Autonomy, Isolation, and Collective Intelligence

Nikolaos Mavridis NIKOLAOS.MAVRIDIS@NYU.EDU
New York University Abu Dhabi
P.O. Box 129188
Abu Dhabi, UAE

Is it total self-sufficiency that we are really after, or harmonious integration into, and facilitation of, ecosystems of intelligent entities, which can participate in wider entities beyond themselves?

In the last years, significant progress of numerous partial aspects of cognitive systems has taken place. We now have various theories, as well as laboratory or real-world examples of implemented subsystems, for such partial aspects as: targeting perception, motor control, inference, planning, affect, as well as many flavors of learning. However, a much smaller amount of research has targeted the further integration of these partial aspects. Even less work has been focused on cognitive architectures exhibiting a considerable degree of completeness and generality, which could potentially be applied to a wide variety of application domains, with minimal manual customization or redesign, and which could adapt themselves to changing environments and needs. At the same time, despite the big successes of artificial intelligence and embodied systems in some specific (and usually narrow) domains (Campbell, Hoane and Hsu 2002), the initial big promises as well as the estimated potential of the field have certainly not been fulfilled and reached yet. At the same time many new theoretical as well as practical tools have become available.

Thus, the review paper of Thórisson and Helgasson, appears at a crucial time for the further development of cognitive architectures, and fulfills an important need. Furthermore, it is built around a clear stance, which is applied consistently throughout its exposition. A very short summary of the basic stance of the paper could read like: "The goal of cognitive architectures is to act as a strong basic framework for creating systems with big generality, high autonomy, and strong capabilities for flexible adaptation and deep learning. An ideal example of what we mean by autonomy is a system that can (a) handle a large variety of environments – be it space, desert, ocean floors, and (b) fulfill its goals, which are given only in a high-level description without pre-encoded domain specific knowledge, while (c) operating in isolation. Furthermore, there are four main themes when designing or analyzing cognitive architectures: real-time operation, learning, resource management, and meta-learning. These four dimensions can be used as an analytical framework for describing existing approaches, and also for cross-comparing them, when we quantify their performance across these dimensions".

The authors indeed introduce their stance clearly, as well as follow it consistently throughout the paper; they analyze and cross-compare nine of the most prominent cognitive architectures of today, while providing interesting insights regarding promising avenues for the future. Having had extensive first-hand experience in designing large-scale systems that integrate multiple aspects of cognition, such as the conversational manipulator robot Ripley and its underlying "Grounded Situation Model" architecture (Mavridis and Roy 2006; Mavridis 2007), I am highly sympathetic as well as appreciative of the author's attempt. Furthermore, from the viewpoint of the wider circle of ideas of close proximity to my background, there are a number of interesting observations to be made that could juxtapose to and potentially enrich the review, and provide avenues for future extensions. In more detail:

1) **Autonomy, Isolation, and Collective Intelligence**

In the central stance of the paper which is summarized above, indeed (b) as well as (a) seem to be very good choices for the requirements for ideal autonomous systems. However, requirement (c), namely operation in isolation, needs further examination. Indeed, one can posit this as a requirement of autonomous systems: total self-sufficiency; a machine that, once created, could be left to even operate in a universe where no other intelligent entities exist. The question though follows: is this a good requirement to impose? How productive would that really be, and how relevant to real-world applications, with the exception of harsh environments with no life and big physical difficulties in communications, such as the outer space?

Human intelligence, seen through the lens of effective intelligence, i.e. the capacity for successful action towards self-selected goals, is very much enhanced due to the social nature of

humans. And it is not only collaborative teamwork that contributes towards this extension; social learning (Thomaz and Cakmak 2009), learner-directed or observation-based, as well as imitation (Nehaniv and Dautenhahn 2007) and en-culturation, are extremely important for the effective intelligence of humans. Examples of individuals having grown up in isolation provide a strong empirical basis for such an argument (Wikipedia 2012; Davis 1947). Furthermore, one can extend the argument much beyond physically and directly communicating humans. Through oral transmission and writing, the knowledge base of humanity is expanding: and past generations are contributing to our current capacity for effective action. Thus, this is yet another way through which collective intelligence is boosting individual intelligence. Furthermore, we are increasingly entering into frequent interactions with intelligent machines (Mavridis 2011a), which become part of our social networks. But how can this be relevant to cognitive architectures?

2) **Social Competencies and Environments**

To exhibit highly effective intelligence, a system implemented in a cognitive architecture should possess competencies for social interaction, learning and adaptation, which would enable it to utilize not only the physical affordances, but also the social affordances that are available in its environment. Thus, it should be able to participate in human-and-machine social networks (Mavridis 2011b), and position itself in relations with high social capital (Putnam 1993). Ideally, such a system should not only be acting towards its own direct short-term goals, but also contributing towards the increase of the resulting collective effective intelligence of the network that it is participating in. This could be achieved not only through participation in the interactions of the network, but also through facilitating structural rearrangements of the social network around it, for example in order to enrich collective social capital (Social Capital 2012). In that respect, for example part (a) of the basic stance of the authors could be extended to "a wide variety of environments, physical as well as socio-technico-cultural", and certainly (c) "operating in isolation" could be replaced with "operating in sustainable symbiotical interaction with the other intelligent entities that are accessible to it".

3) **Language, Non-verbal Communication, Situation Models, Theory-of-mind**

In order for a system to be able to exhibit the social competencies that were discussed above and participate in human-machine social networks, a basic prerequisite, among others, is to be able to support adequate human-machine as well as machine-machine interaction capabilities. Such interaction capabilities, which need to cover natural language as well as non-verbal communication for the human-machine case, necessitate the existence of real-world solutions to the symbol grounding problem (Harnad 1990), as well as situated and embodied language learning capabilities. Thus, one needs to go much beyond capacities for body and affordance discovery (Saegusa, Metta and Sandini 2010; Stoytchev 2005); and extend to conceptual alignment with other intelligent entities (Goldstone and Rogosky 2002), language learning and social affordance discovery, among others.

But then the question follows: what consequences do such competency requirements have, in terms of cognitive architectures? One possible suggestion here is that, in the same way that the demand for explicit attentional mechanisms arises, explicit situation modeling (Zwaan and Radvansky 1998) representations and standardized processes, such as those implemented in Mavridis and Roy (2006); Mavridis (2007), could be highly useful. Such situation modeling mechanisms can facilitate the bidirectional connection of natural language to the senses; the modeling not only of physical, but also of agentive aspects of the situation. Thus it should support theory-of-mind (Premack and Woodruff 1978) and self, as well as hetero-models; and also

directly extend to episodic memories (Mavridis and Petychakis 2010) and predictions. Of course, distinctions between different kinds of memory stores and knowledge bases exist in other cognitive architectures, but rarely there is explicit support for self- and hetero-modeling with theory-of-mind, with the exception of (Friedlander and Franklin 2008). Also, there is very rarely explicit support for the transition between natural language and symbolic representations, and for on-the-fly conceptual and situation-model alignment across agents, which would be highly valuable.

4) **Embodiment, Collective Intelligence and Offloading to Distributed Services**

Furthermore, as thousands of services are becoming available through the internet, in order for a system to take advantage of the full capabilities of the human-machine social network and extend beyond the limitations of its own physical embodiment and processing faculties, it should be able to interact with and utilize remote sensing, actuation, processing, and storage resources. Such services could either be provided by machines (such as those available on a computation or sensing cloud (Amherst, Fox, et al. 2010) or even by humans (such as the real time visual sensing and recognition services provided by humans in the DARPA Network Challenge (10 Red Balloons). Thus, for example, much before the limitations of computational power available by an onboard CPU are surpassed, such a system can harvest the much greater networked processing power of the cloud. Before tasks which are hard for AI but easy for humans become within reach of AI, such system can offload these tasks to networked humans which are part-time crowd-servicing (David 2011). In all those cases where a particular machine sensor or actuator – for example a camera – is not available but a human is, it can utilize the human services to achieve its goal, as in the 10 Red Balloons challenge, and effectively act as if the human sensing subsystems were temporarily part of its own embodiment. This is the idea advocated in the Human-Robot Cloud (Mavridis 2012), which enables the on-the-fly construction and reconstruction of distributed human-machine cognitive systems.

Taking the above four points into account, one resulting extension of the basic stance could be summarized as: “Real-time operation, Resource Management, Learning, and Meta-Learning – but beyond the limitations of the individual system: extending towards the total resources and total embodiment of the human-machine social network of which it is a part. Thus, the system is actually viewing the network as if the sensing, actuation, and processing services offered through it are part of its own body and resources. Starting from within the viewpoint of the individual entity, it needs extending to a more holistic consideration of the resources of the network of which it is part of, manages these resources, and does not only participate with real time considerations, but can also actively maintain and shape them. Also structurally – effectively performing network-wide learning – it can participate in a network-wide self-reflection and meta-learning mechanisms. In this way we extend the four main themes of the review paper from the individual intelligent entity (its own physical and informational resources) to the whole network, but always from within the viewing capabilities of the individual entity, as it pertains to its view of the human-machine network of which it is part of.

In summary: instead of connecting “autonomy” with a requirement of total self-sufficiency and capability of operation in isolation, it is much more reasonable to center our efforts towards positioning the intelligent entities created through cognitive architectures appropriately within human-machine social networks, externally offloading their physical and informational functions when needed, and harmoniously and empathetically integrating within ecosystems of intelligent entities, so that they can participate in much wider entities beyond themselves.

References

Armbrust, M.; Fox, A.; Griffith R; et al. 2010. A View of Cloud Computing. *Communications of ACM*. 53: 50-58.

Campbell, M.,; Hoane, A. J.; and Hsu, F. 2002. Deep Blue. *Artificial Intelligence*. 134: 57-83.

Davis, J. G. 2011. From Crowdsourcing to Crowdservicing. *IEEE Internet Computing*. 15: 92-94.

Davis, K. 1947. Final Note on A Case of Extreme Isolation. *American Journal of Sociology*. 52: 432-437.

Goldstone, R. L.; and Rogosky, B. J. 2002. Using Relations within Conceptual Systems to Translate across Conceptual Systems. *Cognition*. 84: 295-320.

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

Mavridis, N. 2007. *Grounded Situation Models for Situated Conversational Assistants*. Ph.D. diss., MIT.

Mavridis, N. 2011a Growing Robots in the Desert", TEDx Al Ain video available electronically at http://www.youtube.com/watch?v=HNSqQKZiofA

Mavridis, N. 2011b. Artificial Agents Entering Social Networks. In *A Networked Self*, Routledge *291-303*.

Mavridis, N.; Bourlai, T.; and Ognibene, D. 2012. The Human-Robot Cloud: Situated Collective Intelligence on Demand. In *Proceedings of IEEE International Conference on Cyber Technology in Automation, Control and Intelligent Systems*, 360-365.

Mavridis, N.; and Petychakis, M. 2010. Human-like Memory Systems for Interactive Robots: Desiderata and Two Case Studies Utilizing Grounded Situation Models and Online Social Networking. In *Proceedings of 36th Annual Convention of the Society for the Study of Artificial Intelligence and Simulation of Behaviour*.

Mavridis, N.; and Roy, D. 2006. Grounded Situation Models for Robots: Where Words and Percepts Meet. In *Proceedings of IEEE IROS*, 4694-4697.

Nehaniv, C. L.; and Dautenhahn, K. 2007. *Imitation and Social Learning in Robots, Humans and Animals*. Cambridge University Press.

Premack, D.; and Woodruff, G. 1978. Does the Chimpanzee Have a Theory of Mind? *Behavioral and Brain Sciences* 1: 515-526.

Putnam, R. D. 1993. The Prosperous Community: Social Capital and Public Life. *The American Prospect* 13: 35-42.

Saegusa, R.; Metta, G.; and Sandini, G. 2010. Self-Body Discovery Based on Visuomotor Coherence. In *Proceedings of 3rd Conference on Human System Interactions*, 356-362.

Stoytchev, A. 2005. Toward Learning the Binding Affordances of Objects: A Behavior-Grounded Approach. In *Proceedings of AAAI Symposium on Developmental Robotics*, 17-22.

Social Capital, 2012. Available electronically at: http://www.socialcapitalresearch.com/definition.html

Thomaz, A. L.; and Cakmak, M. 2009. Social Learning Mechanisms for Robots. In *Proceedings of International Symposium on Robotics Research* 1-14.

Wikipedia, 2012. Humans Growing in Isolation. Available electronically at: http://en.wikipedia.org/wiki/Feral_child

Zwaan, R. A.; and Radvansky, G. A. 1998. Situation Models in Language Comprehension and Memory. *Psychological Bulletin*. 123: 162-185.

10 Red Balloons, DARPA Network Challenge 2009. Available electronically at: http://en.wikipedia.org/wiki/DARPA_Network _Challenge