

USABILITY OF VISUAL DATA MINING TOOLS

Edwige Fangseu Badjio, François Poulet
*ESIEA Recherche,
Parc Universitaire Laval-Changé
38, rue des Docteurs Calmette et Guérin
53000 Laval*

Keywords: Usability, human-machine interaction, visual data mining

Abstract: Visual data mining is a field of research which needs knowledge from several domains: statistics, data analysis, machine learning, artificial intelligence, human-machine interfaces, data or information visualization. We are interested in visual data mining environment usability (man-machine interaction quality). This paper investigates how usability aspects can be incorporated in visual data mining environment so that usability can be taking into account during the design process of the tool without prototype evaluation tests which are time consuming at design stage. We have defined and we present here a set of criteria for improving visual data mining tools usability.

1 INTRODUCTION

In many research areas needing interfaces for treatments needs, for a long time only technical aspects were taken into account in the design of systems. A lot of problems result from that process, namely the end users difficulties to use the system in order to realize their task. For this purpose, we propose a new method for improving usability of visual data mining tools. In the actual stage of our research, this method consists of a set of criteria and strategies for setting up each criterion. We have made some experimentations of a subset of these criteria but we just present definitions and strategies here.

According to (Fayyad & al, 1996) data mining is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. Visual data mining consists of visualization use as a communication channel for data mining. For (Wong, 1999), visual data mining lies in tightly coupling the visualizations and analytical process into one data mining tool that takes advantage of the strengths of all worlds. Visual data mining is a recent research field.

In man-machine interaction research field, there are usability evaluation techniques (Nielsen, 2000), (Scapin et al., 1993). We took these techniques as a starting point for the development of our method.

The techniques used to improve usability of software in man-machine interface will be presented in the third paragraph, after usability definition, followed in fourth part of this paper by the method which we worked out.

2 USABILITY DEFINITION

There are several usability definitions:

(IEEE, 1990) defines usability as how easily a user can learn to operate, prepare inputs for, and interpret outputs of a system or component.

For (ISO 9241-11, 1998) usability is the way specified goals can be achieved with effectiveness, efficiency, and satisfaction in particular environments.

(Nielsen, 2003) define usability as a quality attribute that assesses user interfaces ease of use. The word "usability" also refers to methods improving ease-of-use during the design process.

Nowadays, in most of research teams, there is a usability evaluation test after the development of a product. That evaluation makes the quality of human-machine interaction evaluation possible.

3 USABILITY EVALUATION

Many user interfaces guidelines can be used to improve usability when applied in the design step. These guidelines could not be applied in visual data mining at all because visual data mining has some particularities. We have to study the visual data mining domain before the definition of criteria which will be appropriate for improving usability. We are going to present the visual data mining process, before this presentation, in the next sub-section, we present some methods for usability design products.

3.1 Design for usability

This part presents some techniques used in order to improve software usability. We have:

- Iterative design and evaluation: this method involves design, evaluation and redesign of the software,
- A method which implies the user in a model or a designed product usability evaluation,
- The last method implies the domain expert in usability evaluation of a model or a designed product.

These methods are time consuming. Our main idea for our work is to take visual data mining tools usability guidelines into account before starting the design process. The study of visual data mining process helped us for this purpose. Knowing that, usability performance is measured relative to users' performance on a given set of tasks, the measures are success rates (whether users can perform the task at all), the time a task requires, the error rate, users' subjective satisfaction (Nielsen, 2001). We have defined a list of usability metrics. Before explaining these metrics we are going to present the visual data mining process in the next sub-section.

3.2 The visual data mining process

There are different stages in the visual data mining process:

1. selection of the data to be exploited,
2. choice of the visualization method to use or passing to stage 4.,
3. data visualization,
4. choice of a visual data analysis method among those proposed by the system,
5. visualization of the results,
6. evaluation of the results (which must be easily understandable), followed by a possible return at stage 1 or 2,

7. analysis of the results considered as new knowledge,
8. return at stage 1 or stop.

Visualization is a stage of visual data mining process, which provides graphical displays, and animation on which investigator observations are based. The user of such a system is not intended to be a data mining or data analysis specialist but an expert of the data domain. In order to perform a significant analysis, the user must be helped because, due to his lack of statistical background, he may not be able to perform the right choices. The usability criteria we define are intended to help users.

4 USABILITY CRITERIA

4.1 Adaptability

4.1.1 Definition

The adaptability is the capacity of the system to be adapted to the user's needs without any explicit intervention from the user or its capacity to be reacted according to the context and the needs and preferences of the users.

4.1.2 Strategy

For setting up this criterion, we thought of the possibility for the user to personalize his interface. The purpose of user interface personalization is to take into account the user strategies or preferences. We also thought of the development of means available for taking into account the experiment level of the user (beginning, tested, occasional) like his profile.

4.2 Curability

4.2.1 Definition

Curability is the user capacity to correct a non-desired situation.

4.2.2 Strategy

Error rate, time required for a task execution are factors usually taken into account during man-machine tools usability evaluation. This criterion recommends visual data mining tools designers to display curatives means for the errors likely to occur

in their environment so that usability evaluation will be successful.

4.3 Errors management

4.3.1 Definition

Error management refers to means allowing on one hand to avoid or reduce errors, and on other hand to correct them when they occur.

4.3.2 Strategy

It is a question here of setting up means to detect and prevent errors. For example, all the possible actions on the interface must be considered and more particularly the accidental supports keyboard keys so that not awaited entries are detected. Another case: if the data analysis method chosen by the user is not successful (method execution is not completed), it is necessary to be able to propose another method to the user without any system crash. The user must be able to execute another algorithm for data analysis, the method selection tool must be able to give not only the most adequate algorithm to the problem resolution but also the list of ranked algorithms. Classification is done according to algorithms evaluation criteria.

4.4 Feedback

4.4.1 Definition

Feedback recommends that after achievement of an action, the system provides an answer to the user informing him about the accomplished action and its result, this, with a deadline for reply suitable and homogeneous according to types of transactions.

4.4.2 Strategy

The visual data mining cycle can be time consuming, depending on the size of the treated data. Some information showing the user that the treatments are going on, the progress report of the treatments should be provided to the user.

4.5 Guidance

4.5.1 Definition

User Guidance refers to the available ways to advise, orient, inform, instruct, and guide the users throughout their interactions with a computer.

Good guidance facilitates learning and use of a system by allowing the users: to know at any time where they are in a sequence of interactions, or in the accomplishment of a task; to know what the possible actions are as well as their consequences; and to obtain additional information (possibly on demand). Ease of learning and ease of use that follows good guidance lead to better performances and fewer errors. (Bastien et al., 1993)

4.5.2 Strategy

In the visual data mining process, users have to select an analysis method for the resolution of their problem. Algorithm selection is an exploratory process highly dependent on the analyst's knowledge of the algorithms and of the problem domain. Our end users are not experts of data mining or data analysis but an expert of the data domain. When making choice of data analysis method to execute, they have to execute the set of available methods and select the most adequate algorithm for the given problem. Running an algorithm for a given task is time consuming, especially when complex tasks are involved. Our strategy here is to provide help to the user for the selection of the most adequate algorithm for a given task. A trivial solution for this problem is to determine the best analysis algorithm. But, the No free Lunch theorem (Wolpert et al., 1996) states that if algorithm A outperforms algorithm B on some cost function then there must exist exactly as many other fonction where B outperforms A.

Given the wide variety of analysis method available the selection of the right algorithm for a problem is an important issue. There are some research works from that field. For example we have the StatLog and the METAL projects. As far as METAL is concerned, several approaches have been used. These approaches investigate the problem of using past performance information to select an algorithm for a given problem. For this purpose, knowledge about past performance information are stored and the authors use the approaches such as: ontologies, case based reasoning, induction algorithm to predict the performance of a given algorithm on a task. For new cases these approaches proceed by successive approximations and so lead to a loss of information.

We chose a multi agents system for the evolutionary needs of the system. We thus will be able to use the assets of this paradigm, and more particularly the autonomy of the agents as well as the possibilities to distribute our treatments.

4.6 Multiplicity of returned

4.6.1 Definition

This criterion refers to the system capacity to provide several visualization methods.

4.6.2 Strategy

Current state-of-the-art data visualization or information visualization propose many data representation techniques: geometric techniques, icon-based, pixel oriented techniques, hierarchical techniques, graph-oriented techniques, distortion techniques, and dynamic or interaction techniques. Everyone agrees on the fact that none of these methods is better than the others in all cases. For the same set of data, it is a question of envisaging several possible methods of visualization.

4.7 On-line help

4.7.1 Definition

This criterion relates to the documentation availability of the user.

4.7.2 Strategy

Data visualization is based on graphical methods. A possible approach to set up this criterion is to make appear contextual texts to the screen to inform the user or to provide him explanations associated with the visualisation method used or about the choice of a split-criterion for decision tree.

4.8 Plasticity

4.8.1 Definition

This criterion refers to the system ability to dynamically react to fluctuation on resources while preserving ergonomic continuity.

4.8.2 Strategy

The transition from one stage to another one in the visual data mining process must be perceived by the system as well as the transition to the analysis of a data base different from the previous one.

4.9 Training data re-use

4.9.1 Definition

Training data re-use criterion refers to the possibility of pursuing a data mining process. Particularly, the outputs of the system can be used like data input.

4.9.2 Strategy

Visual data mining process included preprocessing, treatment, postprocessing. For example a decision tree can be interactively constructed (instead of the usual automatic approach). The first representation of the data corresponds to the initialization of the decision tree construction algorithm. The tree growing can be stopped at any level of the construction. An important and untreated aspect in this type of environment is the possibility for the user to go on with a previously stopped treatment. Indeed, if the user stopped the visual data-mining algorithm before the task is finished, the idea here is to enable him to continue in the process without having to start again at the initial stage.

5 CONCLUSION

As we have seen, data mining tools users can be data mining experts, visualization experts or expert of the data domain. Our visual data mining approach is dedicated to the experts of the data domain. We have presented some work about the usability of visual data mining tools, in order to develop a software program able to help that type of users, and to avoid any redesign step generating waste of time and high production costs (without however guaranteeing the performances). To get this usability, we have established criteria having to be taken into account for a development of reliable and useful software. These criteria are applied to visual data mining. We have started the development of the corresponding software program: a visual data-mining environment dedicated to the data specialist. The criteria we have defined can be used as a basis for other achievements.

REFERENCES

- Bastien, J. M. C., Scapin, D. L., & Leulier, C. 1999. The Ergonomic Criteria and the ISO 9241-10 Dialogue Principles: A pilot comparison in an evaluation task.
- Bastien, J.M.C., Scapin, D.L. 1993. *Critères ergonomiques pour l'évaluation d'interfaces*

utilisateurs. Rapport technique INRIA n° 156, Juin 1993, INRIA : Le Chesnay.

David H. Wolpert, William G. Macready, 1996. No Free Lunch Theorems for Optimization, *IEEE Transactions on Evolutionary Computation*, 1:67-82.

Hilario M. and Kalousis A., 2002. Fusion of meta-knowledge and meta-data for case-based model selection. *In Proceedings of the 5th European Conference on Principles of Data Mining and Knowledge Discovery (PKDD-01), Springer-Verlag, Freiburg, Germany.*

Nielsen J., 2000. Designing Web Usability: The Practice of Simplicity, New Riders Publishing, Indianapolis.

Nielsen J., 2003. Usability 101, (<http://www.useit.com/alertbox/20030825.html>)

Wong P. C., 1999 Visual data mining, *IEEE Computer Graphics and Applications* 19(5), p 20-21.



SciTeP
Science and Technology Publications