CONTEXT OF USE ANALYSIS
Activity Checklist for Visual Data Mining

Edwige Fangseu Badjio, François Poulet
ESIEA Pôle ECD, 38 Rue des Docteurs Calmette et Guerin, 53000 Laval France

Keywords: Visual data mining, activity theory, context of use, requirements.

Abstract: In this paper, emphasis is placed on understanding how human behaviour interacts with visual data mining (VDM) tools in order to improve their design and usefulness. Computer tools that are more useful assist users in achieving desired goals. Our objective is to highlight quality in context of use problems with existing VDM systems that need to be addressed in the design of new VDM systems. For this purpose, we defined a checklist based on activity theory. The responses provided by 15 potential users are summarised as design insights. The users respond to questions selected from the activity checklist. This paper describes the evaluation method and shares lessons learned from its application.

1 INTRODUCTION

Computer capabilities offer means to store very large databases. All these databases are not useful if at least a part of information they contain is not extracted. It is the goal of Knowledge Discovery in Databases (KDD) process. According to (Fayyad et al., 1996), KDD is the non-trivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. Several KDD packages offer means to visualize data and KDD results.

In this paper, emphasis is placed on understanding how human behaviour interacts with visual data mining (VDM) tools. A considerable effort has been done to enhance KDD tasks performance. High performance algorithms have been created (Grossman and Yike, 2002), (Freitas and Lavington, 1998). A considerable effort has been done to enhance KDD tasks performance. Also, decision support systems for the appropriate selection and parameterization of such techniques have been provided (Michie et al., 1994), (Fangseu Badjio and Poulet, 2004a). In spite of good results obtained by VDM tools, there is less interest on human factors. Few investigations have been done about for example what would occur when these powerful systems will be transferred from the research laboratories to a real use and on a large scale? What think the end users to who are intended the data mining tools? (Whiteside et al., 1988) and (Wolf, 1989) found that although many products performed well in their laboratory experiments, they did not work when transferred to the real work. They put this down to the fact that the research often overlooked something crucial to the context in which the product would be used (Maguire, 2001).

Recommendations for more usable VDM tools and methods allowing the evaluation of this type of tools using a combination of human-centred, task, environment oriented approaches, and general knowledge of Human Computer Interaction (HCI) design have been proposed (Fangseu Badjio and Poulet, 2004b, 2005a, 2005b). In this contribution, we address the VDM context of use analysis for usability evaluation. Taking human factors into account in software evaluation involves considering not only users but also tasks and context of use. Consequently, the evaluation of VDM tools requires the analysis of the context of use to understand the impact of the artefact.

The objective here is to improve the quality of VDM tools in the design step, increase user productivity and decrease user errors. For this purpose, we use a social science theory named activity theory (AT) which is a philosophical framework used to analyse and model human activity. Activity theory provides a robust analytical framework and a common vocabulary for describing human activity in context (Nardi, 1996). Context of use analysis is a technique that assists software engineering. It is performed in order to resolve

Fangseu Badjio E. and Poulet F. (2006).
CONTEXT OF USE ANALYSIS - Activity Checklist for Visual Data Mining.
In *Proceedings of the Eighth International Conference on Enterprise Information Systems - HCI*, pages 45-50
DOI: 10.5220/0002456100450050

problems in the software development process. VDM tools design could benefit from activity theory approach to analyse the transformative relationship between users of a computer system and the activity in which there are engaged. For this purpose, we have to establish the means by which the concepts presented in activity theory can be incorporated in VDM tools design.

The responses provided by 15 potential users are summarised as design insights. The users answer questions selected from the activity checklist.

The overview of this paper is the following: firstly, we present the VDM domain and the means allowing an overall analysis of such software and the activity checklist. Lastly, there is a case study before conclusion and future works.

2 THE VDM DOMAIN AND QUALITY OF USE

The first research works treating VDM appear at the end of 1990s (Cox et al., 1997), (Inselberg, 1998). VDM relates to the use of visualisation as communication channel for the discovery of correlations in data. Being given the increasing quantity of the data available in the world, a point of interest of the field consists in the development of visual representation techniques for massive data sets and innovative computing techniques. For example, (Keim, 1996) has proposed a pixel based method from which an interactive method for decision trees construction is derived (Ankerst et al., 1999). Other visualisation methods (self organizing map (Deboeck and Kohonen, 1998), 2D matrices (Witten and Eibe, 2000), and parallel co-ordinates (Yujin et al., 2004)) have been used for VDM.

Except some few works, the VDM field first results relate much more to technical aspects development. (Grinstein et al., 1997) for example was interested in the technical quality of visual representations used in data mining field. Recently, an interest was carried towards VDM tools usability (quality of use). Indeed, in spite of their necessity, VDM tools have utility only if the end-users accept to use them. In general, the software acceptability is related to its quality. In the process of determining the software quality, the end-user is the most indicated. The utility of a VDM tool relates to the adequacy existing between the functions provided by the system and those necessary to the user in order to achieve the VDM tasks assigned to him.

There are many design guides ensuring the quality of software, we have for example HCI standards (ISO, 1998), and a set of ergonomic criteria (Nielsen and Landauer, 1993), (Bastien et al., 1999). Software quality assessment is done by evaluation which should be considered at all life cycle stages (design techniques, prototyping and implementation techniques). The evaluation can be completed by an expert, the end user or can be model or task based.

We study the qualitative analysis of a VDM system for which there could be several approaches: visual representation oriented approach, data oriented approach, task oriented approach, interface oriented approach and then context (activity, event, regulation) oriented approach. We are interested in the last one. According to (Suchman, 1987), context can be seen as a resource upon which users can draw. It is important to evaluate computer systems in context. Contextual analysis helps explain the reasons for an outcome, clarifies a situation, make a situation more specific. Many authors recommended representative evaluations in context; we have for example (Bevan and Macleod, 1994), (Beyer and Holtzblatt, 1999).

3 ACTIVITY THEORY

3.1 Definition

Activity theory is a theoretical framework that provides concepts and a vocabulary to analyse and understand human activity in context. Activity theory provides an alternative formulation to information processing as to how people learn and society evolves, from a material perspective, based on the concept of human activity as the fundamental unit of analysis. According to (Nardi, 1996), activity theory is a powerful and clarifying descriptive tool rather than a strongly predictive theory. The pioneers of activity theory are Vygotsky and Leont’ev (Leont'ev, 1978), further development have been done by Engeström (Engeström, 1987). Two basic ideas animate activity theory: (1) the human mind emerges, exists, and can only be understood within the context of human interaction with the world; and (2) this interaction, that is, activity, is socially and culturally determined (Kaptelinin et al., 1999).

There are six main principles within activity theory:

Unity of consciousness and activity (state that human learn by doing and the human consciousness is formed by interaction with external world),

Object-orientedness (every activity has an object (purpose) and is performed in order to achieve a goal),

Mediation (the principle of mediation states that in any activity there will be tools involved, both physical and psychological),

Internalisation/Externalisation (Internalisation is the process by which mental representations are formed by carrying out external actions. Externalisation is the opposite: where mental representations are manifested in external actions),

Hierarchical structure (each activity can be decomposed into actions and operations),

Development (this principle explains that activity can only be understood through analysis of its developmental transformations).

3.2 Activity Theory and VDM

VDM tools cannot be introduced in KDD domain as a powerful decision support tool without analyzing the impact from the users' point of view. Activity theory is concerned with understanding the relationship between consciousness and activity (Nardi, 1996). We must understand how visual data miners can perceive the VDM tool and their impact on the work to be achieved. The activity theory is a general framework for studying different forms of human activity as development processes (Kuutti, 1996). Within our context, the activity theory is particularly interesting in that it postulates that an activity has to be analyzed both as an individual process and a social process. Activity theory has the potential to provide a shared vocabulary for designers and to resolve some of the problems facing the VDM field. More precisely, activity theory offers tools for defining user needs and evaluating usability. Current trends in VDM can then be understood by: identifying the structure of activities that undergo transformations, revealing the most important contradictions typical of the current stage of development of the above activities, analyzing the ways technological affordances and limitations can influence the above contradictions, considering possible scenarios of resolving the contradictions with the help of technology, and anticipating further contradictions.

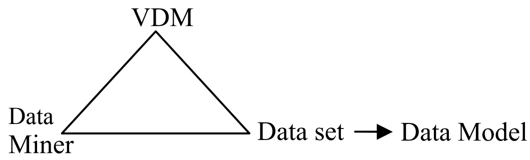

Figure 1: Structure of the activity (Visual Data Mining).

The Figure 1 presents the application of the basic activity theory framework to VDM domain. The subject of the activity would be the data miner, the mediation tool would be VDM, the object to be transformed would be a dataset and the result would be data models ready for usage.

In order to understand the VDM system's context of use, we focus on the triad described by Figure 1. We analyse the impact of VDM (considered as tool) on Data Miner and specify in a systematic way the characteristics of the users, the tasks there will carry out, and the circumstances of use. First steps are description of the VDM software and characterisation of its context of use. The Figure 1 shows also that the treatments of datasets will have an impact on Data Miners. The activity theory helps to capture the context of software. Features of activity theory that have implications for visual data mining tools include recognition of actions, mediator, historicity, constructivism, dynamics and others. Finally, Activity Theory offers a promising avenue for providing a framework and theories to deal with the developmental and dynamic features of human practices.

The features and benefits of the activity theory are: identify the stakeholders in the process; ensure that technology is designed to benefit the user; work toward alignment between users' rewards and business needs; work toward alignment between the rewards of the designers of the device and the business needs.

By using activity theory conceptual framework, we ensure that quality studies reflect the context of use.

Usually for context of use analysis, a range of people who have a stake in the development bought together at the context meeting. Instead of contextual meeting, our proposition suggests the analysis of existing tools in order to find out if the current user needs are satisfied. For this purpose, a contextual analysis checklist has been defined. It is a kind of requirement elicitation process. A requirement is a criterion that a system must meet; a desired feature, property, or behaviour of a system. There are functional and non-functional requirements. Functional requirements describe the interactions between the system and its environment independent from implementation. Non-functional requirements are the user visible aspects of the system not directly related to functional behaviour.

For recommendations elicitation, we analyze existing software usefulness and usability. Our objective is to highlight quality of use problems with existing systems that need to be addressed in the design of new systems. The goals are: minimizing human information processing, minimizing cognitive demand on the users and avoiding errors or poor performance. We performed post evaluation of existing VDM tools which objectives are: assessing whether stated development goals have been met and suggesting strategies for future design changes.

The next section presents the context of use analysis approach.

4 ACTIVITY THEORY BASED CONTEXTUAL ANALYSIS METHOD

The proposed analysis approach based on activity theory is a checklist which helps to ask meaningful questions about context of use analysis in VDM field. The following paragraphs present the main topics of the activity checklist knowing that context of use analysis aims at identifying and addressing user needs that may not be obvious. For this purpose, during the task, we are interested in the following subjects concerning the users: understand task and purpose, choose appropriate strategy, attention, anticipation and prediction, comprehension assistance, hesitation, confusion, new method integration. After the task, we are interesting in user assimilation and competent feelings.

Object examination: Every visual data mining session has an objective and is performed in order to achieve a goal. In this part, we examine if there are tasks that users will want to perform that are not currently supported by visual data mining tools.

Support for Internalisation/Externalisation and Learning: Internalisation is the process by which mental representations are formed by carrying out external actions. Externalisation is the opposite: where mental representations are manifested in external actions. The sub-topics of this topic are: user training for software usage, contain of documentation, documentation and software coverage, user background requirement.

Support for Actions and Operations: in this topic, we are interested in how the user associates the correct action with the effect to be achieved, how the user notice the correct action is available. Other points of interest concern on-line help, paper based user guide, tool installation (network, floppy or CD) without assistance.

Support for Mediation: the principle of mediation states that in any activity there will be tools involved, both physical and psychological.

Development: users activities can only be understand through analysis of its development transformation.

5 CASE STUDY

5.1 Users

The evaluation is based upon the use of VDM software by 15 master degree students in computer science and business administration, there are volunteers. During the recruitment process, these volunteers were asked to specify the following details: experience and training with data mining and VDM, experience and training with WEKA (Witten and Eibe, 2000). They were also asked to specify experience and training with graphical representation and interaction with visualisations, attitude to task and product.

The selected volunteers have no experience with the product and basic knowledge about data mining, VDM, graphical representations and visual interaction.

5.2 Task

The evaluation consisted of a single task: interactive construction of a decision tree starting from representations of the datasets described in table 2 from the UCI (Blake and Merz, 1998). Data sets for this kind of evaluation can also be found in other repositories (Jinyan and Huiqing, 2002). The decision trees allow partitioning a great quantity of data in small groups or parts by application of a series of decision rules.

Table 1: Datasets characterisation.

Dataset name	Nb of records	Nb of Attributes	Nb of classes
Ionosphere	351	32	2
Vehicle	846	18	4
Segmentation	2310	11	7
SatImage	6435	36	6
Letters	20000	16	26

The purpose of this context of use study was to assess through a set of interviews the design of a WEKA module for VDM named UserClassifier and to identify areas for improvement. The volunteers answer questions whether they were satisfied with UserClassifier or not and they stated some improvements directions. Interviews were conducted following a semi-structured guide extracted from the activity checklist. Some examples of the semi-structured guide questions are:

Describe your use of UserClassifier for VDM.

What information and functions of the UserClassifier module do you find most useful?

What information needs are not currently being met by the UserClassifier module?

How could any of these unmet information need be met by the UserClassifier module?

5.3 Environment

Every user worked alone and assistance is provided about the operating system if requested, although, no assistance is given about WEKA.

5.4 Evaluation Results

According to the volunteers, more than 40% of their needs are not obvious in UserClassifier module.

Object examination: tasks needed by the end users are not currently provided by the tool. For example, only one algorithm and only one visualisation method are implemented in UserClassifier module for VDM. The users can not assess preferred analysis methods or visualisation tools. It is not possible to access various data set formats, only the arff format is supported by the tool.

Support for Actions and Operations: The users are not oriented (guided), it misses the on line help, the contextual menus (focus, overview, detail on demand), the user manual.

Support for Mediation: the elements disposal on the screen is very good; graphics and colours are well used but it is not possible to reuse training data sets. The users' workload is high for the treatment of very large data sets. Only one 2D matrix (representing 2 attributes and the class) can be displayed at the same time on the screen, it is impossible to have the overall contextual information in the data sets in the same visualisation. It is impossible to obtain the correlations between the attributes in the data set without a lot of data explorations.

Development: the system ease of use and ease of learning is recognised by the evaluators. The most difficult for the volunteers was to achieve the construction of the decision tree with very large data sets and to obtain an appropriate tree.

The results of this evaluation enable the designers of UserClassifier module to improve the aspects related to the context of use usability (assistance modules, user manual, several alternatives possible with regard to data analysis methods and data visualisation, cognitive aspects of visualisation for data mining, user preferences).

6 CONCLUSION

We have proposed an innovative approach: an activity checklist based on activity theory for context of use in VDM field. As stated by (Maguire, 2001), there are several benefits of context of use in software design: it provides an understanding of the circumstances in which a product will be used, it helps to identify user requirements for a product, it helps address issues associated with product usability and provides contextual validity of evaluation findings.

Visual data mining tools are useful and necessary. In KDD domain, it is innovative and stimulating to be able to treat data sets with millions of observations. The available algorithms that aim at performing this kind of treatments are developed in laboratories. Generally, the final end users of those algorithms are not the designers. In our works, we are interested in all the things which could happen after passing VDM tools from laboratories context to a real context of use, in the absence of the tool designer who is able for example to modify the program in order to take account of new functionalities. It is then necessary to develop a set of standards for the development of these tools, by taking account of user, task, activity, and context of use.

REFERENCES

Ankerst M., Elsen C., Ester M., Kriegel H.-P., 1999. Visual classification: An interactive approach to decision tree construction. In *Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp.392-396.*

Bastien J.M.C., Scapin D.L., Leulier C., 1999. The ergonomic criteria and the ISO/DIS 9241-10 dialogue principles: a pilot comparison in an evaluation task. In *Interacting with Computers, vol. 11(3), pp.299-322.*

Bevan N., Macleod M., 1994. Usability measurement in context. In *Behaviour & Information Technology, vol. 13(1-2), pp.132-145*.

Beyer H., Holtzblatt K., 1999. Contextual design. In *ACM* interactions, vol. 6(1), pp.32-49.

Blake C., Merz C., 1998. UCI Repository of machine learning databases, [www.ics.uci.edu/~mlearn/MLRepository.html]. Irvine, *University of California, Department of Information and Computer Science*.

Cox K.C., Eick S.G., Wills G.J., Brachman R.J., 1997. Visual Data Mining: Recognizing Telephone Calling Fraud. In *Data Mining and Knowledge Discovery, vol. 1, pp. 225-231*.

Deboeck G., Kohonen T., 1998. *Visual Explorations in Finance with self organizing maps*, Springer-Verlag.

Dillon A., Morris M., 1966. User acceptance of information technology: theories and models. In *M. Williams (ed.), Medford, NJ: Information Today, Vol. 31*.

Engeström, Y., 1987. *Learning by Expanding: An Activity-Theoretical Approach to Developmental Research*. Helsinki: Orienta-Konsultit Oy, Finland.

Fangseu Badjio E., Poulet F., 2004a. A decision support system for data miners. In *AISTA'04, International Conference on Advances in Intelligent Systems - Theory and Applications in cooperation with IEEE*.

Fangseu Badjio E., Poulet F., 2004b. Usability of Visual Data Mining Tools. In *ICEIS'04, 6th International Conference on Enterprise Information Systems, vol.5, 254-258*. ICEIS Press.

Fangseu Badjio E., Poulet F., 2005a. Towards usable visual data mining environments. In *HCII'05, 11th International Conference on Human-Computer Interaction*.

Fangseu Badjio E., Poulet F., 2005b. Visual data mining tools: quality metrics definition and application. In *ICEIS'05, 7th International Conference on Enterprise Information Systems, vol. 5, pp.98-103*. ICEIS press.

Fayyad U. M., Piatetsky-Shapiro G., Smyth P., 1996. (ed) *Advances in Knowledge Discovery and Data Mining*. AAAI Press / MIT Press, Menlo Park, CA.

Freitas A., Lavington S. H., 1998. *Mining Very Large Databases with Parallel Processing Series,* International Series on Advances in Database Systems, vol. 9.

Grinstein G. G., Hoffman P., Laskowski S. J., Pickett R. M., 1997. Benchmark Development for the Evaluation of Visualization for Data Mining. In *Issues in the Integration of Data Mining and Data Visualization*, Workshop, Newport Beach, California.

Grossman R. L., Yike Guo, 2002. Parallel Methods for Scaling Data Mining Algorithms to Large Data Sets. In *Handbook on Data Mining and Knowledge Discovery, Jan M Zytkow, editor, pp.433-442*. Oxford University Press.

Hasan H., 2001. An Overview of Different Techniques for applying Activity Theory to Information Systems. In *Information Systems and Activity Theory: Theory and Practice (Ed, Hasan, H.)* University of Wollongong Press.

Inselberg A., 1998. Visual Data Mining with Parallel Coordinates. In *Computational Statistics Vol. 13(1), pp.47-63*.

ISO (International Organization for Standardization), 1998. *ISO 13407: Human-Centered Design Process for Interactive Systems*.

Jinyan L., Huiqing L., 2005. Kent Ridge Bio-medical Data Set Repository. *http://sdmc.lit.org.sg/GEDatasets*, accessed the 2nd October 2005.

Kaptelinin V., Nardi B. A., Macaulay C., 1999. The Activity Checklist: A Tool For Representing the "Space" of Context. *Interactions, Vol.6, pp. 27-39.*

Keim D.A., 1996. Pixel-oriented Visualization Techniques for Exploring Very Large Databases. In *Journal of Computational and Graphical Statistics*, *vol. 5(1), pp.58-77*.

Kuutti K., 1996. Activity Theory as a Potential Framework for Human-Computer Interaction Research. In *Nardi, B.A., (1996) (Ed) Context and Consciousness: Activity Theory and Human-Computer Interaction*. MIT Press.

Leont'ev A. N., 1978. *Activity, Consciousness, Personality*. Englewood Cliffs, NJ, Prentice Hall.

Maguire M., 2001. Context of use within usability activities. In *International Journal Human-Computer Studies vol. 55*.

Marghescu D., Rajanen M., Back B., 2004. Evaluating the Quality of Use of Visual Data-Mining Tools. In *ECITE'04, 11th European Conference on Information Technology Evaluation, pp. 239-250.*

Nardi B., (Ed.), 1996. *Context and Consciousness. Activity Theory and Human Computer Interaction*. MIT Press.

Nielsen J., Landauer T. K., 1993. A mathematical model of the finding of usability problems. In *INTERCHI'93, 4th International Conference on Human-Computer Interaction, pp. 206-213*. ACM Press.

Suchman L.A., 1987. *Plans and situated actions: The problem of human-machine communication*. Cambridge University Press.

Whiteside J., Bennett J., Holtzblatt K., 1988. Usability engineering: our experience and evolution. In *M. Helander, Ed. Handbook of Human Computer Interaction, pp.791-817*. Amsterdam: Elsevier.

Witten I. H., Eibe F., 2000. *Data Mining: Practical machine learning tools with Java implementations*. Morgan Kaufmann, San Francisco.

Wolf C. G., 1989. The role of laboratory experiments in HCI: help, hindrance or Ho-hum?. In *CHI'89, 6th conference on Human Factors in Computing Systems, pp.265-268*. ACM Press.

Yujin C., Qingyuan Z., Jianming W., 2004. Visual Data Mining Based on Parallel Coordinates and Rough Sets. In *ICITA'04, 2nd International Conference on Information Technology for Application*.