Hybrid Neural Symbolic Agent Architectures for Multimedia

Stefan Wermter
University of Sunderland
School of Computing and Information Systems
Sunderland SR6 0DD, United Kingdom
stefan.wermter@sunderland.ac.uk

1 Introduction

Recently there has been a lot of interest in adaptive symbolic and neural agents for different tasks, for instance speech/language integration and image/text integration in various multimedia applications [1, 7, 11, 10, 14, 2, 8, 4]. Hybrid neural symbolic methods have been shown to be able to reach a level where they can actually be further developed in real-world scenarios. A combination of symbolic and neural agents is possible in various neural symbolic processing architectures, which contain both symbolic and neural agents appropriate for to a specific task, e.g. integrating speech, text and images for multimedia.

In this paper we concentrate on general principles of neural and hybrid architectures for multimedia in general. From the perspective of knowledge engineering, hybrid symbolic/neural agents are advantageous since different mutually complementary properties can be combined. Symbolic representations have advantages with respect to easy interpretation, explicit control, fast initial coding, dynamic variable binding and knowledge abstraction. On the other hand, neural agents show advantages for gradual analog plausibility, learning, robust fault-tolerant processing, and generalization to similar input. Since these advantages are mutually complementary, a hybrid symbolic neural architecture can be useful if different processing strategies have to be supported.

2 Different types of neural architectures for multimedia

In figure 1 there is an overview of different possibilities for integration. Continuous connectionist representations are represented by a circle, discrete symbolic representations by a square. Symbolic interpretations of connectionist representations are shown as squares with dotted lines.

Connectionist structure architectures are the first type of symbolic/neural architectures. They rely solely on neural representations and symbolic knowledge arises by an interpretation process of the neural representations. Often specific knowledge of the task is built into the neural structure architecture. Much early work on structured connectionism can be traced back to early work by Feldman and Ballard, who provided a general framework of structured connectionism [5]. This framework was extended in many different directions, for instance for the so-called *NTL, neural theory of language* [6].

Hybrid transfer architectures transfer symbolic representations into neural representations or vice versa. Using a transfer architecture it is possible to insert or extract symbolic knowledge into or from a neural architecture. The main processing is performed by neural representations but there are automatic procedures for transferring from neural representations to symbolic representations or vice versa. Hybrid transfer architectures differ from neural structure architectures by the automatic transfer into and from symbolic representations. While certain units in neural structure architectures may be interpreted symbolically by an observer, only hybrid transfer architectures allow the knowledge transfer into rules, automata, grammars, etc. A good example is the work

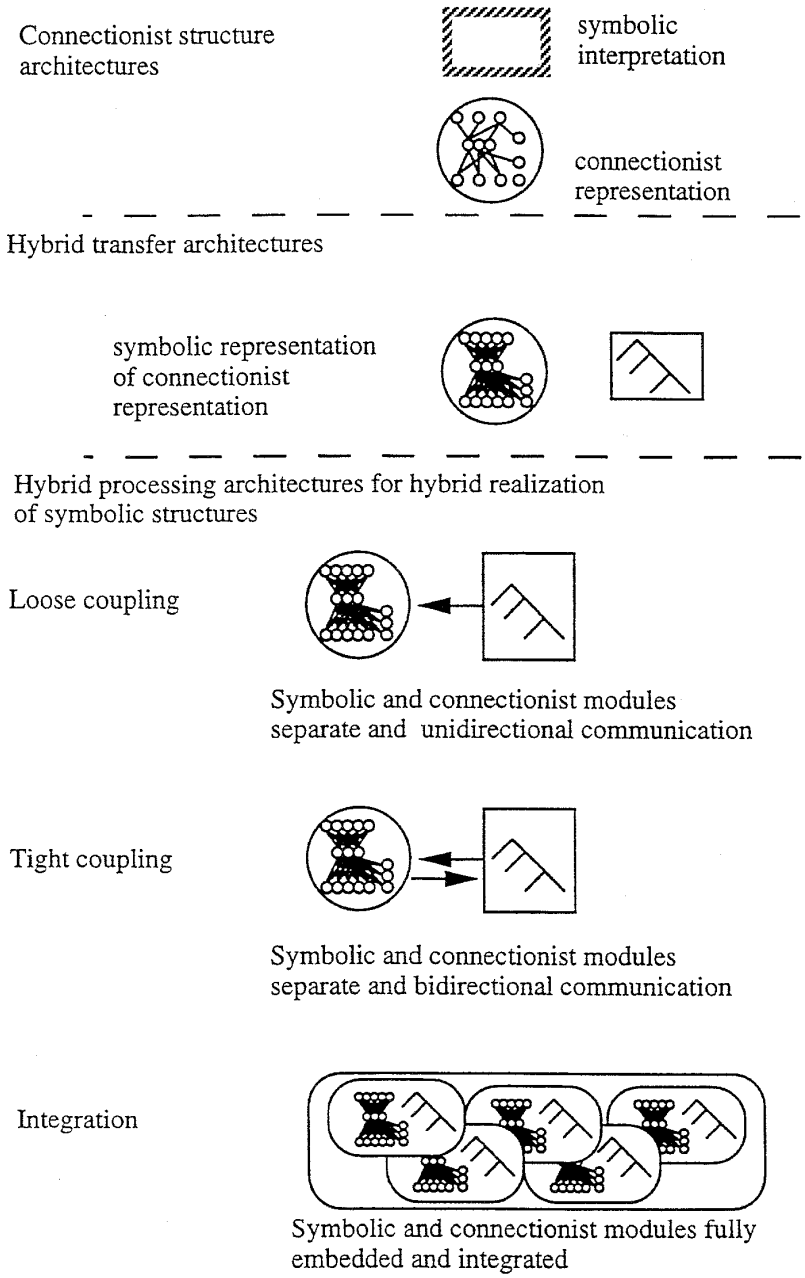

Figure 1: Overview about various types of symbolic/connectionist integration

on activation-based automata extraction from recurrent networks [9] or a weight-based transfer between symbolic rules and feedforward networks [12].

A *loosely coupled hybrid architecture* has separate symbolic and neural agents. The control flow is sequential in the sense that processing has to be finished in one agent before the next agent can begin. Only one agent is active at any time, and the communication between agents is unidirectional. An example architecture where the division of symbolic and neural work is loosely coupled has been described in a model for structural parsing within the SCAN framework [13].

A *tightly coupled hybrid architecture* contains separate symbolic and neural agents, and control and communication are via common internal data structures in each agent. The main difference between loosely and tightly coupled hybrid architectures is common data structures which allow a bidirectional exchange of knowledge between two or more agents.

In an *integrated hybrid architecture* there is no discernible external difference between symbolic and neural agents, since the agents have the same interface and they are embedded in the same architecture. The control flow may be parallel and the communication between symbolic and neural agents is via messages. Communication may be bidirectional between many agents, although not all possible communication channels have to be used.

3 Examples and future work in neural networks in multimedia

In this section we would like to point out two very different examples of hybrid architectures in order to illustrate different principles. DETE was a very challenging project of using only neural networks for the integration of visual/motoric knowledge with language knowledge [3]. DETE is restricted to a two-dimensional world of simple objects (circle, triangle, rectangle) which differ in color, size, and position. A moving "eye" can focus on or zoom into certain parts of this world triggered by motoric commands. A movable "finger" can point on certain objects. Together with the changing two-dimensional world DETE receives 1) a sequence of motor commands for controlling eye and finger and 2) a sequence of verbal input as a description. The task is either to generate a verbal output for a visual/motoric input or vice versa to generate a visual/motoric output for a verbal input. Such architectures have some potential for modeling the integration of language and vision in a monolithic interpretable framework but currently are restricted to relatively small domains and applications. While they can model context effects, short term memory integration and dynamic spatial processes it is not yet clear whether these architectures can scale up to real-world systems and applications for multimedia.

One example of an integrated hybrid architecture was developed for exploring integrated hybrid processing for speech and text analysis [15]. SCREEN is an integrated architecture since it does not crucially rely on a single neural or symbolic agent. Rather, there are many neural and symbolic agents, but they have a common interface, and they can communicate with each other in many directions. From an agent-external point of view it does not matter whether the internal processing within an agent is neural or symbolic. This architecture therefore exploits a full integration of symbolic and neural processing at the agent level. Furthermore, the agents can run in parallel and produce an analysis in an incremental manner.

Currently it seems coupled or integrated hybrid neural architectures are particularly useful for the complexity of integration and the required robustness. The demonstrated robustness and adaptive learning capabilities of neural networks makes them a particular promising candidate for processing and generating multimedia documents adaptively.

References

[1] Frederique Biennier, Michel Guivarch, and Jean-Marie Pinon. Browsing in hyperdocuments with the assistance of a neural network. In *Proceedings of the ECHT'90 European Conference on Hypertext*, Navigation and Browsing, pages 288–297, 1990.

[2] Hsinchun Chen, Marshall Ramsey, and Terry R. Smith. Visual SOM. In *Proceedings of the 20th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Posters, page 336, 1997.

[3] M. G. Dyer. Grounding language in perception. In V. Honavar and L. Uhr, editors, *Artificial Intelligence and Neural Networks: Steps towards principled Integration*, pages 455–482. Academic Press, San Diego, 1994.

[4] A. Faro, D. Giordano, and C. Santoro. Link-based shaping of hypermedia webs assisted by a neural agent. *J.UCS: Journal of Universal Computer Science*, 4(7):630–648, July 1998.

[5] J. A. Feldman and D. H. Ballard. Connectionist models and their properties. *Cognitive Science*, 6:205–254, 1982.

[6] J. A. Feldman, G. Lakoff, D. R. Bailey, S. Narayanan, T. Regier, and A. Stolcke. L_0 - the first five years of an automated language acquisition project. *AI Review*, 8, 1996.

[7] Alain Lelu and Claire Francois. Hypertext paradigm in the field of information retrieval: A neural approach. In *Proceedings of the Fourth ACM Conference on Hypertext*, Information Retrieval, pages 112–121, 1992.

[8] Y. Matsuyama. Multiple descent cost competition: Restorable self-organization and multimedia information processing. *IEEE Transactions on Neural Networks*, 9(1):106–122, January 1998.

[9] C. W. Omlin and C. L. Giles. Extraction of rules from discrete-time recurrent neural networks. *Neural Networks*, 9(1):41–52, 1996.

[10] A. Pedotti, G. Ferrigno, and M. Redolfi. Neural network in multimedia speech recognition. In E. C. Ifeachor and K. G. Rosen, editors, *Proceedings of the International Conference on Neural Networks and Expert Systems in Medicine and Healthcare*, pages 167–73, Plymouth, UK, 1994. Univ. Plymouth.

[11] S. Ramanathan and P. V. Rangan. Adaptive feedback techniques for synchronized multimedia retrieval over integrated networks. *IEEE/ACM Transactions on Networking*, 1(2):246–260, 1993.

[12] J. Shavlik. A framework for combining symbolic and neural learning. In V. Honavar and L. Uhr, editors, *Artificial Intelligence and Neural Networks: Steps towards principled Integration*, pages 561–580. Academic Press, San Diego, 1994.

[13] S. Wermter. *Hybrid Connectionist Natural Language Processing.* Chapman and Hall, Thomson International, London, UK, 1995.

[14] S. Wermter, E. Riloff, and G. Scheler. *Connectionist, Statistical and Symbolic Approaches to Learning for Natural Language Processing.* Springer, Berlin, 1996.

[15] S. Wermter and V. Weber. SCREEN: Learning a flat syntactic and semantic spoken language analysis using artificial neural networks. *Journal of Artificial Intelligence Research*, 6(1):35–85, 1997.