VidaMine: User-Centred Development of a Visual Mining Environment

Stephen Kimani, Stefano Lodi**, Tiziana Catarci*, Giuseppe Santucci*, Claudio Sartori***

* University of Rome "La Sapienza", DIS, Via Salaria 113, 00198 Rome, Italy
{kimani,catarci,santucci}@dis.uniroma1.it

** University of Bologna, DEIS, Via Risorgimento 2, 40136, Bologna, Italy
{slodi,csartori}@deis.unibo.it

Abstract

Tremendous technological breakthroughs have virtually revolutionized the world. One major consequence is that humans are confronted with ever-increasing already massive amounts of data at virtually every turn. Visual data mining (VDM) intends to exploit effective visual strategies in the entire process of mining knowledge. Though there have been many efforts in this area, most of them have paid little attention, if any, to usability studies. While ensuring that a VDM system involves the user in the entire mining process is indeed significant, ensuring that the user is involved in the system development process needs prior consideration. This work describes how a user-centered approach was adopted in the development of a VDM system named *VidaMine*.

1 Motivation

Tremendous technological breakthroughs have virtually revolutionized the world. One major consequence is that humans are confronted with ever-increasing already massive amounts of data at virtually every turn. On the other hand, there have not been corresponding advances in techniques for extracting knowledge from the data. It therefore comes as no surprise that data still present formidable challenges to effective and efficient mining of knowledge.

Since the human-visual system enables both recognition and understanding of overwhelming data at an instant [4], it is an outstanding resource for detecting and extracting knowledge from data. Tapping into the human-visual system would primarily entail exploiting relevant and effective visual strategies within the user interface. Most mining efforts have employed such visual strategies only at the beginning and at the end of the discovery process [12]. Human involvement in the entire mining process is crucial. Toward that, a human-user architectural component should be designed and positioned at a strategic place in an open overall discovery framework. Such an approach constitutes a

great step toward according the user a central place in the entire discovery process since the aforementioned human user's outstanding visual system becomes much more available for exploitation across all the phases of the discovery process.

This research, which is part of the project D2I (Data to Information http://www.dis.uniroma1.it/~lembo/D2I), focuses on the investigation and exploitation of strategies that are instrumental toward the realization of a visual interaction environment that supports the human user throughout the entire process of mining knowledge. Our research findings culminated in the realization of *VidaMine* (VIsual DAta MINing Environment) [1, 5-10], a visual data mining system that exploits various visual strategies thereby offering a visual interface that allows or enables the user not only to process data, but also to steer, guide or direct the entire process of data mining.

It should be acknowledged that there have been various research efforts in the field. Nonetheless, most of the efforts have paid little attention, if any, to usability evaluation. It is the user that can really or ultimately determine the worth or the usefulness of a particular mining system/system results in a specified environment or domain. When it comes to a visual data mining system worth its salt, there is no way round usability evaluation. There simply is no substitute for usability evaluation. Ensuring that a VDM system involves the user in the entire mining process is certainly very significant; most of the VDM research efforts come up short even in this respect. Notwithstanding, ensuring that the user is involved in the design process of the system life cycle needs prior consideration. This would shield the VDM system early in the system life cycle from exhibiting drawbacks later and especially after the system becomes operational. The input and recommendations of the users-to-be should be collected. Such information is instrumental in guiding the interface design and system development process.

The rest of the paper is organized as follows: Section 2 gives an overview of the VDM system. Section 3 describes the usability evaluation techniques that were employed, the obtained evaluation results and the corresponding analyses. Future work and a conclusion of the paper are given in Section 4.

2 The Proposed Visual Data Mining System (*VidaMine*)

The proposed system [1, 5-10] is based on an open knowledge discovery framework and aims at providing the user with a consistent, uniform, flexible and intuitive visual interaction environment.

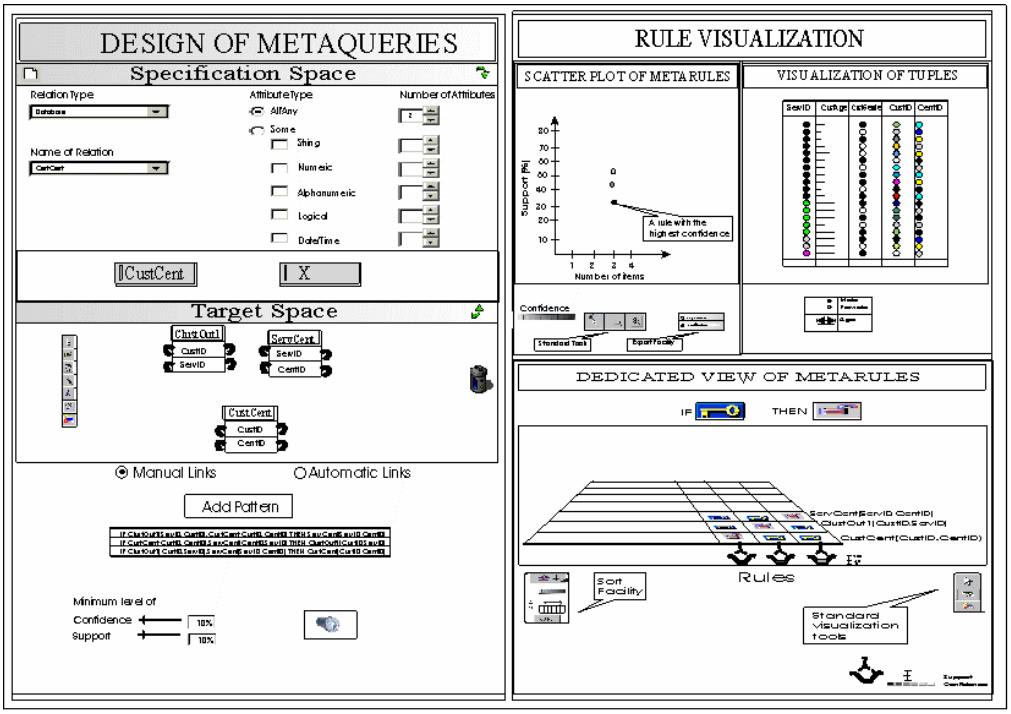

Figure 1: Metaquery Environment

At present, the system, *VidaMine*, supports, but is not limited to the following data mining (DM) strategies: metaqueries, association rules, and clustering. The architecture of the system comprises two main layers: *user layer* and *abstract DM engine layer*. The architecture supports the incorporation of new components and the modification of pre-existing components. In the architecture, the result of some DM task can be used as input to another DM task. The proposed architectural arrangement is pivotal toward enabling and empowering the user not only to process data (and knowledge), but also to drive, guide and control the entire discovery process.

VidaMine exploits various visual strategies to support the user in carrying out activities prior to the selection and application of DM algorithms (data preparation) e.g. visual construction of the target dataset and visualization of raw data. As regards target dataset construction, the user directly interacts with the actual data and other relevant information (e.g. metadata or previous mining results). The target dataset may comprise one or more relations. The system uses an environment comprising two intuitive interaction spaces: 'Specification Space' and 'Target Space'. The 'Specification Space' provides mechanisms, tools and resources necessary for visually building the target dataset. The 'Target Space' holds or hosts the relations that are part of the target dataset. As for visualizing raw data, *VidaMine* exploits a visualization subcomponent of the architectural user component. Through the subcomponent, the system can build a visualization of the data.

VidaMine supports visual construction of the mining query. For this, the user directly interacts with the actual data, other relevant information and other parameters (e.g. thresholds). The metaquery environment offers a novel visual approach through which the user can intuitively link attributes in order to define patterns (actually metarules) of interest. Figure 1 shows the metaquery environment. The interface also provides visual “baskets” for the construction of association rules.

VidaMine supports visual presentation of and interaction with mining results. With regard to clustering, some of the mining result aspects whose meaning should be made apparent include: density, homogeneity and separation measures. Regarding rules (metarules and association rules), aspects of interest include: measures of interestingness (e.g. confidence and support) and the items participating in a particular rule. The system offers various novel visualizations for the output of mining tasks. *VidaMine* can also exploit existing visualizations that are known to be effective by plugging in the corresponding tools/systems through the visualization subcomponent.

Since the user's information quest is not static, the interface design addresses the issue e.g. by ‘advertising’ relevant prior knowledge and providing interaction widgets that enable the users to adjust their mining approach.

3 Usability Evaluation

The ISO 9241-11 Ergonomic requirements for office work with visual display terminals: Guidance on usability (1998), defines usability as ``the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use.” In fact, the evaluation of goal quality includes the use of effectiveness, efficacy and satisfaction indicators, and security [11]. The importance of usability studies cannot be overemphasized.

3.1 Heuristic Evaluation

As a way of getting started, we carried out a heuristic evaluation [2, 3]. A detailed analysis of how heuristic principles have been applied in our interface design can be found in [6]. It should be mentioned that heuristic evaluation would need to always be referred to (or re-applied) throughout the life cycle or course of the project.

3.2 Mock-up and Walkthrough

Moreover, we initially designed a mock-up interface. We presented it before the whole team of DM experts participating in the project and carried out a simulation of their typed tasks.

We got encouraging results from the tests and even suggestions on how to improve the interface. For instance, the data mining experts suggested that the interface should allow the user to specify more generic relations when constructing the set of target data. They also suggested that the interface should offer the possibility to put constraints on such a generic relation (e.g. by specifying the number of its attributes). Consequently, the foregoing suggestions have been included by introducing some ``Specification Space" and ``Target Space" in the current design of the interface.

3.3 User Tests

We later carried out another evaluation that entailed performing user tests on the current version of the prototype with five selected users from the universities of Bologna (http://www-db.deis.unibo.it), Ferrara (http://www.unife.it), and Modena and Reggio Emilia (http://www.unimo.it).

One of the tested users is a multi-media post-doctorate student with some basic knowledge in data mining. Another user is an image analysis graduate with very little knowledge in data mining. Two of the users are data mining researchers. The last user is a post-doctorate student in pattern recognition with some little knowledge in data mining.

The experiment involved: the prototype as an application, a case study, user tasks corresponding to the case study, data schema corresponding to the data used in the case study, and a questionnaire. Each user was expected to run and interact with the interface of the prototype with reference to the accompanying documents (case study, user tasks, and data schema). After the experiment, the user would fill out the questionnaire.

The first part of the questionnaire had closed questions pertaining to the simplicity or complexity of carrying out user tasks. The respective results are seen in Figure 2, in which for each level of simplicity/complexity, the tasks are analyzed. It can be deduced that at least 60% of the users found each of the supported tasks fairly easy to carry out. In particular, all the users found the clustering output very easy to understand and interact with. 80% of the users found specifying target dataset and clustering input very easy to perform. 40% of the users found it very easy to construct metaqueries and association rules. 60% of the users found it fairly easy to understand and interact with the metarule and association rule visualization(s). 60% of the users also found it fairly easy to perform a mining task in general. As for using the output of a particular task as the input to another task, 40% of the users found it fairly easy to perform.

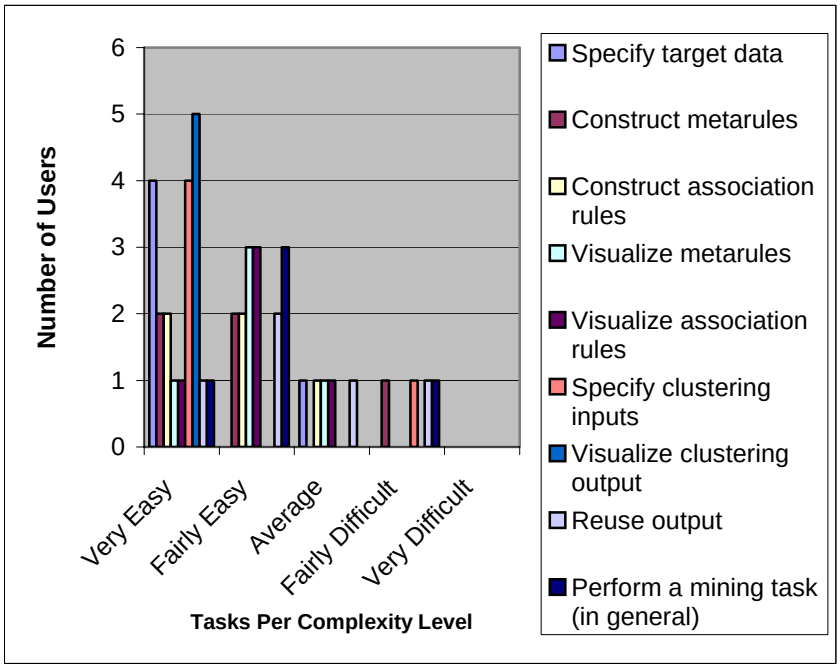

Figure 2: Simplicity/Complexity of User Tasks

The second part of the questionnaire also had closed questions. The part aimed at assessing interface design aspects. It should be mentioned that this being a simulation experiment, the subjects were not in a position to assess some of the aspects e.g. progress information, help and documentation. At least 60% of the users found each of the tested design aspects reasonably well-adhered to. 60% of the users observed that consistency was very well-applied in the interface design. All the users found the interface perceptiveness fairly well-adhered to. 80% of the users felt that the user's language was fairly well-applied. 80% of the users noted that it was fairly easy to remember (or acquire) the aspects relevant to a particular mining task. Moreover, the same percentage observed that the system responded to user operations in a reasonably valuable way. 60% of the users found the interface elements fairly well organized.

The last part of the questionnaire had open questions pertaining to strengths, weaknesses, and capability of the system/interface. It also had room for extra/other comments. It is interesting to realize that for the most-liked interface features, some of the subjects highlighted the same features. 40% of the users mentioned consistency as one of the main features they liked most about the interface. 40% of the users mentioned good layout/organization. 40% of the users mentioned the support for visual exploration as one of the main features they liked most about the interface. Based on the feedback obtained from the (closed and) open questions, there is an indication that the interface is reasonably

strong in consistency, layout/organization and visual exploration. On the other hand, some of the users stated some aspects that they considered negative. 40% of the users found some of the visual elements unnecessarily small/big. The foregoing issues will be carefully considered in future prototype versions. As for the supported system functionalities, 60% of the users were satisfied. 40% of the users called for an explicit and visual way of expressing joins while constructing the target dataset.

4 Future Work and Conclusion

At the moment, the current version of the prototype is being modified/improved in line with the results that were obtained from the walkthrough and questionnaire. After that, we intend to perform controlled experiments on the revised version. Consequent results will be instrumental in determining further relevant interface improvements and modifications.

In this paper, the pivotal role of usability studies in the interface design process of the system life cycle has been emphasized. Various usability methods and results have been presented coupled with a discussion on how they refined the design of the VDM system.

References

1. F. Angiulli, T. Catarci, P. Ciaccia, G. Ianni, S. Kimani, S. Lodi, M. Patella, G. Santucci and C. Sartori: "An Integrated Data Mining and Data Presentation Tool". Proceedings of the International Conference on Data Mining Methods and Databases for Engineering, Finance and Other Fields, 2002.

2. J. Nielsen. “Heuristic Evaluation”, In Usability Inspection Methods. J. Nielsen and R.L. Mack, Ed. John Wiley and Sons, 1994.

3. J. Nielsen and R. Molich. “Heuristic Evaluation of User Interfaces”, In Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, ACM Press, 1990.

4. S. K. Card, J. D. Mackinlay and B. Shneiderman. "Readings in Information Visualization-Using Vision to Think". Morgan Kaufmann Publishers,1999.

5. S. Kimani: "An Effective Visual Data Mining Environment". Doctoral Posters of the International Conference on Very Large Data Bases (VLDB), 2002.

6. S. Kimani, S. Lodi, T. Catarci, G. Santucci and C. Sartori: "VidaMine:A Visual Data Mining Environment". Journal of Visual Languages and Computing 15 (1):37-67, Elsevier, 2004.

7. S. Kimani, S. Lodi, T. Catarci, G. Santucci and C. Sartori: "Visual Data Mining with VidaMine". Proceedings of the Italian Symposium on Advanced Database Systems (SEBD), 2003.

8. S. Kimani, T. Catarci and G. Santucci: "A Visual Data Mining Environment: Metaqueries and Association Rules". Proceedings of the International Conference in Advanced Visual Interfaces (AVI), 2002.

9. S. Kimani, T. Catarci and G. Santucci: "A Visual Data Mining Environment". Proceedings of the CODATA Workshop on Information Visualization Presentation and Design, 2002.

10. S. Kimani, T. Catarci and G. Santucci: "A Visual Data Mining Environment". Proceedings of the ECML/PKDD Workshop on Visual Data Mining, 2002.

11. T. Catarci. “What Happened When Database Researchers Met Usability”. Information Systems, 25 (3), 177-212, 2000.

12. U. Fayyad, G. G. Grinstein and A. Wierse. "Information Visualization in Data Mining and Knowledge Discovery". Morgan Kaufmann Publishers, 2002.