User Interface Adaptation through Data Mining

Mária BIELIKOVÁ and Martin PILKA

Department of Computer Science and Engineering
Slovak University of Technology
Ilkovičova 3, 812 19 Bratislava
bielik@elf.stuba.sk, pilkam@decef.elf.stuba.sk

Abstract. In this paper we focus on adapting menu-oriented user interface using data mining. The adaptation model is created using the process of knowledge discovery. In the process of designing a method of adaptive presentation an application area should be considered. We experimented with the domain of medical examination cases. We combine the knowledge and experience of experts with machine learning in order to present information most relevant to the particular situation. We describe a software tool called EMMA dedicated to the collection and management of electromyography examination cases. The EMMA tool features an adaptive user interface, which helps a physician to make a decision about examinations performed.

Keywords: adaptive user interface, menu-driven user interface, electromyography examination cases, decision tree.

1 Introduction

A typical interactive application provides the same interface to all users regardless of user knowledge and previous work with the application. Some applications enable different views on the system. They offer users a way to initialize the user model through, for example, "Options" menus. Several systems allow personalized views based on user-selected stereotypes like "beginner" or "expert". Such systems are called *adaptable*. This kind of personalizing or adapting the user interface requires knowledge about the user (user model) and his/her manual setting.

An *adaptive* system is a system in which the adaptation is done automatically instead of being done by the user. Adaptive systems are valuable especially when the system's user interface is complex. It is useful also in situations where the user has to make a decision regarding a next action. A lot of research in the last years has been done in the area of adaptive hypermedia [3]. Hypermedia can be considered as a system offering information organized in nodes related through links. Because readers often have to comprehend a large hyperspace, the adaptation of hypermedia can help with this problem.

The methods and techniques for adaptation of hypermedia and "classic", often menu-oriented applications are based on the same principles. Certainly there are differences regarding the application (or hypermedia) space, methods of moving (selecting) different actions (or information to be read).

M. Bieliková (Ed.): DATAKON 2001, Brno, 20.-23. 10. 2001, pp. 310-318.

In this paper we focus on menu oriented user interface adaptation by means of data mining. We experimented with a proposed approach in the domain of medical examination cases collection and management. We combine the knowledge and experience of experts with machine learning in order to present information most relevant to the particular situation.

The outline of the rest of the paper is as follows. First, in Section 2, we analyze possibilities of the menu oriented interface adaptation. We were inspired by promising results achieved on the adaptive hypermedia [3], [4], [5]. Next, in Sections 3 we describe selected application domain – computer support of electromyography (EMG) examination cases collection and management. In Section 4 our approach to the realization of adaptation by using results of a data mining process is discussed. We finish the paper with our conclusions and give some proposals for future work.

2 Techniques for Menu-Driven User Interface Adaptation

Menu selection is one of the basic interaction styles used in many applications. In menu-selection systems, users read a list of items, select the one most appropriate to their task, and observe the effect.

The adaptation techniques used in this kind of systems should be different than those used for hypermedia. The main difference lays in the fact that hypermedia have rich content (e.g., educational material). Menu interface content on the other hand is briefer. It displays possibilities of working within the system for the user. The user wishes to find a particular action, or decide during complex form filling easily. Adaptation can be performed on two levels:

- content of menu items is adapted, i.e. the users are confronted only with items, which are relevant to their state (the user model);
- selections inside the forms contain adapted elements, i.e. those which are useful for the users.

Adaptive presentation uses some data about the user (e.g., his/her knowledge about a particular problem, previous work with the system) or about the context of the presentation (e.g., time) to present the content, which is the most interesting for the user.

In order to ensure that the content of the user interface provides a user with appropriate information, the adaptive system conditionally shows, hides, highlights or dims fragments of the interface when presenting it. This is done in order to ensure that the current screen of the application contains the appropriate information for the given user at the given time.

The adaptive content presentation enables the information content of elements to adapt to the user's expectations or requirements. Adaptive systems take into account the fact that users differ in knowledge, cognitive skills and reasons for working with the system. By keeping a model of some aspects of user characteristics and requirements, the system can adapt to and aid the user to work effectively with the system.

Two basic approaches to user related data gathering are following:

1. While a user is working with an application all his/her selections are registered. Based on these observations the adaptive system maintains a model of the user's knowledge about the application.
2. The adaptation is based on results of the previous work with the application, i.e. data entered into the application are considered.

Several techniques recognized as useful for adaptive hypermedia [3] can be used also in the context of menu oriented user interface adaptation. Namely:

- *conditional inclusion of items*: the adaptive system determines which items should be displayed. This technique is equivalent with menu items removal (non-relevant items are removed);
- *stretchtext*: the adaptive system determines which items and groups of items should be stretched (i.e., shown) and which items should be shrunk (i.e., hidden in the hierarchy of menu items). This decision determines the initial presentation of the menu. The user can stretch or shrink menu items through mouse clicks depending on his/her requests;
- *graytext*: this technique is often used in order to disable some user selections which are inappropriate in the current state of the application (e.g., if the MS Word is in the state 'no table selected', all menu items related to a table modification are grayed, only "Draw Table", "Insert Table", "Sort", "Formula" and "Show Gridlines" are black and clickable, the rest of items such as "Delete Cells", "Merge Cells" are visible, but gray and nonclickable);
- *explanation variants*: the same information (menu items) is presented in different ways. Providing explanation variants seems to be irrelevant to the menu interface adaptation in most cases because menu items should be brief and they represent concise names for particular commands. There is no sense in describing a command in different wording for different users. However, language variants and techniques for accommodating interface in different languages are developed by means of this technique;
- *reordering information*: the order of menu items and values for selection can be changed. Different orders can indicate different recommendations for the selection. Items are typically sorted from the most to the least relevant (this technique is known from information retrieval systems). Typical bases for sequencing items include [12]: time (chronological ordering), numeric ordering (ascending or descending), physical properties (volume, temperature, weight, velocity, etc.), alphabetic sequence of terms, grouping of related items (with blank lines or other demarcation between groups), most frequently used items first, most important items first (importance may be difficult to determine and may vary among users).

We adopted the last technique and used it for adaptation of the user interface for the EMG examination cases collection tool. The aim of our work is to assist a physician during the EMG examination in the decision process of the test being performed.

Our approach to the user interface adaptation is based on the assumption that users working with a system insert their "knowledge" into the system. Then the whole community can benefit from their knowledge, when we can comprehend them. This comprehension can be done through user interface adaptation. Novice users will benefit from the working habits of experts. In this case we have only one user model, which is changed in the course of application use by the experts.

A different approach is adapting the user interface according to the user model. In such a case the application is adapted for each user (or group of users) in order to achieve an improvement in their performance when working with the application. A combination of both mentioned approaches is also possible.

3 EMG Examination and Case Collection Tool

Electromyography (EMG) is a standard method for monitoring neuromuscular activity at the level of bioelectrical signals. The EMG diagnosis is based on a systematic acquisition of numeric and symbolic data. The domain of EMG is broad, covering more than a hundred existing diagnoses, exploiting knowledge belonging to various kinds of data, including about four thousand tests of nerve or muscle anatomical structures [15].

The relative complexity of EMG examination raises possible problems for physicians. Computer based information processing supporting diagnostic process in neurophysiology together with an adaptive user interface may enhance the physician's ability to make appropriate decisions and find a correct diagnosis.

The specificity of the EMG diagnosis process lies in the grounded evocation of diagnosis hypotheses and in the grounded selection of tests, depending on a collection of environmental and patient-dependent factors [7]. The tests to be performed are situated in a multi-facetted context involving the diagnostic hypothesis at hand, as well as various patient-dependent and device-dependent conditions. The process is furthermore prescriptive, in the sense that further diagnosis hypotheses are formulated on the basis of the current information available.

The EMG examination process can be described as follows. First of all, the practitioner consults the medical records about the patient and starts examining and questioning the patient. At the end of this stage, the different clinical signs are stated. Practitioner is allowed to consider one or several hypotheses about a diagnose. Then the hypothesis evoked most often is chosen. Next, a test procedure is associated with it from which several tests are selected. Each test is performed according to two selections: an anatomical structure and an examination technique. The results of these tests make it possible to confirm or reject the initial hypothesis (so-called referral diagnosis).

We have developed a software tool for EMG examination cases collection and management called EMMA [2]. This software tool serves as a basis for exchange of expertise between physicians. The EMMA tool is developed in the Visual C language in the Windows operating system. The user interface is menu-driven. It is comprised of several windows and a set of forms for filling and maintenance the EMG case data (see Figure 1).

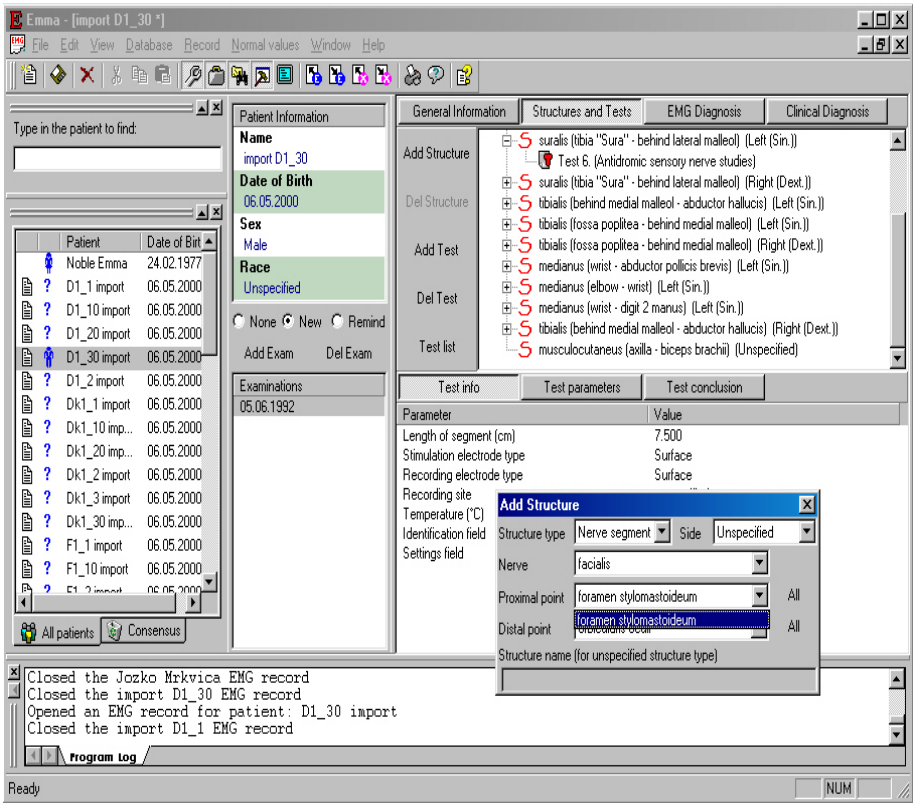

Figure 1. Example of EMMA user interface.

We use several data formats during the EMG examination cases collection:

- binary format: ESTEEM Communication Protocol EPC/ECCO 3.2 [8],
- relational format [1],
- exchange XML format [1].

ECCO binary format is designed as an efficient format from the point of storage view (developed during the ESTEEM project [9]). Several existing tools use it and it is retained in the EMMA in the form of export/import capabilities. The relational format is advantageous because of its flexibility, extensibility and portability. It is used for storing EMG examination cases. It also forms a basis for data mining. Finally, the XML format is suitable for distribution of EMG cases between physicians (the users of the system). It is important in the process of consensus exercises. Moreover, the XML format is used for communication with the EMG machine (automatic data gathering).

The EMG case base contains data structures covering all anatomical structures, examination techniques, test conditions and parameters used during the EMG examination, together with data structures important during the diagnostic process and

interpretation of the examination data: symbolic parameter values, pathophysiological test conclusions, pathophysiological structure conclusions, EMG diagnoses and clinical diagnoses.

4 Data Mining for the Case Collection Tool Adaptation

The system discovers knowledge by observing the expert's selections of anatomical structures examined together with examination techniques used. The adaptation model follows discovered knowledge. It is represented by a decision tree. The decision tree is created in the data mining step, where modification of the ID3 algorithm [11] is used.

Without an indication of what we are looking for, it is difficult to infer an adaptation model from the data. We had available more than 1 000 EMG examination cases represented in the ECCO format (collected during the ESTEEM project). Our first steps were oriented toward data preparation. We selected only cases without error (some 5% of examinations were discarded). Next, we converted the EMG data into the relational database proposed for the EMMA tool. Relational database helped in effective experimentation with data. In fact erroneous examinations were found during the process of data converting from binary to the relational format.

In order to determine *condition* and *decision* attributes we did several statistics over the EMG database [10]. Results of this statistical analysis are important. We found out which data in the database are incomplete or missing. The problem with the current state of the EMG data is that there is still a large amount of non-structured information, which is entered into the examination cases collection tool as free text.

Considered *condition attributes* were: date of birth, sex, weight, height, acquiring department information, and referral diagnosis. *Decision attributes* were anatomical structure and examination technique. After statistical analysis of the EMG database we constrained the condition attributes set: attributes weight and height were removed. Moreover we replaced attribute referral diagnosis by the attribute EMG diagnosis.

The process can be broken into two basic phases [6]:

1. *learning phase*, where the system analyses a database and recognizes similarities among data entities. The result of this analysis is a decision tree, which determines the order of decision attributes according to values of condition attributes;
2. *using phase*, where the created decision tree is used in the EMG examinations.

At using time of the EMMA tool, new EMG examination cases are collected. New cases affect the decision tree, which is used for examination process support. We had two possibilities: using off-line (non-incremental) learning or using on-line (incremental) learning. In non-incremental learning all of the training examples are presented simultaneously to the algorithm. In contrast, using incremental learning the examples are presented one at a time.

We decided to use a combination of these approaches. Basically, off-line learning is used. However, learning is provided in an incremental basis. This means that the

decision tree is updated periodically based on new added examinations (after the last decision tree update). However, at the moment our learning algorithm needs all cases for decision tree building (also the previously considered ones). The decision tree is not updated, it is always created from scratch. For the selected application area of the EMG examinations this approach is sufficient enough because the number of examination cases is small. (At the moment we have a database of about 1 000 cases.) This database was built during several years, so we do not suppose extensive increase of database size. In order to extend our algorithm to incremental learning the approach of [13] could be used.

We had to solve a problem of inconsistent data, i.e. different physicians decide in the same situation (given by selected decision attributes) differently. In this case our learning algorithm weighted both solutions in the same way.

Leaves of the decision tree are not simple decisions in the form of a value of an attribute. We have two decision attributes: anatomical structure and examination technique. In fact, the decision comprises several anatomical structures with one or more examination techniques for each. Many such decisions (much more than the number of actual examination cases) can be constructed. In order to decrease the number of possible attribute combinations, we defined an equivalence relation on the decisions. It is based on comparing specified number of performed tests and occurrence of a particular anatomical structure and examination technique.

Decisions (classes) are computed using cluster analysis technique *before* the data mining process starts. *After* data mining, the anatomical structures and examination techniques in the decision are ordered according to their occurrence in particular cases related to this decision in the decision tree.

Figure 2 illustrates a simple example of the decision tree. For simplicity we consider just one condition attribute, *age*. In fact such a decision tree is useless, because the attribute with the highest entropy is the attribute *EMG diagnosis*, which should be always considered in data mining.

5 Conclusions

In this paper we discuss different approaches to user interface adaptation. We proposed approach to adaptation of the user interface through results of a data mining process. Knowledge about the decisions (menu selections) is discovered via rule induction.

We experimented with a proposed approach in the field of electromyography examinations using more than 1 000 EMG examination cases and expert consultations from physicians. The result of our work is the EMG examination cases collection tool with an adaptive user interface. The system currently uses a database of examination cases and acts as a structure and technique selection support in addition to a static user interface. It uses an adapted ID3 algorithm for data mining. Adaptation is concerned with specific nature of a decision attribute: a decision consists of a set of anatomical structures each related to a set of examination techniques. A decision serves as a recommendation for a physician according to the selection of test parameters. Resulting anatomical structures and examination techniques are also ordered according to their occurrence in all relevant examination cases from a decision tree.

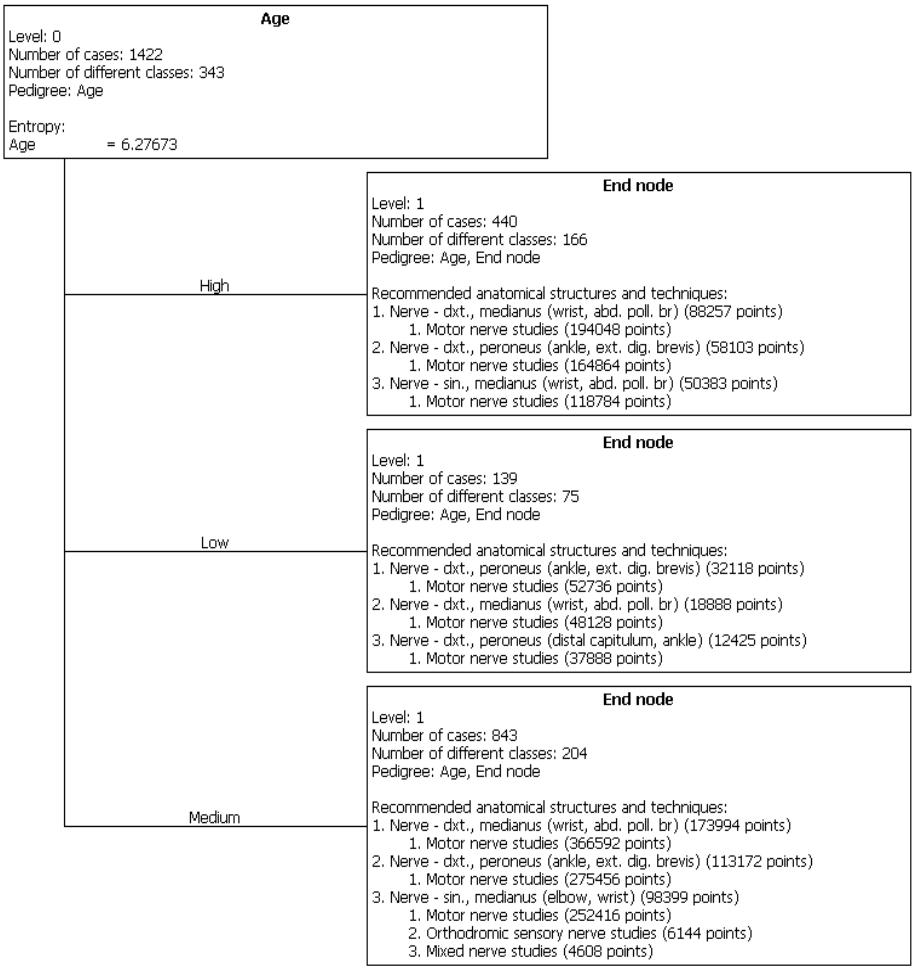

Figure 2. An example of a decision tree.

The problem with the current state of adaptation is still a large amount of non-structured information, which is entered into the database as a free text. The structured clinical information from the MYOSYS [14] can be used to enhance the possibility of data mining over created databases.

Acknowledgement

The work reported here was partially supported by project INCO Copernicus, No. 977069, EMG-Net and by Slovak Science Grant Agency, grant No. G1/7611/20.

References

1. Bieliková, M., Návrat, P. and Smolárová, M.: Reengineering of Legacy Data and Knowledge: Designing Data Representation for Electromyography Studies. In: *Proc. of ADBIS'2001*. (to appear).
2. Bieliková, M., Návrat, P. and Smolárová, M.: Collection and Management of EMG Examination Cases. *ERCIM News*. No. 45 (April 2001), 43-44.
3. De Bra, P., Brusilovsky, P. and Houben, G.J.: Adaptive Hypermedia: From Systems to Framework. *ACM Computing Surveys* **31**, 4es (December 1999) 1-6.
4. Brusilovsky, P. and Schwarz, E.: User as Student: Towards an Adaptive Interface for Advanced Web-Based Applications. In: *User Modeling - UM97*, A. Jameson, C. Paris, C. Tasso (Eds.), Springer Wien New York (1997). Available on-line from http://um.org.
5. Brusilovsky, P.: Methods and Techniques of Adaptive Hypermedia. *User Modeling and User Adapted Interaction* **6**, 2-3 (1996) 87-129.
6. Cios, K., Pedrycz, W. and Swiniarski, R.: *Data Mining: Methods for Knowledge Discovery*. Kluwer Academic Publishers, 1998.
7. Garbay, C.: Data Mining in EMG: An Agent-Centered Perspective. In: *Proc. of Workshop on Application of Advanced Information Technologies to Medicine*, G. Agre and D. Ziébelin (Eds.), Varna (2000), 72-83.
8. Jakobsen, L.S., Fog, B. and Talbot, A.: *ESTEEM Communication Protocol*, ver. 3.2. ESTEEM report Advanced Informatics in Medicine (A2010), 87p., 1993.
9. Johnsen, B., Fuglsang-Frederiksen, A.: *USER-5: EMG Dataset Specification* version 3. ESTEEM report Advanced Informatics in Medicine (A2010), 44p., 1994.
10. Pilka, M.: *Preparing EMG CaseBase for Data Mining*. Final Bc. project, Department of Computer Science and Engineering, Slovak University of Technology in Bratislava, 2001 (in Slovak).
11. Quinlan, J.R.: Learning Efficient Classification Procedures and Their Application to Chess End Games. In: *Machine Learning: An Artificial Intelligence Approach*, R.S. Michalski et. al. (Eds.), Tioga, Palo Alto, Calif. (1983).
12. Schneiderman, B.: *Designing the User Interface: Strategies for Effective Human-Computer Interaction*. Addison-Wesley, Third Edition, 1998.
13. Utgoff, P.E.: Incremental Induction of Decision Trees. *Machine Learning*, No. 4 (1989),161-186.
14. Ziébelin, D. and Vila, A.: Medical Planning Environment. In: *Proc. of 6th Int. Conf. on AI in Medicine*, Grenoble, Springer, LNCS 1211 (March 1997), 343-348.
15. Ziébelin, D. and Villa, A.: Building Knowledge Base in Electromyography through the WWW. In: *Proc. of Workshop on Application of Advanced Information Technologies to Medicine*, G. Agre and D. Ziébelin (Eds.), Varna (2000), 117-124.