Improving Control in an Agent for User Modeling

Maria Virvou and Katerina Kabassi

Abstract-- This paper describes the User Modeling Agent (UM Agent) of an intelligent Graphical User Interface that manipulates files. The intelligent GUI is called IFM and it monitors users while they work; in case a user has made a mistake, it intervenes automatically and offers advice. The reasoning of the UM Agent is largely based on an adaptation of a cognitive theory, called Human Plausible Reasoning (HPR). The UM Agent observes users during their interaction with the system, maintains and manages the user profiles and provides relevant information whenever other agents request it. The main focus of this paper is on examining how the UM Agent’s control can be improved by the combination of Human Plausible Reasoning and stereotypical knowledge.

Index Terms--Human-Computer Interaction, Intelligent User Interfaces, Multi-agent systems, Intelligent assistance, User Modelling.

I. INTRODUCTION

Research in the area of Human-Computer Interaction is trying to meet the challenge of making user interfaces more usable. However, user interface researchers have put little emphasis on improving the underlying structures that represent the intelligence of these interfaces. Systems can be more usable and more useful when they provide users with experiences fitting their specific background knowledge and objectives [7]. Therefore, as Rist et al. [20] point out the next major step in the evolution of interfaces is very likely to focus on highly personalized interfaces.

In this paper we describe an intelligent Graphical User Interface called Intelligent File Manipulator (IFM) that manipulates files, such as the Windows 98 Explorer [14]. The objective of IFM is to allow a user direct manipulation and, at the same time, to protect him/her from his/her own errors by providing intelligent help. Generally, IFM tries to act as a human expert who watches the user over the shoulder, reasons about his/her actions and provides spontaneous advice in case of an error. However, creating a system that intervenes appropriately requires inferring the user’s intentions. Diagnosis should be conducted as non-invasively as possible because continually interrupting users to determine their current intentions and to ferret out their misconceptions would be rather annoying. Therefore, IFM has assigned agents to observe the user while s/he is actively engaged in their usual activities and provide spontaneous advice in case of an error.

A multi-agent system is composed of a group of agents that are autonomous or semiautonomous and which interact or work together, to perform some tasks or achieve some goals [11]. The design of individual agents within a multi-agent system has the advantage of being independent of the design of other agents. This significantly contributes to the breakdown of complexity [6].

It is widely agreed that basing decisions on an accurate cognitive model of the user is important for effective prediction of user intent [17]. Therefore, we have adapted and implemented a part of a cognitive theory called Human Plausible Reasoning theory [3] (henceforth referred to as HPR) and its certainty parameters in order to make inferences about possible users’ errors based on evidence from users’ interaction with the system. HPR has been used in IFM to simulate the users’ reasoning, which may be correct or incorrect (but still plausible) and thus may lead to “plausible” user errors. This is a similar approach to another intelligent help system for users of the UNIX operating system, called RESCUER [22], [23]. The adaptation and incorporation of HPR into an intelligent GUI is quite important since it aims at rendering the interaction with the user more human-like.

The main focus of this paper is on the UM Agent of the system. The UM Agent observes users during their interaction with the system, maintains and manages the user profiles and provides relevant information whenever other agents request it. Although agents are successful at being able to learn their user’s behavior and assist them, a major drawback of these systems is the fact that they require a sufficient amount of time before they can be of any use [10]. Indeed, in previous versions of IFM [25] one problem that came up was that no sufficient information about the user could be obtained before s/he had interacted with the system for quite a long period of time. A solution to this problem may be the incorporation of stereotypes [18], [19] in the system to provide default assumptions about users until the user model acquired sufficient information about each individual user. As Kay [9] points out, stereotypes constitute a powerful mechanism for building user models that may be used before sufficient information about each individual user is acquired.

However, a system that intends to provide assistance to users’ errors needs to perform error diagnosis. Therefore, the UM Agent has to select the most appropriate hypothesis about the cause of an error among many possible hypotheses. To
improve control in the UM Agent, we have taken an innovative approach of combining stereotypes with the certainty parameters of HPR.

II. RELATED WORK

. Human Plausible Reasoning

HPR is based on an analysis of people’s answers to everyday questions about the world and tries to formalize plausible inferences that occur in people’s responses to different questions [3]. The theory consists of a formal representation of plausible inference patterns that are frequently employed in answering everyday questions, a set of parameters that affect the certainty of people’s answers to such questions and a system relating the different plausible inference patterns and the different certainty parameters. An example of a plausible inference is presented below. Let’s suppose that the question asked was whether coffee is grown in Llanos region in Colombia. The answer would depend on the knowledge retrieved from memory. If the subject knew that Llanos was in a savanna region similar to that where coffee grows, this would trigger an inductive, analogical inference, and generate the answer yes.

According to the theory a large part of human knowledge is represented in “dynamic hierarchies”. These hierarchies are used to model the reasoning of people with patchy knowledge. There are four kinds of relation between objects in hierarchies: generalization (GEN), specialization (SPEC), similarity (SIM) and dissimilarity (DIS). Statement transforms are the simplest class of inference patterns.

However, the theory also introduces certainty parameters. These certainty parameters can affect the statement transforms. The degree of similarity ($\sigma$) represents the similarity of one set to another one. The degree of typicality ($\tau$) represents how typical a subset is within a set (for example, the cow is a typical mammal). Dominance ($\delta$) indicates how dominant a subset is in a set (for example, elephants are not a large percentage of mammals). Finally the only certainty parameter applicable to any expression is the degree of certainty ($\gamma$) or belief that an expression is true.

. Agents for User Modeling

Most of the agent-based systems that perform user modeling, attempt to model the user in order to provide more individualized interaction. Agent-based systems such as I-HELP [2], [21], Baghera [26], Fab [1], Amalthea [15], and MyEnglishTeacher [4] attempt to model some of the users’ characteristics.

I-HELP [2], [21] is an example of a system with similar functionality to IFM. I-HELP models workers so that it can assist one worker in identifying a peer who can help them. However, the problem with this approach is that there is not always someone available to help. Individual support is also provided by Baghera [26] but in a quite different domain, in problem solving. However, this approach, similarly to I-HELP, instead of really helping the user it tries to choose the appropriate problem solver to send the student’s solutions.

Emphasis on user modeling is given in MyEnglishTeacher [4], which supports users during their learning process. MyEnglishTeacher is a multi-agent, long-distance teaching environment for academic English. The main problem with this approach is that the system uses two agents for gathering the information about the user. The Global Agent (GlA) stores general features on students where as the Personal Agent (PA) stores the private features of each student. The division of information in two different agents may delay the reasoning process.

De Andrade et al. [5] propose a framework for the use of Information Technology in education based on Vygotsk’s socio-cultural theory. In this approach, what they call ZPD agents are responsible for observing the real development of the students. However, this approach, similarly with the ones mentioned above, has the problem that the agent requires a sufficient amount of time before it starts working in its full functionality. The UM Agent, in IFM, uses stereotypes in order to confront this problem.

III. OPERATION OF THE GUI

Intelligent File Manipulator (IFM) is an Intelligent Graphical User Interface that works in a similar way as a standard file manipulation program, such as Windows 98/NT Explorer but it also incorporates intelligence. IFM is meant to help users during their navigation and manipulation of the file store and provides advice in case this is considered necessary. In general, IFM tries to act as a human expert who watches the user over the shoulder and offers advice spontaneously.

IFM’s architecture consists of three agents, namely, the UM Agent, the Advising Agent and the Speech-driven Agent. Agents could be viewed as black boxes whose operations are abstracted to the services they provide [6]. Therefore, we present the system’s operation with respect to what every agent contributes to the reasoning process of IFM.

The UM Agent is mainly responsible for observing the user while s/he is actively engaged in his/her usual activities and keeping his/her profile updated. So every time a user issues a command (e.g. selects an icon or a command from a menu), the UM Agent reasons about it in order to categorize it in one of four categories, namely “expected”, “neutral”, “suspect”, “erroneous”. A command is categorised as expected if it is compatible with the user’s hypothesized goals. On the other hand, a command that contradicts the UM agent’s hypotheses about the user’s goals is categorised as suspect. Erroneous are the commands that are wrong with respect to the user interface formalities. Finally, a command is considered neutral if it cannot be assigned to one of the former categories. If the action is categorised as expected or neutral, it is executed normally. However, if the action is categorised as suspect or erroneous then the UM Agent informs the Advising Agent that the user is probably involved in a problematic situation.

Before starting the generation of advice the Advising Agent requests information about the user from the UM Agent. Meanwhile the Advising Agent is involved in the transformation of the given action, which is based on HPR, so that similar alternatives can be found. After receiving the
as Mitrovic et al. [13] point out there may be different explanations of observed incorrect user’s actions. Therefore, there is a need to attach priorities to different explanations so
In the previous section, based on the results of the empirical study that was described, it was found that some may be preferred over others. For example, expected actions have priority over neutral ones. However, even this process may result in the generation of many alternative commands and this may confuse the user rather than really helping him/her. The Advising Agent uses the certainty parameters of HPR to determine the priority among actions belonging to the same category. The values of the certainty parameters are provided by the UM agent. Certainty parameters of HPR have been adapted to fit IFM’s requirements. The certainty parameters of HPR used in IFM are the following:

- Degree of certainty ($\gamma$)
- Degree of typicality ($\tau$) of an action in the set of all actions issued by the student.
- Degree of similarity ($\sigma$) of a set to another set.
- Frequency ($\varphi$) of an error in the set of all actions
- Dominance ($\delta$) of an error in the set of all errors.

In IFM, the degree of similarity ($\sigma$) is used to calculate the resemblance of two commands or two objects. It is mainly used to show possible confusions that a user may have made between commands. The typicality ($\tau$) of a command represents the estimated frequency of execution of the command by the particular student. The degree of frequency ($\varphi$) of an error represents how often a specific error is made by a particular student or the students of a particular stereotype. The particular student’s weaknesses can be recognized by the dominance ($\delta$) of an error in the set of all errors. Finally, all the parameters presented above are combined to calculate a degree of certainty related to every alternative command generated by IFM. This degree of certainty ($\gamma$) represents the system's certainty that the student intended the alternative command generated.

The degree of certainty is calculated as a sum of all certainty parameters, with each parameter being multiplied to a weight, which is determined with respect to how important the particular certainty parameter is. The calculation of the degree of certainty required specifying the weights that each certainty parameter should be multiplied with. The weights that were selected are shown in the formula (1).

$$\gamma = 0.4 \cdot \sigma + 0.3 \cdot \delta + 0.2 \cdot \varphi + 0.1 \cdot \tau \quad (1)$$

The weights of each certainty parameter were estimated based on the results of the empirical study that was described in the previous section.

VI. COMBINATION OF USER STEREOTYPES AND HPR

Stereotypes may serve as a tool to model the beliefs and preferences that the users of a system may have. The main reason for the application of stereotypes is that they provide a set of default assumptions, which can be very useful during hypotheses generation about the user. Generation of default assumptions can prove very effective for modeling a large set of users. However, this approach has many problems as well. For example, despite the similar behavior that users of the same group may have, every one is an individual that differentiates in many aspects from all the others. Therefore, stereotypes should be used for initializing the user model, until there is more information.

In IFM, users are classified into one of three major classes according to their level of expertise, namely, novice, intermediate and expert. Each one of these classes represents an increasing mastery in the use of the particular file manipulation system. Such a classification was considered crucial because it would enable the system to have a first view of the usual errors and misconceptions of a user, belonging to a group. For example, novice users are usually prone to command errors whereas expert users do not make mistakes in the command use.

One might not expect expert users to make mistakes but this does not correspond to reality. There are some experts that are very prone to accidental slips due to their carelessness. As a result, another classification that was considered important was dividing users into two groups, careless and careful.

A stereotype usually has a set of trigger events. Trigger events are boolean expressions that once one of them becomes true, the corresponding stereotype must be activated for a particular user. The UM Agent infers information about the user by watching him/her during his/her interaction with IFM; a user that makes a lot of errors is probably a novice whereas someone who only makes few errors that can be considered as accidental slips is probably an expert.

However, the UM Agent cannot decide where to categorize a user before s/he has executed a satisfactory number of commands. The empirical study revealed that an early conclusion could be drawn only after the execution of twenty commands. A novice user will have made a large number of errors after the execution of twenty commands, whereas an expert in the same amount of commands will have made only a few errors that can be regarded as accidental slips.

Triggers for categorizing the users according to their carelessness when executing certain tasks are constructed similarly to the triggers of the stereotypes that correspond to the user’s proficiency. Again a user is not categorised before having executed 20 commands and the system takes into account the number of accidental slips a user may have made.

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Dominance</th>
</tr>
</thead>
<tbody>
<tr>
<td>NOVICE</td>
<td>0.3</td>
<td>0.5</td>
</tr>
<tr>
<td>CommandError</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>StructureError</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>INTERMEDIATE</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CommandError</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>StructureError</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 1. Default assumptions of stereotypes relating to the level of expertise of a user.

However, a problem encountered when using stereotypes is that a user does not necessarily remain to a certain stereotype forever. Users’ skills and behavior change while they interact with the system. This is why the system must regularly check whether the right stereotype is activated or
not. After 50 commands at a time the UM Agent checks whether the activated stereotype is still appropriate. If the UM Agent has observed an improvement in skills or an attitude modification for a specific user, it revises its previous conclusions and deactivates the activated stereotype in order to activate the one that fits best the user in the present conditions.

The default assumptions for users of the novice, intermediate and expert stereotype are presented in Table 1. We can see that novice users are more prone to command errors, which are considered to be their weak point, rather than structure errors. Intermediate users are still committing command and structure errors, although these are not such usual errors for them. On the other hand, expert users do not make such errors.

Stereotypes that classify users according to their degree of carelessness include default assumptions about the errors made due to carelessness. This kind of stereotype contains information about the kind of accidental slips, a user may make and the frequency s/he makes such errors. For example, a user, who is considered by the system as careless, usually makes 30% mouse errors, 30% spelling errors and only a few errors are due to confusion between objects with identical names. In Table 2, one can see the default assumptions for careless and careful stereotype.

### Table 2. Default assumptions relating to the carelessness of a user.

<table>
<thead>
<tr>
<th>Stereotype</th>
<th>Frequency</th>
<th>Dominance</th>
</tr>
</thead>
<tbody>
<tr>
<td>CARELESS</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpellingErrors</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>MouseError</td>
<td>0.3</td>
<td>0.4</td>
</tr>
<tr>
<td>IdenticalNamesError</td>
<td>0.1</td>
<td>0.2</td>
</tr>
<tr>
<td>CAREFUL</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SpellingErrors</td>
<td>0.1</td>
<td>0.45</td>
</tr>
<tr>
<td>MouseError</td>
<td>0.1</td>
<td>0.45</td>
</tr>
<tr>
<td>IdenticalNamesError</td>
<td>0.025</td>
<td>0.1</td>
</tr>
</tbody>
</table>

VII. INDIVIDUAL HISTORY OF USERS

As mentioned in the introduction, stereotypes are used in IFM by the UM Agent only for capturing the initial impression of a user. A stereotype is activated after implicitly acquiring information by observing the user while interacting with the system.

In the beginning, information is acquired only by the stereotype. However, the corresponding agent is also constantly collecting information about a particular user’s behavior and errors and informs the individual user model of the user. As the UM Agent collects more and more evidence about a user, information is acquired in part by the stereotype and in part from the individual user model. The percentage of information acquired by the stereotype diminishes as the percentage of acquisition by the individual user model increases. In case a conflict appears, the UM Agent always favors information acquired from the individual user model.

VIII. EVALUATION OF THE UM AGENT

For the evaluation of the UM Agent 16 users were selected, 8 novice and 8 experts. They had diverse backgrounds and interests and constituted a representative sample of expert and novice users. All 16 users were asked to interact with IFM, as they would normally do with a standard file manipulation program. IFM worked in its full functionality and provided spontaneous advice in case of an error. However, the system had only limited information about each user since users had interacted with IFM for a short period of time. Therefore, the only available information about users’ errors, habits and tendencies could have been acquired by the corresponding stereotype. The experiment aimed at finding out how successful IFM was at the categorization of the users to stereotypes and how successful the default assumptions of the stereotypes were in the generation of advice. In order to do so we compared IFM’s reactions to those of the human experts.

Below, we illustrate an example of a user protocol. The user of this example had already been categorised by IFM as novice and careless. The categorization had been based on the user’s first interactions with the system and it had been in accordance with the opinion expressed by the majority of human experts. In the example, we show what IFM’s reactions were to the user’s actions and how these were compared to the reactions of the human experts that participated in the experiment and analyzed the protocols. The user’s initial file store state is illustrated in figure 1.

![Figure 1. The user’s initial file store state](image)

The user issued the following actions:
1. delete(A:\Windows1\Active.bak)
2. delete(A:\Windows1\Log.bak)
3. delete(A:\Windows\)

IFM had no problem with the first two actions and executed them normally. However, the system judged that there may have been a problem with the third action. This was due to the fact that the folder A:\Windows\ contained three subfolders with many useful files. Therefore, IFM generates the following alternative actions each of which is compatible to the user’s hypothesized intentions:

*Alternative action 1:*
delete(A:\Programs\)

*Alternative action 2:*
delete(A:\Windows1\)
the generation of alternative actions, the Advising Agent calculates the degree of certainty for each alternative action. As mentioned above, the only available information are the default assumptions of the corresponding stereotype, which are presented in Table 2.

The UM Agent taking into account those values and using the formula (1) calculates the degree of certainty for each alternative. Hence, the degree of certainty of the first alternative was 0.43 whereas the degree of certainty for the second alternative was 0.83. Therefore, the second alternative was proposed. Indeed, the comparison with the human experts revealed that 90% of the human experts proposed the second alternative action.

Concerning the comparison of IFM’s responses to human experts’ responses, the results were encouraging. In cases when there was a total agreement of human experts’ opinions, IFM produced either a very similar or exactly the same advice to that of the human experts. This usually corresponded to cases where the error was “obvious” to human advisors such as the error presented in the example, where there was 90% of agreement among the experts. However, there were cases where there was a diversity of human experts’ opinions. In those cases IFM’s advice was either identical to the advice of the majority of human experts or in fewer cases it was compatible to the advice provided by a minority of experts. In general, the degree of compatibility between the majority of human experts and IFM’s advice was 63%, which was quite satisfactory.

IX. CONCLUSIONS

In this paper, we have described a multi-agent system for the provision of intelligent assistance. The main focus of this paper has been on the UM Agent of the system. The UM Agent is responsible for observing the user, updating the user models and providing other agents with information about the user. Compared with the traditional architectures, having an agent for the user modeling is advantageous in terms of simplicity and flexibility in the design and implementation. However, previous versions of the system revealed that the UM Agent could not provide other agents with information about the user before having observed him/her for quite a long period of time. This had resulted in the generation of many competing hypotheses about the user’s beliefs.

Therefore, the classification of users into meaningful groups (stereotypes) enhanced the system’s performance. The stereotypes of users were combined with principles of a cognitive theory called Human Plausible Reasoning. This novel combination is quite important for the initialization of user modeling and the provision of assistance adapted to the needs of each user. In this way, the interaction is made more human-like in the sense that the system may provide to a large extent automatic assistance similar to a human advisor who watches a user over the shoulder.

REFERENCES