The Collective Pet Unconscious: Balancing Intelligence and Individuality in Populations of Learning-Enabled Virtual Pets

Ben Goertzel
Novamente LLC
1405 Bernerd Place
Rockville MD 20851 USA
ben@novamente.net

Cassio Pennachin
Novamente LLC
1405 Bernerd Place
Rockville MD 20851 USA
cassio@novamente.net

CHI 2009, April 4 – 9, 2009, Boston, MA, USA
ACM 978-1-60558-246-7/09/04.

Abstract
A conceptual architecture is presented in which a population of learning agents share a dynamic collective knowledge base but also retain individual memories, biases and learning systems. Knowledge learned by agents goes into the collective base, but each agent also has a "personality filter" that controls how the collective base affects its individual base. An application in virtual worlds is discussed, in which behaviors are taught by users to virtual dogs and the shared knowledge is used to speed up learning.

Keywords
Learning, virtual worlds, knowledge sharing

ACM Classification Keywords
I.2.6. Artificial Intelligence: Learning

Introduction
In the real world, different organisms can share their knowledge and learning with each other only via low-bandwidth mechanisms like imitation and linguistic communication. Artificial agents are in a different

situation, as knowledge can be extracted from one agent's knowledge base (KB) and shared with others. Knowledge sharing has been used to transfer learned behavior from virtual agents to robots [6], and to coordinate simultaneous learning by agent groups [1].

One could create a population of knowledge-sharing agents that operate as a sort of "borg mind," with intelligence beyond what any of the agents could achieve with its own resources, with each giving contextually appropriate expression to the same knowledge and intelligence. However, in some cases preserving a significant degree of separation between the KBs of different agents may actually be optimal in terms of advancing the total intelligence of the population; this is similar to the reasons why in evolutionary learning one sometimes uses an "islands model" [5] consisting of a set of separately evolving subpopulations with limited interaction, rather than one big population. In many practical applications what the end users of virtual agents want is also an agent with its own strengths and weaknesses, and its own learning process that the agent's owner gets to participate in.

In this paper we explore these issues in a generally applicable way, but focus on the example of virtual pets. We present a conceptual system architecture involving a group of virtual agents, each possessing their own individual KB and learning processes, hooked up to a group knowledge store that one might whimsically call, in the pet application, a "Collective Pet Unconscious." This centralized KB may be associated with its own learning processes. However, its impact on the individual agents must be carefully controlled, and hence in our architecture each agent is supplied with a "personality filter" component that determines which knowledge from the group mind enters the individual mind. The filter allows the balance between rapidity of learning and distinctiveness of the agent to be configured based on the goals of the designers of the agent and the virtual world in which it is embedded.

A Conceptual System Architecture for Balancing Collective Intelligence and Individuality

The conceptual system architecture we describe here has currently been implemented in the context of virtual pets interacting in a dog park, but has much more general applications. Generally conceived it comprises a population of virtually embodied agents, embodied in the same world(s), where each agent possesses an individual knowledge store; a learning unit; a personality model that distinguishes it from other agents; and a personality filter. Additionally, there are software objects corresponding to a "coherent group" of different agents (possibly all the agents in the population, or possibly a subset thereof). The agent groups also have their own knowledge store (the centralized KB) and learning unit.

In our current application of the architecture, the agents reside in a virtual, controlled by the Novamente Cognition Engine (NCE), an integrative AI system [2].The NCE contains a single KB, as well as specific knowledge objects for agents and agent-groups. Thus, each agent's KB is a subset of the collective KB. A set of processes act periodically on each agent's knowledge store, including the personality filtering process.

The same architecture applies beyond the domain of virtual worlds. One could use it to control physical robots that connect wirelessly to a centralized KB. It also applies to agents that lack “bodies” in the traditional sense, such as search agents that navigate a large textual or relational knowledge base, each using their own particular biases to guide their search.

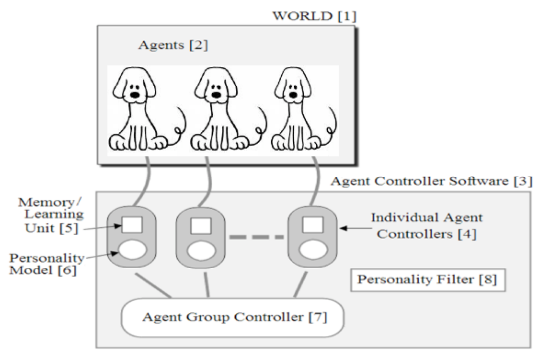

Figure 1. A Conceptual System Architecture for Balancing Collective Intelligence and Individuality

Figure 1 gives a simple graphical depiction of the architecture. In the figure, (1) represents the world, in which computer-controlled simulated agents (2) are situated, along with similar agents controlled by humans. The agents (2) are collectively controlled by agent controller software (3), which internally consists of a number of components (4-8). Each agent has its own agent controller (4), which contains a memory and learning unit (5) and a personality model (6). There is also an agent group controller (7) that controls the population, and stores the collective KB. The personality filter component (8) controls which of the collective knowledge in (7) is imported by individual agent controllers (4), based potentially on multiple constraints, but including the constraint that each individual agent (2) must continue to display its own individual personality even after absorbing appropriate items of collective knowledge from (7).

Toward a Collective Pet Unconscious

The above architecture is quite general, but we have initially explored it in a very specific context: the Novamente Pet Brain architecture [3-4], which uses a simplified version of the NCE to control virtual dogs embodied in online virtual worlds. Figure 2 is a screenshot of two Pet Brain controlled pets in Multiverse, together with an indicator pane showing the emotional and physiological status of one of the pets. While they have separate bodies, their minds are stored in the same Pet Brain software system on the same server, and hence may share knowledge as much as the system's configuration allows them to. They can be taught new tricks and behaviors through an imitation/reinforcement learning process.

To illustrate “personality filtering” in the Pet Brain, suppose that Jane teaches her pet Fido to sit on command, and it takes her 5 trials to do it. Once this behavior is taught, it goes into Fido’s memory, and also into the collective memory. Afterwards, suppose that Jack wants to teach his pet Princess to sit on command. Once Jack gives Princess one example of sitting, Princess could use the knowledge Fido put in the collective memory, her personality filter allowing, to figure out what Jack wants. Princess will use Fido's experience to learn much faster than Fido did. However, Jack will not really get the fun of watching his dog learn to sit. If Princess’s personality filter is set more restrictively, then she may not access Fido’s knowledge about sitting at all, or she may use it only weakly, to bias her learning rather than to strongly guide it, in which case it might take her 3 trials to learn to sit, instead of the 1 that would come from a pure "borg mind" strategy, or the 5 that would come from pure individuality.

Figure 2. Two Pet Brain controlled pets in Multiverse,

This process works even if Jack and Jane call the same behavior totally different things. The collective-memory-based matching uses the examples of sitting behavior that Jack and Jane have provided, not just the words they have associated with the behavior.

In this scenario, collective memory is an intelligence advantage and the only purpose of personality filtering is to preserve the pleasure of pet individuality for the end user. However, in the cases of more complex behaviors there may be an intelligence advantage in maintaining a significant degree of agent individuality, as it would allow subsequent learners more liberty to try variations on already learned behaviors. They may then discover better ways to execute those behaviors. In this case, it would seem appropriate for the personality filter to restrict collective knowledge more strictly. Thus, the crafting of an application-appropriate personality filter may be quite subtle.

There are many possibilities to be explored here, and this aspect of virtual pet psycho-engineering is at its very beginning. Our goal here has been to report some of the simple experiments we've carried out so far, and to describe a general architecture which we believe will have far-reaching applications as virtual pets and virtually-embodied agents more generally advance.

References

[1] García-Martinez, R., Borrajo, D., Maceri, P. and Britos, P. Learning by Knowledge Sharing in Autonomous Intelligent Systems. *LNCS 4140*, Springer (2006).

[2] Goertzel, B. Patterns, Hypergraphs and General Intelligence. *Proc IJCNN 2006*.

[3] Goertzel, B., Pennachin, C. and de Souza, S.A. An inferential dynamics approach to personality and emotion driven behaviour determination for virtual animals. *The Reign of Catz and Dogs Workshop at AISB-08*, 2008.

[4] Goertzel, B., Pennachin, C., Geissweiller, N., Looks, M., Senna, A., Silva, W., Heljakka, A. and Lopes, C. An Integrative Methodology for Teaching Embodied Non-Linguistic Agents, Applied to Virtual Animals in Second Life. *Proc. AGI-08*, IOS Press (2008), 161-175.

[5] Whitley, D., Rana S. and Heckendorn, R.B (1999). The island model genetic algorithm: On separability, population size and convergence. *Journal of Computing and Information Technology* 7 (1999), 33-47.

[6] Yamaguchi, T., Masubuchi, M., Tanaka, Y. and Yachida, M. Propagating learned behaviors from a virtual agent to a physical robot in reinforcement learning. *Proc IEEE International Conference on Evolutionary Computation*, IEEE (1996), 855-859.