication, but more capacious, multifaceted and preferential for many practical situations.
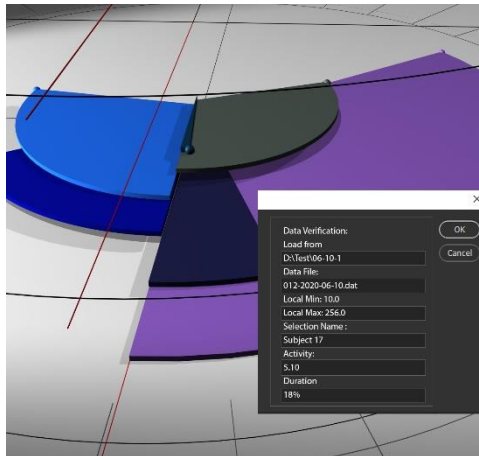

Fig. 3. Visualization of numeric multidimensional data using traditional representation.

7   Development of visualization tools

Focusing on the similarity between speech and visual communication, developers of information visualization tools, used among other things in control decision support systems, get a promising direction for the development of information visualization tools. Its feature is structured sequential transmission to the consumer (recipient) of information volumes that exceed the capabilities of both language messages and traditional computer graphics. However, the achievement of this result faces at least several problems that require special research.

Firstly, the developer of visualization tools must get systematic and deep understanding of the features of user communication with visualized information. This will allow to move from the extensive complication of visualization tools to communication focused on the user's perception and thinking.

Secondly, the question of how to overcome the communicative barrier associated with different levels of preliminary informing users, their cultural (including linguistic) and physiological characteristics remains practically unexplored. Existing visualization methods circumvent these issues by using a standardized (for historical and economic reasons) set of expressive techniques, which is currently becoming a negative factor that hinders the development of visual analytics.

Thirdly, the need naturally arises to develop techniques for visual data representation both borrowing effective means of verbal communication (imagery, comparisons, metaphors, attention retention, emotional coloring) and using unique visualization capabilities (parallel presentation; memorizing and operating not with attributes but images; interpretation of movement, etc.). Experimental evaluation of the effectiveness of new visualization techniques is still another independent and laborious task.

Based on the statements made, benefits arising from independent improvement of some characteristics of visualization tools corresponding to various aspects of the semiotic approach have been assessed. The most significant results obtained when solving a series of test problems for multidimensional data analysis (Table 1) are associated with improvement in visualization pragmatics as a result of organizing interactive communication between the user and visual analytics (Figure 4).

Table 1. Comparison of effectiveness of visualization tools with various organization of visual communication.

Series A (traditional representation), time, sec	Series B (interactive communication), time, sec	Effect, %
121,2	88,4	27,0
153,2	80,3	47,6
185,3	79,6	57,0
157,6	97,7	38,0
143,6	91,6	36,2
160,4	120,8	24,7

In addition, transition to the use of visualization systems with high information saturation makes it more important to solve a number of problems related to perceptual

psychology. The most common of them include passivity of thinking, leading to resistance to the new or unexpected, subjective fears and prejudices, changing emotional background, fatigue, loss of concentration or motivation [11].
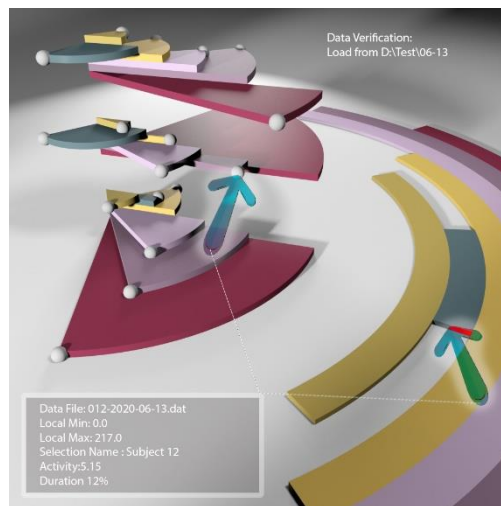

Fig. 4. Visualization tool supplemented with elements of interactive communication.

8　　Conclusion

The paper proposes an approach aimed at changing the development process of visual analytics tools, which are currently widely used both in business processes and in scientific research. The goal of the semiotic approach is to increase practical efficiency of existing visualization tools by explaining and reducing the number of erroneous results of both interpretation of visualized data and interaction of an unprepared user with specialized information processing systems. One of the significant conclusions obtained during the development of the semiotic approach applied to visual analytics systems is the possibility of more active use of knowledge and experience belonging to the user.

References

1. Blascheck, T., John, M., Kurzhals, K., Koch, S., Ertl, T.: VA^2: A Visual Analytics Approach for Evaluating Visual Analytics Applications. IEEE Transactions on Visualization and Computer Graphics. 22(1), 61–70 (2016). doi: 10.1109/TVCG.2015.2467871
2. Sacha, D., Sedlmair, M., Zhang, L., Lee, J.A., Weiskopf, D., North, S., Keim, D.: Human-Centered Machine Learning Through Interactive Visualization: Review and Open Challenges. In: ESANN 2016 proceedings, European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, pp. 641–646. i6doc.com publ, Bruges, Belgium (2016).

3. Massel L.V., Massel A.G., Ivanov R.A.: Integration of Visual Analytics, Cognitive Graphics and Semantic Modeling in Semiotic Intelligent Systems for Decision Support. In: Proceeding of International Workshop on Contingency Management, Intelligent, Agent-Based Computing and Cyber Security in Critical Infrastractures CM/IA/CS/CI-2016. pp. 19–22. Melentiev Energy Systems Institute; Russian Academy of Science Siberian Branch, Irkutsk, Russia (2016).
4. Crouser, R.J., Franklin, L., Endert, A., Cook, K.: Toward Theoretical Techniques for Measuring the Use of Human Effort in Visual Analytic Systems. IEEE Transactions on Visualization and Computer Graphics. 23(1), 121–130 (2017). doi: 10.1109/TVCG.2016.2598460
5. Batch A., Elmqvist, N.: The Interactive Visualization Gap in Initial Exploratory Data Analysis. IEEE Transactions on Visualization and Computer Graphics. 24(1), 278–287 (2018). doi: 10.1109/TVCG.2017.2743990
6. Levontin P., Walton Jo.L.: Visualising Uncertainty: A Short Introduction. AU4DM, London, UK. 2020.
7. Schofield D.: Courting the Visual Image: The Ability of Digital Graphics and Interfaces to Alter the Memory and Behaviour of the Viewer. In: Kurosu M. (eds.) Human-Computer Interaction. Theories, Methods, and Human Issues. HCI 2018. Lecture Notes in Computer Science, vol 10901, pp. 325–344. Springer, Cham (2018). doi: 10.1007/978-3-319-91238-7_27
8. Podvesovskii, A.G., Isaev, R.A.: Constructing Optimal Visualization Metaphor of Fuzzy Cognitive Maps on the Basis of Formalized Cognitive Clarity Criteria. Scientific Visualization 11(4), 115–129 (2019). doi: 10.26583/sv.11.4.10
9. Sacha, D., Stoffel, A., Stoffel, F., Kwon, B., Ellis, G., Keim, D.: Knowledge Generation Model for Visual Analytics. IEEE Transactions on Visualization and Computer Graphics 20(12), 1604–1613. (2014).
10. Zakharova, A., Vekhter, E., Shklyar, A., Zavyalov D.: Visual Detection of Internal Patterns in the Empirical Data. In: Kravets A., Shcherbakov M., Kultsova M., Groumpos P. (eds.) Creativity in Intelligent Technologies and Data Science. CIT&DS 2017. Communications in Computer and Information Science, vol 754, pp. 215–230. Springer, Cham (2017). doi: 10.1007/978-3-319-65551-2_16
11. Zakharova, A.A., Vekhter, E.V., Shklyar, A.A.: The Applicability of Visualization Tools in the Meta-Design of an Educational Environment. European Journal of Contemporary Education 8(1), 43–51 (2019). doi: 10.13187/ejced.2019.1.43