Software agents that learn through observation (Short paper)

Jaçanã Machado
ADETTI / ISCTE
We, the Body and the Mind Research Lab
+351 21 782 64 84
Jacana.Machado@we-b.mind.org

Luís Botelho
ADETTI / ISCTE
We, the Body and the Mind Research Lab
+351 21 782 64 84
Luis.Botelho@we-b-mind.org

ABSTRACT

In this paper, we present an architecture for software agents that enables them to learn vocabulary through the observation of each other bodies and actions. Besides sensors, effectors, and action control, the architecture provides the equivalent of a body with visual appearance. The agent soft visual appearance is designed to be seen by other software agents, not by people. The paper describes the agent software body with visual appearance, the learning mechanism, the demonstration scenario and presents some results showing that the agent software body allows agents to learn vocabulary through observation and to ground the meaning of the symbols they learn. The paper emphasizes the important role of social interaction in learning processes.

Categories and Subject Descriptors

I.2.6 [**Artificial Intelligence**]: Learning – *language acquisition.*

General Terms

Algorithms, Experimentation, Theory

Keywords

Learning vocabulary through observation, Embodiment, Software body, Software visual appearance, software agents

1. INTRODUCTION

One of the most remarkable features of living bodies is their visual appearance, which enables them to communicate and to learn through mutual observation.

Having a body enables agents to benefit from the advantages embodiment accomplishes for living beings [1]. This work proposes a component based agent architecture with software visual appearance that enables agents to learn simple vocabulary (names of body parts and actions) through observation. This is a short version of the paper *"Software agents that Learn through observation"*[1]

Our work is centred on the problem of learning simple vocabulary (nouns and verbs) through observation. We have implemented some learning experiments that provided some evidence supporting the validity of our proposal. In our learning experiments there are only two agents: the Observed Agent, and the Observer agent. The Observer Agent learns the words spoken by the Observed Agent and grounds their meanings in its own body and actions using a vocabulary map. The Observed Agent shows parts of its body (e.g., hand) and executes simple actions (e.g., it shakes its hand) while it says words that represent nouns (e.g., "hand") or words that represent verbs (e.g., "shake") naming the shown body parts or executed actions. Nouns are related to parts of the body, verbs are related to actions of the body [3]. This process requires that agents have body that can be observed by other software agents.

The reminder of the paper is organized as follows. Section 2 describes related work. Section 3 describes the demonstration scenario, and the results that have been achieved in our learning experiments. Section 4 describes our proposal. Section 4.1 describes the component-based architecture which sensors, effectors, action control, Learning Mechanism, social skills, and soft visual appearance. Finally, in section 5, we present conclusions and describe future directions for further research.

2. RELATED WORK

Artificial Intelligence research was founded under the influence of classical cognitive theory and the issue of symbolic representation: the mind is a symbol system and cognition is symbol manipulation. Generation of complex behaviour through symbol manipulation was successfully demonstrated by classical artificial intelligence [4].

Modern cognitive science has reviewed old concepts of pure mental processes and started considering the importance of the body in cognitive processes. For modern AI researchers and philosophers, there cannot be intelligence without a body. Modern artificial intelligence researchers, like Tom Ziemke[1], Brooks[2], Kushmeric[5], Varela[7], and Mataric[6] have been studying the role of the body on the generation of intelligent behaviour in natural and artificial systems. Principles of software embodiment architectures have been proposed by Kushmeric [5], and Botelho and Figueiredo [1]. Kushmeric has developed a computational framework for analyzing embodiment and identified several ways embodiment may reduce the complexity of the methods used in the agent's tasks. Botelho and Figueiredo have proposed concrete architectural principles for embodied software agents [1], a class of software agent capable of exhibiting some of the beneficial roles embodiment accomplishes in living beings, such as facilitating learning processes and social interactions.

AAMAS'06, May 8-12, 2006, Hakodate, Hokkaido, Japan.

[1] iscte.pt/~luis/papers/AAMAS2006Learning.pdf

3. DEMONSTRATION SCENARIO

In our demonstration scenario, the observed and observer agents are anthropomorphic. Their bodies are composed of the following components: mouth (i.e., textual language effector), ear (i.e., textual language sensor), eye (static and dynamic visual sensor), hand, action control, learning mechanism, action capturing mechanism, vocabulary map, and graphic user interface. Even though they could have different bodies, we used two agents with identical bodies. The Observed Agent teaches two concepts: *hand* and *shake hand*. It shows its hand and says the word "hand", and it shakes its hand and says the phrase "shake hand". In the described experiments, "saying" and "hearing" are accomplished through the exchange of textual messages.

The proposed approach can be used for agents of different nationalities. Therefore it can be used in cases in which a French speaking agent needs to learn English or some other language. When learning the word "hand" and its meaning, the Observer Agent identifies two body parts as possible candidates to match the Observed Agent part that is named by the word "hand": *pied* (French word for foot) and *main* (French word for hand). Given the ambiguity, the Observer Agent asks the Observed Agent for help. The Observed Agent knows that *hand* is part of *arm* and *foot* is part of the *leg* therefore it can help the Observer Agent to choose the correct part. After disambiguating the meaning of the learnt word, the Observer Agent imitates the observed action: it shows its hand, generates the textual expression "*I have a hand*" and sends it to the Observed Agent. The Observed Agent recognizes the success of the learning experiment and sends a positive textual message to the Observer Agent. Finally, since the Observer Agent knows the learning has been successful, it creates an entry in its vocabulary map to the noun "hand" and its grounding: the source of information is the Observed Agent; the observer agent class for "hand" is called *hand*; and its own class for hand is called *main*.

In the second experiment, the Observed Agent shakes its hand and says the phrase "shake hand". The Observer Agent identifies its body parts *pied* and *main* as possible candidates to perform the action named by the phrase "shake hand". The Observed Agent knows that *hand* is part of *arm* and *foot* is part of *leg* therefore it can help the Observer Agent to choose the right part. Then, the Observer Agent performs the actions that can be done with the class *main*. This process ends when it determines that the action *agiter* (French word for "to shake") performed by its *main* matches the observed action (through the visual appearance of the Observed Agent). Therefore the Observer Agent shakes its hand, and generates the expression "To shake the hand" and sends it to the Observed Agent. This one recognizes that the learning process was successful hence it sends a positive textual message to the Observer Agent. Finally, the Observer Agent creates a new entry for the verb "to shake" and its grounding in its vocabulary map: the source of information is the Observed Agent; the observed agent class of the body part involved in the observed action is *hand*, its own class of the corresponding body part is *main*; the observed agent method that implements the action is called *shake*; and its own method that implements the observed action is called *agiter*.

These two learning experiments show that the proposed architecture can support simple forms of vocabulary learning through observation. The architecture supports both noun learning and verb learning. The experiments also emphasize the role of social interaction for disambiguating the result of the learning process. The experiments also show that the architecture enables the agent to ground the learnt symbols in its own body and actions. Although the used algorithms exhibit some limitations (e.g., they do not work for learning actions performed by several parts of the body), the general processes are independent of the agents specific make up. The same algorithms could have been applied to agents with FTP and SMTP sensors and effectors instead of hands and feet.

4. LEARNING SIMPLE VOCABULARY THROUGH OBSERVATION

The agent architecture supporting the described learning experiments includes the mechanisms required to create the software-equivalent to the agent visual appearance and the mechanisms required for learning through observation. The agent visual appearance can be seen (consulted) by its owner and by other agents.

4.1 Agent Architecture and Software Visual Appearance

The proposed agent architecture is a component-based architecture with several components providing the following functions: sensors, effectors, vocabulary map, control mechanism, learning mechanism, and software visual appearance (static composition and dynamic behavior).

Sensors support visual and textual language perceptions. Visual sensors are used to capture visual appearance while the textual language sensor is used to capture textual messages. The visual sensor captures dynamic and static information of the observed body. The learning mechanism is the algorithm used by the Observer Agent to learn through observation. The vocabulary map represents the words known to the agent together with their meanings. It grounds the meaning of each word through a relationship between known word and its type (verb / noun), the source of information from where the word was acquired (in our case, this is the identification of the Observed Agent or the special symbol *built-in*), the classes of the component named by the acquired words or the classes of the component that executed the action named by the acquired words (in the Observed Agent's body and in the Observer Agent's body), and the signatures of the methods of the agent program encoding the action named by the stored word (on the side of the observed agent and on the side of the observer agent). All these attributes are mandatory, except method signatures which are mandatory only when the type of word is verb.

The agent software visual appearance consists of static information regarding the composition of the agent body and dynamic information about the actions the agent executes. Software visual appearance (both static and dynamic information) is made publicly available to other agents by the agent software visual appearance component. The agent software visual appearance is a component that captures information regarding its static body composition, and information about its dynamic body actions. Static visual appearance includes information regarding the composition, properties and roles of the agent visible components. The information regarding the agent's components and their roles is used by Observer Agents to create a static mental image of the parts integrating Observed Agents. This mental image is interpreted as the static visual appearance of the Observed Agent. Dynamic elements of the body are used by the Observer Agent to create a dynamic mental image of the actions of the Observed Agent. Dynamic visual image relates to what the agent is doing at each instant of time.

"Seeing another agent" means consulting its software visual appearance and performing the required inferences. The dynamic information maintained by the agent visual appearance is captured by a crosscutting mechanism called the Action Capturing Mechanism. The Action Capturing Mechanism captures all relevant elemental operations performed by the agent involving the visible parts of its body, including get and set operations (i.e., get and set the value of a visible attribute of any of the visible agent components), entry and exit operations (i.e., invoking and exiting a method that performs a visible action). Composite actions are made up of sequences of elemental actions therefore they can also be seen through the agent software visual appearance.

4.2 The Learning Mechanism

Learning is facilitated through textual interactions among the two agents. The Observed Agent initiates the interaction by showing a part or an action of its body, and saying a word naming the shown part or action. When the Observed Agent shows its body parts and actions, the relevant part/action is marked, in its soft visual appearance. Therefore the Observer Agent knows exactly what is being shown, and the Observed Agent does not need to have salience building mechanisms. In real life scenarios, learning vocabulary through observation would require that the Observed Agent would be capable of building salience on the parts and actions it shows because, when the Observer Agent looks at the Observed Agent body, it sees the whole body, not only the relevant parts and actions intended to be shown. We overcame this salience building problem through a simplification consisting of marking the salient parts of the visual appearance. The visual appearance does not contain the information about the type of word (noun or verb). There is no visual difference between showing a body part or perform an action: both are actions. The difference is given by the textual message said by the Observed Agent.

The Observer Agent triggers its learning mechanism when it hears and sees the Observed Agent showing its body parts or executed actions while saying its names. The Learning mechanism identifies the part of the body/action shown by the observed agent associating it to the heard word. It compares the observed part/action with its own body parts/action in order to select those that more closely mirror the observations. When the learning agent discovers the part or the action of its body matching the observations, it grounds the meaning of the learnt words in its own body and actions. When the Observer Agent learns a word, it generates sentences using the learnt word, using pre-coded sentence schemata and the vocabulary already mastered by the agent. The Observed Agent provides feedback regarding the accuracy of the learning results. When learning is perceived to be accurate, the Observed Agent generates a positive textual reaction. When learning is perceived to be inaccurate, the Observed Agent generates a negative textual reaction. This feedback allows the Observer Agent to know whether its learning was correct or incorrect.

5. CONCLUSION AND FUTURE WORK

The strongest contribution of this paper is to present a concrete software agent architecture exhibiting some of the features required to have a body with software visual appearance that can be seen by other software agents (not by people) through which it is possible to improve several cognitive tasks, in particular "learning simple vocabulary through observation". The proposed approach can be used to overcome some criticisms put forth against classical AI. Namely, through the proposed approach, it is possible for an agent to ground the meanings of acquired vocabulary. That is, in addition to being understood by third parties (agent designers and observers), learnt symbols acquire meaning to the agent itself. The proposed architecture and algorithms allow the agent to ground the meanings of the acquired vocabulary on the parts and actions of its body. This allows agents to create first person meanings for the symbols they process. The defined and implemented demonstration scenario shows that the proposed architecture and the used learning algorithms are general enough to support learning through observation in diverse application domains. The demonstration also shows that social interaction is an important condition for learning through observation. This is most apparent when the learning process generates ambiguity. The proposed learning process can also be used in different situations such as those in which an agent must learn a different communication language.

A direction for future work is to build the mechanisms necessary to increase the salience of the shown body parts and actions. We will explore the role of emotion as the basis for those processes.

6. REFERENCES

[1] Botelho; L. and Figueiredo, P. 2004. "What Your Body and Your Living Room Tell My Agent". Proceedings of the AAMAS 2004 Workshop "Balanced Perception and Action in Embodied Conversational Agents"

[2] Brooks, R. 1991. *Intelligence without Representation*, Artificial Intelligence 47(1991) 139-159 Elsevier

[3] Cangelosi, A., Riga, T., Giolito, G., and Morocco, D. Language Emergence and Grounding in Sensorimotor Agents and Robots. *First International on Emergence and Evolution of Linguistic Communication,* May 31 – Jun 1 2004, Kanazawa Japan.

[4] Harnad, S. 1990. The Symbol Grounding Problem. Physica D 42: 335-346.

[5] Kushmerick, N. 1997. "Software agents and their bodies". Journal of Minds & Machines, 7(2):227-47

[6] Mataric, M. Studying the Role of Embodiment in Cognition. In Cybernetics and Systems, Special issue on Epistemological Aspects of Embodied AI, 28(6):457-470.

[7] Varela, F., Thompson, E. & Rosch, E. (1991) *The Embodied Mind - Cognitive Science and Human Experience*.Cambridge, MA: MIT Press.

[8] Ziemke, T. 1999. "Rethinking grounding". In Riegler, Peschl & von Stein (eds.) Understanding Representation in the Cognitive Sciences (pp. 177-190). Plenum Press