

# Goal-Based Autonomous Social Agents: Supporting Adaptation and Teaching in a Distributed Environment\*\*

Julita Vassileva

ARIES Laboratory, Department of Computer Science, University of Saskatchewan,  
1C101 Engineering Bldg, 57 Campus Drive, Saskatoon, S7N 5 A9 Canada  
Email: jiv@cs.usask.ca

**Abstract.** This paper proposes a theoretical framework that allows goal-based agents attached to networked applications and learning environments to support users' work and learning. Users, learners, applications and learning environments are represented by autonomous goal-based social agents who communicate, cooperate, and compete in a multi-system and multi-user distributed environment. This allows for a uniform approach to support the user while working, by adaptation to the user's goals, preferences, level of experience and available resources; as well as teaching the user using various teaching paradigms (constructivist or instructivist). In addition it allows one to take into account the user's /learner's motivation and affect; as well as enabling a coherent discussion of teaching strategies.

## 1 Trends in the Development of Adaptive and Teaching Systems

Two major trends can be observed in the development of learning environments, which follow from the rapid development of networking and communication technologies:

- An integration of working and learning environments.
- No virtual difference between humans and application agents

Nowadays nearly all commercial applications (most prominently CorelDraw, Toolbook etc.) are equipped with training programs, which provide an introduction into the main features and basic working techniques, as well as with on-line help, which is in some cases context-sensitive and even adaptive (MS-Office 97). This means that the user is working and learning at the same time, and can switch from "working" mode to "learning" mode at will. It is easy to switch from some type of teaching or demonstration, which will help the user learn about something specifically needed at the moment, and then switch back to the "working" mode to try and use the newly acquired knowledge in practice.

---

\* Published in Proceedings of ITS'98, San Antonio, Texas, pp.564-573

From the other side, learning environments specifically designed for educational purpose in some subject are often inspired by constructivist and Vygotskian theories of learning, which focus on context-anchored learning and instruction, that takes place in the context of solving a realistic problem. There is a tendency in learning environment design philosophies towards integrating work and learning; work being the source of problems and motivation for learning.

In general, one can observe a convergence between working environments and learning environments. For example, instead of adapting to sub-optimal learner behavior, the system may decide to *teach* the learner (instruct, explain, provide help, etc.) how to do things correctly, i.e. to make the user adapt to the system by learning. In reality every adaptation is bi-directional. Every participant in an interaction process adapts to the other participant(s). The system learns about the user and adapts to him/her; the user learns about the system and adapts accordingly. An adaptive system should support the user's learning about the system [8]. It has to be able to decide whether to adapt to the user or to teach something instead (i.e. to make the user adapt to the system). It must decide whether to be reactive or proactive. In this way the system will be an active participant in the interaction, an autonomous agent, which can decide in the course of interaction and not just follow embedded decisions made at design time (normative decisions).

An attempt to build such a system which can take decisions about whether to teach or to coach the student depending on the context of interaction and state of student model has been designed and implemented using reactive planning techniques [8]. However, we feel that such a pedagogically "competent" system has to be able to negotiate its decisions with the learner and not just impose them, since no matter what expertise is underlying these decisions, there is always uncertainty about the correctness of this knowledge and about the student model.

Therefore we decided to model the pedagogical component in an intelligent learning environment as an autonomous agent that pursues certain teaching goals. These goals can be cognitive (subject- and problem-specific), motivational, and affective (learner- and subject-specific). We call these agents "Application agents", since they are associated with an application which can be in a special case, a learning environment. Since the user / learner is also an autonomous agent pursuing his / her own goals, the decision of which and whose goals will be pursued (the pedagogical agent's or the learner's) is made interactively, in a process of negotiation and persuasion.

In pursuing its goals, an application agent uses explicitly its relationship with the user/ learner. It can modify the parameters of the relationship, so that it can adopt user goals or learner goals and provide resources for them (achieving in this way explorative learning), infer and adapt to the learner's goals (to provide adaptive help or coaching) or try to make the learner achieve the teaching goals of the agent (to instruct the user / learner how to do something).

The second major trend in the development of teaching systems is that there is no virtual difference between humans and application agents. It is no longer necessary that the teaching system is an almighty teacher knowing the answer to any question that may arise during the interaction /learning session. Networking provides a possibility to

find somewhere else a system or a human- partner who can help the learner with his/her problem and explain him/her themes that the system itself can not. This trend can be seen in the increasing work on collaborative learning systems, which are able to find appropriate partners for help or collaboration, to form teams and support goal-based group activities [1], [2]. For this purpose, it is imperative that teaching systems (and other computer applications providing adaptive help) be able to communicate information about their users (user models) and about their available resources and goals in order to find an appropriate partner. We can imagine application agents, attached to every application or learning environment, which have an explicit representation of the user's or application's goals, plans, and resources. These agents communicate and negotiate among themselves for achieving their goals. This means that we need an appropriate communication language about goals and resources, which would allow these agents to share information. This communication has to be on a higher level than the level of knowledge communication (as in KQML or KIF), since it has a different purpose. While KQML and KIF have to define how the agents communicate their knowledge, this higher level of communication has the purpose to define who will be contacted, about what, when and how communication will take place (i.e. in which direction etc.). This level of communication has to be also transparent for humans, since some of the partners may be human-agents.

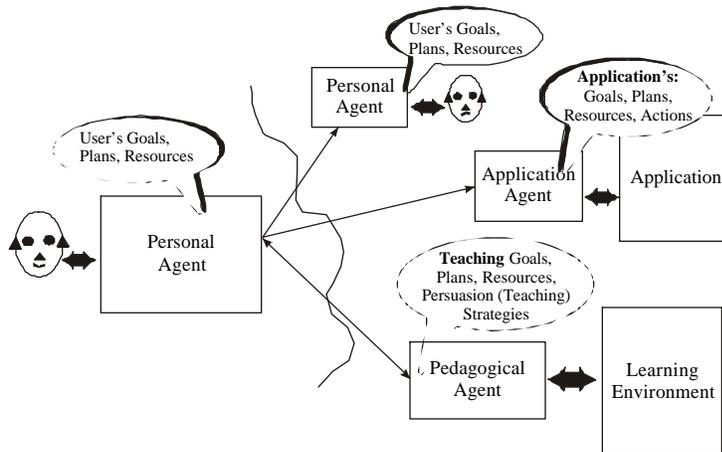
## 2 Goal-Based Agents

We propose creating "application agents" associated with applications, tutors, coaches and learning environments (see Figure 1). These agents will possess an explicit representation of the goals for which the application has resources and plans. Usually, these goals are embedded implicitly in the application at design time and can be *achievement goals* (for example, creating a table in Word), *typical user tasks* (which the application supports), and *user preferences* (normally also embedded at design time). Teaching applications have normative *teaching goals* (i.e. what the application is supposed to teach), which can be further classified into content goals, presentation goals / tasks, psycho-motor and affective (motivational) goals. Every application is provided, at design time, with resources and plans for achieving these goals (data, knowledge, functions of the application).

Human agents also possess goals, resources and plans. Classifications of human goals have been proposed by Schank & Abelson [5] and later by Slade [6]. Slade also proposes various dimensions for goals, like polarity, persistence, frequency, deadline etc., which influence the goal-importance. Human resources can be divided into two categories: tangible (money, time, objects, skills, credentials, rank etc.) and cognitive (memory, attention, knowledge, affects, moods). The resources can be classified with respect to whether they are perishable, expendable, interchangeable, transferable etc.

Humans communicate their goals, available resources and plans to their "personal agents", which serve as mediators in search for other application agents or personal agents that can provide resources and plans for achieving the goals of the human users. In this way, a human user and a software application appear in a symmetric

position, they possess goals, resources and plans, and they can adopt each other goals (i.e. help each other achieve their goals) mediated by the "pedagogical agents" and the "personal agents" (see Figure 1.)



**Fig. 1.** Personal and Application Agents

### 3 A Goal-Theory of Agents

According to Slade's [6] theory of goals, the behavior of a goal-based agent (for example, humans) follows the principles:

**Principle of Importance** *The importance of a goal is proportional to the resources that the agent is willing to spend in pursuing this goal.*

Humans possess not only tangible resources, like money, time, hardware configuration, etc., but also cognitive resources, like attention, memory, affect, moods etc. In order to infer each other's goals, the agents use the following

**Principle of Investment** *The importance of an active goal is proportional to the resources that the agent has already expended in pursuit of that goal.*

The place of the human -agents specific resources (attention, affects and moods) in the goal framework is explained below.

**Motivation** *An agent's motivation to pursue a given goal is equivalent to the importance of the goal for the agent.* So, the motivation of an agent to pursue a given goal is proportional to the resources that the agent is willing to spend in pursuing that goal.

Attention is the amount of processing time spent by the agent in pursuing some goal. The importance of a goal is proportional to the attention / amount of processing time which the agent is willing to expend in pursuit of that goal.

Affects The importance of a goal is proportional to the degree of affective response to the status of that goal. This means that the difference among happiness, joy and ecstasy relates to the importance of the goal that is achieved or anticipated. Happiness does not depend only on the world, but also on one's idiosyncratic goal hierarchy. Knowing the human's emotion (via an appropriately designed interface e.g. buttons allowing the user to directly communicate his / her emotion), the personal agent can infer the importance of the goal, on which the user has just failed or succeeded. Knowing the importance of the user's goal, the system can predict the affective response to goal-achievement or -failure.

Moods Goal persistence is reflected in persistent affective states, or "moods". The intensity of the mood reflects the importance of the related goals. An agent is in a good mood when after having achieved an important persistent goal. According to the principle of importance, the agent might be prepared to expend considerable resources to achieve this goal. Now these resources are free for other goals. So an agent in a good mood effectively has excess resources that could be used for new goals. An agent in bad mood has a lack of resources, and as such will be much less open to the pursuit of any new goals. This fits with our everyday experience. There is a heuristic that suggests that people try to put someone in a good mood before delivering bad news or making a request. This can be used in order to decide whether a system should teach the user (i.e. make him/her adopt the system's teaching goal) or whether it should adapt to the user (i.e. the system adopts the user's goal).

## 4 Relationships

An agent must act in a world populated by other agents. Many of an agent's goals require the help of another agent. In this way, relationships among agents can be viewed as another kind of resources for achieving goals.

Principle of Interpersonal Goals Adopted goals are processed uniformly as individual goals, with a priority determined by the importance and context of the relationship.

Parameters of Relationships: Inter-agent relationships can be characterized by the following parameters:

*Type of the other agent involved in the relation:*

- Human
- Computer

*Type of Goal Adoption*

- Goal Assignment
- Goal Inferring (adapting)
- Goal Development (teaching)

*Symmetry of Relationship*

- Asymmetric: dominated
- Asymmetric: dominating
- Symmetric

*Sign of Relationship*

Positive <[-----[-----]-----] > Negative  
collaboration cooperation competition adverse

The *importance* and the *context* of the particular relationship for a given agent determine both what goals the agent will adopt and what importance will be assigned to these goals. The principle of importance applies to adopted goals, meaning that an agent will expend resources in pursuit of an adopted goal in proportion to the importance of the adopted goal. The same applies to cognitive resources. For example, the agent will spend more attention (time thinking) about the interests or problems of a close friend than those of an acquaintance.

If an agent A wants that another agent B adopts A's goal as an important goal, it uses persuasion strategies. These are inter-agent planning strategies that aim at increasing the relative goal importance for another agent of an adopted goal. Persuasion strategies may, for example, exploit the importance of the relationship between A and B ("if you don't do this for me, I won't play with you anymore"), they can increase the importance of the goal B by bargaining resources, or offering that A will achieve a goal of B in exchange ("if you do this for me, I will do that for you"). A set of persuasion strategies was proposed by Schank & Abelson [5]. Persuasion strategies are particularly important when the type of goal adoption is "Goal Development", i.e. when agent A has a teaching goal and it wants that agent B adopts this goal as an important goal (thus B will be motivated to achieve the goal, i.e. to learn). We can consider some teaching strategies used by human teachers as a special case of persuasion strategies aimed at motivating the student to achieve some teaching goal. One can easily find parallels between the conditions for successful persuasion formulated by Slade [6] and some conditions for successful teaching. Slade's conditions are the following:

- the agent B should be aware of A's goal,
- B has to be in possession of resources and plans for pursuing A's goal,
- the relationship between A and B has to be perceived as important by B,
- B must not have any important goals that conflict with A's goal (i.e. with respect to some resource),
- the goal of A has to be intrinsically important.

For example, the first condition states that the student has to know exactly what is the goal he /she is trying to achieve. The second condition states that the student has to have the necessary prerequisite knowledge and cognitive resources free at the moment in order to be able to pursue a given teaching goal.

Another important characteristic of a relationship is its closeness. The closeness of relationship denotes how well the agents understand (are aware of) each other's goals. When talking about closeness, we always mean closeness from the point of view of a certain agent: it can well be the case that one agent understands the goals of the other

one well, but the second one is not aware of the goals of the first one. Agents learn about each other's goals from three main sources:

- Normative knowledge about other agents' goals (stereotypes embedded at design time)
- Statement of a goal by an agent (direct goal communication via some language)
- Inferred knowledge arising from observation of an agent's behavior (diagnosis as in user modeling systems).

The closeness of the relationship between a human user and his / her personal agent should be as high as possible. This means that a personal agent must use not only normative and directly communicated knowledge about the user's goals, but should be able to infer user goals from his /her behavior (methods for diagnosis in user modeling and plan recognition could be applied), as well as infer the user's affect and moods (communicated in some way from the user to his /her personal agent).

*Definition: An agent, who is able to represent explicitly, reason about and modify the parameters of its relationships with other agents, when it is able to create and destroy relationships according to its goals, is called a social autonomous goal-driven agent.*

Personal agents and application agents (which can be pedagogical agents if the application is a learning environment) are examples of social autonomous goal-driven agents.

A personal agent can be related with the human user with an asymmetric goal-assignment type of relationship (i.e. the application receives and executes commands from the user). In this case, the personal agent searches among the available relationships with application agents the one that is most appropriate for achieving the current user's goal. If no such relationships are available, it will contact a broker-agent that has a large number of relationships to various application agents (including pedagogical agents). The broker will find and contact an application agent that can fulfill the goal. The personal agent will negotiate with the application agent (see Figure 1) in order to find reasonable conditions for obtaining the service. When agreement is achieved, the application agent adopts the user's goal and provides its normative resources and plans for achieving the user's goal.

Pedagogical agents have a set of normative teaching goals and plans for achieving these goals as well as persuasion plans (i.e. teaching strategies) and associated resources in the learning environment. The persuasion plans can involve a modification of the parameters of the relationship between the pedagogical agent and the user. For example, it can change the sign of the relationship from collaborative to competitive (game-like) or to cooperative (simulating a peer-learner), or the symmetry of the relationship from user-dominated (a constructivist type of learning environment), to symmetric (coach), or pedagogical agent dominated (instructivist tutor).

If a personal agent is able to reason about and modify the relationship with the user, it can decide not only to fulfil the user's orders, but also to change the relationship to become more symmetric, or to take initiative and teach the user something suggested by the corresponding application agent. For this to happen, the personal agent should have adopted a goal from a pedagogical agent of some learning environment (who has managed to persuade the personal agent that it is in possession

of resources and plans for pursuing a teaching goal related to the current goal of the user). In this case, the personal agent has to make a decision between two conflicting goals (the achievement goal of the user and the adopted teaching goal from the pedagogical agent of the learning environment). In order to be able to make decisions, the personal agent needs to be able to reason about the relative importance of the goals and to plan resources.

## 5 An Architecture for Autonomous Goal-Based Social Agents

Ideally, an autonomous cognitive agent will possess a reactive, reasoning, decision-making and learning capability. Therefore it has to contain processes implementing these capabilities. We propose an architecture for autonomous goal-based social agents (see Figure 2) which contains the following components:

- *Explicit Representation of Goals, Plans, Resources and Actions* available to the agent (Elements);
- *Explicit Representation of the Agent's Relationships* with other agents and ensuring that all the parameters of the relationship can be manipulated by the agent (Relationships);
- *Internal Processes* which can be applied to the elements, but also on the parameters of the relationships of the agent and for enforcing inter-agent goal-adoption (Processes).

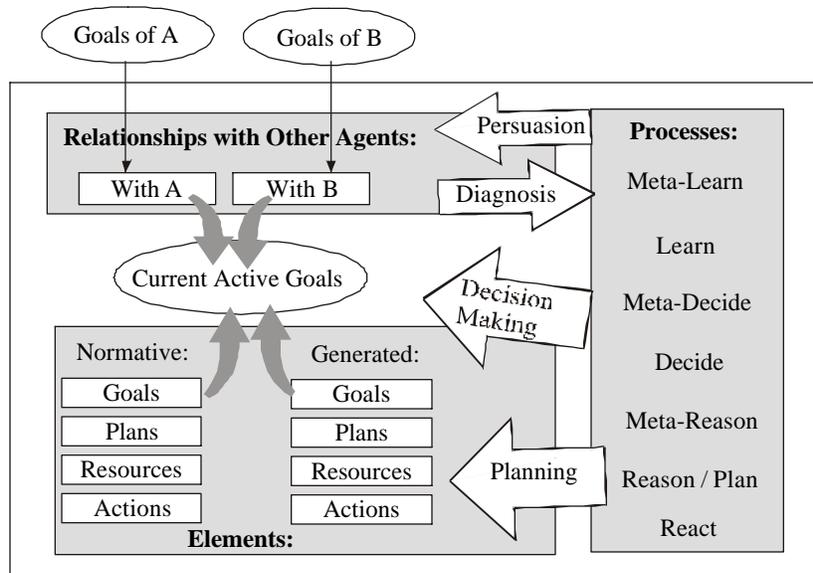


Fig. 2. An Architecture of Intelligent Personal /Application Agent.

A personal agent with this architecture should have the following properties:

- *Taking into Account User Motivation, Emotions, and Moods*: According to the principle of importance, the motivation can be naturally taken into account by assessment of the relative importance of goals.
- *Unified Model of Collaborative, Cooperative, Competitive, and Adverse Behavior*: A truly adaptive system should be able to modify its type of relationship with the user. The explicit representation of the sign of the relationships will allow the agents to reason and dynamically modify their behavior with respect to other agents.
- *The Agent Decides about the Style of Behavior* (passive or proactive) it will have with other agents (and with the user).

An agent defined in this way fulfils Nwana's [4] definition of a "smart" agent and is an ideal to which one could strive. However, to bring this architecture to a computational framework, one has to find new techniques for reasoning and decision-making about inter-agent relationships.

We are currently working on the implementation of this agent-based framework as a basis for the Intelligent Helpdesk (I-Help) project [3]. Students and applications (discussion forum, a WWW-based pool of teaching materials and an on-line help system) in this architecture are represented by agents who can communicate among themselves by exchanging messages about their goals, needs and available resources. This implementation is in JAVA, the agents are implemented as threads communicating with messages via TCP/IP on the Internet. We are defining a taxonomy of goals and resources for a selected domain, which is to be used by the agents. As a first version, the agents will communicate with a broker-agent who keeps information about the goals and resources the agents need or have to offer. Gradually, the agents will develop relationships, which will allow them to select "partners" without contacting a central service. As a start, the agents will be kept "slim", i.e. unintelligent. They will "borrow" reasoning, planning, persuasion and negotiation services (advice) from existing planning systems or specially developed for this purpose applications, represented in the frameworks by their application agents. Starting with a minimal communication scheme, we will extend the "intelligence" of agents by adding capabilities of reasoning about goals, inferring the importance of the goals of other agents, decision making and inter-agent planning (persuasion). The basic agent architecture can be completely unintelligent and serve purely communicative functions; i.e. not including any reasoning, planning, decision-making and learning capabilities. However, when it is necessary and possible, one can further develop these capabilities in any single agent, without adversely affecting the general scheme of interaction.

## 6 Summary

We are proposing an agent-based framework for supporting adaptation and teaching/learning in a distributed environment. Every application (working or teaching system) is provided with an agent, called "application" or "teaching" agent

which possesses an explicit representation of the goals of this application (achievement goals or teaching goals), the available resources, plans and relationships. Human users are provided with "personal" agents, which contain representation of the user's goals, available resources, plans and relationships (stored in a user model; inferred or assigned directly by the user). The representation of goals, resources, relationships and plans is based on a special taxonomy, using Slade's theory of goals [6]. When a user faces a problem and needs a specific adaptation in his/her working or learning environment, the personal agent searches for an appropriate adaptive application or a teaching program in the distributed environment (e.g. Internet). This is done by negotiating with the application agents and teaching agents of other applications or with the personal agents of other human users to find one that is appropriate, available, and willing to help.

**Acknowledgement:** Special thanks to Jim Greer for comments on earlier drafts of this paper.

## References

1. Hoppe H.U. The use of multiple student modeling to parameterize group learning. In Artificial Intelligence and Education: Proceedings of AI-ED 95, AACE, (1995), 234-241.
2. Collins J., Greer J., Kumar, V., McCalla, G., Meagher P., Tkatch, R. Inspectable User Models for Just-In-Time Workplace Training, in User Modeling, Proceedings of UM97, Springer (1997) Wien NewYork, 327-338.
3. Greer, J., McCalla, G., Cooke, J., Collins, J., Kumar, V., Bishop, A., Vassileva, J. The Intelligent Helpdesk: Supporting Peer-Help in a University Course, Proceedings ITS'98 (this volume), (1998) , Springer Verlag: Berlin-Heidelberg.
4. Nwana H. (1996) Software Agents: An Overview, *Knowledge Engineering Review*, 11, 1-40.
5. Schank, R., Abelson, R. *Scripts, Plans, Goals and Understanding*. Lawrence Erlbaum Assoc., (1977) Hillsdale, NJ.
6. Slade, S. *Goal-Based Decision Making: An Interpersonal Model*. Lawrence Erlbaum Associates Inc. (1994) Hillsdale, NJ.
7. Vassileva J. Reactive Instructional Planning to Support Interacting Teaching Strategies, in *Proceedings of the 7-th World Conference on AI and Education*, 334-342, (1995) AACE: Charlottesville, VA.
8. Vassileva, J. A task-centered approach for user modeling in a hypermedia office documentation system, *User Modeling and User Adapted Interaction*, (1996) **6**, (2-3), 185-223.