Chapter 5
A High Level Model of a Conscious Embodied Agent

Jiří Wiedermann
Academy of Sciences of the Czech Republic, Czech Republic

ABSTRACT

In this paper, the author describes a simple yet cognitively powerful architecture of an embodied conscious agent. The architecture incorporates a mechanism for mining, representing, processing and exploiting semantic knowledge. This mechanism is based on two complementary internal world models which are built automatically. One model (based on artificial mirror neurons) is used for mining and capturing the syntax of the recognized part of the environment while the second one (based on neural nets) for its semantics. Jointly, the models support algorithmic processes underlying phenomena similar in important aspects to higher cognitive functions such as imitation learning and the development of communication, language, thinking, and consciousness.

INTRODUCTION

Hyakujo wished to send a monk to open a new monastery. He told his pupils that whoever answered a question most ably would be appointed. Placing a water vase on the ground, he asked: "Who can say what this is without calling its name?" The chief monk said: "No one can call it a wooden shoe." Isan, the cooking monk, tipped over the vase with his foot and went out. Hyakujo smiled and said: "The chief monk loses." And Isan became the master of the new monastery.

The introductory quotation related to the life of Hyakujo Ekai (also known as Baizhang Huaihai in China) (720-814) indicates that the masters of the Zen philosophy knew the key to the problem known in the artificial intelligence as the *symbol grounding problem*. This problem is related to

the question of how words get their meanings, and of what meanings are (Harnad, 1990). Artificial intelligence started to be interested in this problem especially in nineteen eighties when this problem has become popularized as the Chinese room thought experiment designed by philosopher John Searle (1980). This experiment pointed to the problem whether a computer can understand the words (i.e., the symbols) by which it communicates with people. Searle highlighted the fact that with the symbols a computer cannot do anything else than, following some rules, to transform them into other symbols, eventually into sequences of zeros and ones whose semantic a computer cannot know either. Henceforth, this cannot be the way for computers to get the meaning of words, or "semantic knowledge". This thought experiment has started an ongoing discussion which has (among other things) lead to the development of theories claiming that in order the AI systems to "understand" their actions they need to have a body – they need to be "embodied". It is the body, usually in the interaction with the environment, which gives these systems a facility for understanding their own actions and those of other embodied agents, and for communicating with each other one. By the way, this was the facility used by Isan, the cooking monk from our quotation: not having a body and there not being a real water vase, Isan could not kick the vase with his foot, the vase would not tip over, it would not emit the corresponding characteristic sound and the monks could not see this event. All this, perceived by the sensors of all participants, would not invoke in their minds the set of corresponding memory and sensory associations which all together represent the semantics of the word "vase". Saying it poetically, Isan stroke a chord in the minds of bystanders as the unspoken world "vase" would do.

Returning to the Searle's Chinese room experiment we see that, in order to communicate "intelligently" with people, the computer should obviously have information how the people's

world looks like, what could be the abilities of people under various circumstance, etc. In short, Searle's computer should have possessed a kind of internal model of the external world (inclusively that of the self), represented in whatever useful way. This model should represent a part of rules by which the computer communicated with people.

The idea that non-trivial cognitive systems should build and exploit some form of internal world models has been around practically since the dawn of AI. However, efforts for controlling behavior by formal reasoning over symbolic internal world models have failed. Consequently, during nineteen nineties the mainstream research turned towards biology inspired behavior-based designs of cognitive systems. The respective approach has stressed the necessity of embodiment and situatedness used in sensorily driven control of behavior of simple robots, cf., Brooks (1991). This paradigm worked well with so–called subsumption architecture using incrementally upgraded layers of behavior realized by a task specific robot programming, cf., Pfeifer et al. (1999). Nevertheless, after a series of promising successes, mostly in building various reaction–driven robots, it has appeared that such a framework has its limits. Especially in humanoid robotics a further progress towards higher levels of intelligence turned out to be impossible without introducing further innovations into the basic architecture of cognitive systems with machine consciousness being the ultimate goal, cf., Holland (2003).

Nowadays, it is generally believed that in order to open the road towards higher brain functions we need automatic mechanisms that will augment the previously acquired semantic knowledge. These mechanisms often make use of internal world models. Presently, prevailing trends seem to prefer other than symbolic representation of the internal worlds, in most cases variants of neural nets. For an overview of the recent state-of-the-art and a discussion on internal world models, cf., Holland et al. (2003) or Cruse (2003).

Related Content

Route-Planning Algorithms for Amusement-Park Navigation

Hayato Ohwada, Masato Okadaand Katsutoshi Kanamori (2014). *International Journal of Software Science and Computational Intelligence (pp. 78-92).*

www.irma-international.org/article/route-planning-algorithms-for-amusement-park-navigation/127015

Artificial Intelligence Techniques for Unbalanced Datasets in Real World Classification Tasks

Marco Vannucci, Valentina Colla, Silvia Cateniand Mirko Sgarbi (2012). *Machine Learning: Concepts, Methodologies, Tools and Applications (pp. 414-427).*

www.irma-international.org/chapter/artificial-intelligence-techniques-unbalanced-datasets/56154

An Alternative Backward Fuzzy Rule Interpolation Method

Shangzhu Jin (2014). *International Journal of Software Science and Computational Intelligence (pp. 47-71).*

www.irma-international.org/article/an-alternative-backward-fuzzy-rule-interpolation-method/133258

Machine Learning for Biometrics

Albert Ali Salah (2010). *Handbook of Research on Machine Learning Applications and Trends: Algorithms, Methods, and Techniques (pp. 539-560).*

www.irma-international.org/chapter/machine-learning-biometrics/37003

Sentiment Analysis of COVID-19 Tweets Using Adaptive Neuro-Fuzzy Inference System Models

Sabri Sabri Mohammed, Brahami Menaouer, Abid Faten Fatima Zohraand Matta Nada (2022). *International Journal of Software Science and Computational Intelligence (pp. 1-20).*

www.irma-international.org/article/sentiment-analysis-of-covid-19-tweets-using-adaptive-neuro-fuzzy-inference-system-models/300361