

# Special issue on integration of symbolic and connectionist systems

Paolo Frasconi<sup>a</sup>, Marco Gori<sup>b,\*</sup>, Franz Kurfess<sup>c</sup>, Alessandro Sperduti<sup>d</sup>

<sup>a</sup>University of Florence, Dipartimento di Sistemi e Informatica, via di Santa Marta 3, 50139 Firenze, Italy

<sup>b</sup>University of Siena, Dipartimento di Ingegneria dell'Informazione, via Roma 56, 53100 Siena, Italy

<sup>c</sup>California Polytechnic State University, Computer Science Department, San Luis Obispo, CA 93407, USA

<sup>d</sup>University of Pisa, Dipartimento di Informatica, Corso Italia 40, 56125 Pisa, Italy

---

The integration of symbolic and connectionist systems appears as a fundamental research topic for the development of intelligent and efficient systems capable of dealing with tasks whose nature is neither purely symbolic nor sub-symbolic. It is common opinion in the scientific community that quite a wide variety of real-world problems require hybrid solutions, i.e. solutions combining techniques based on neural networks, fuzzy logic, genetic algorithms, probabilistic networks, expert systems, and other symbol-based techniques.

In the last few years we have seen significant progresses in the theoretical understanding and the construction of neural systems capable of representing and processing structured knowledge in an adequate way, while maintaining essential capabilities of neural networks such as learning, tolerance of noise, treatment of inconsistencies, and parallel operation (see, e.g. Honovar & Uhr, 1994; Sun & Bookman, 1994). Giles et al. (1992) and, immediately afterwards, Towell and Shavlik (1993) stimulated a number of people towards the exploration of different approaches to inject and extract knowledge

into neural networks, in the attempt of constructing complex input/output maps to be refined later by learning from examples.

When comparing those hybrid systems with the symbolic counterpart, a clear difference which emerges immediately is that they were not adequate to process truly dynamical structured data. This was early recognized by Pollack (1990) who introduced the RAAM architecture for processing special trees with information located into the leaves. The research by Frasconi, Gori, and Sperduti (1998) is basically in the same direction. The emphasis is on the construction of appropriate schemes for learning in graphical domains, where symbolic and sub-symbolic information live together.

This special issue reports novel contributions on symbolic and connectionist systems focussing mainly on tight integrations. This is likely to open doors to interesting cognitive models, which can provide further insights on the emergence of symbols from distributed numerical representations.

Bruno Apolloni et al. present their view on how symbolically explained classification rules can be extracted from data. In their vision, data undergo a journey across a multi-layered architecture, starting from the soft processing by a neural network and ending in the hard symbolic world of symbolic

---

\*Corresponding author. Tel.: +39-057-723-3610; fax: +39-057-723-3602.

E-mail address: marco@dii.unisi.it (M. Gori).

formulas, where symbols represent (boolean) properties of data. The whole architecture is not static and the mapping from properties of data to symbols may change over time in search for a stable configuration. Key features of the approach are: (i) the use of entropic rules to guarantee the absence of information waste; (ii) an induction/deduction mechanism which combines usual learning algorithms at the sub-symbolic level, with an inferential procedure at symbolic level; (iii) the use of feedback (at different levels) to guarantee a fruitful interaction between the sub-symbolic and symbolic parts. However, the plasticity of the interaction is controlled via the optimization of a fitness function.

The paper by Giovanna Castellano, Anna Maria Fanelli, and Corrado Mencar (University of Bari, Italy) addresses the problem of human understandability of rules generated as a result of training neuro-fuzzy networks. After formally identifying some properties required for human understandability of a fuzzy knowledge base, it is argued that unconstrained learning approaches cannot guarantee that such properties will hold for the trained system. The neuro-fuzzy architecture introduced in the paper relies on a learning algorithm that maintains understandability by constraining the admissible values of tunable parameters. The effectiveness of the proposed method is demonstrated on two simple problems.

Barbara Hammer considers recent work in connectionist systems for processing of structured data. Structured data can naturally be processed by a symbolic system. However, there are some computational tasks which cannot be easily and naturally addressed by a symbolic system. For example, in chemical applications, it is often required to predict some physical or biological property of a chemical compound by just looking at its molecular structure. This type of property is naturally expressed as a real number and thus symbolic systems have problems in dealing with this kind of tasks, while a connectionist system is often able to reach a satisfactory solution. Different connectionist approaches to processing of structured data are reviewed. In particular, Barbara Hammer shows that while encoding of trees can be realized by small networks, the decoding function needs an a priori unlimited amount of resources. Moreover, she concludes that also for structured

domains it is possible to derive explicit sample bounds. However, these bounds, due to the nature of structured data, have to take in account structural features of the training data.

Although neural networks perform certain tasks admirably well, their human users sometimes have reservations about their usage in sensitive applications. This stems to a large degree from the low degree of comprehensibility, caused by a high degree of interdependence in the underlying processing model, and also often by the high input dimensionality of neural networks. Among several techniques used for the analysis of neural networks, rule extraction has the advantage of yielding results in a format that is more easily comprehended by humans. These approaches, however, usually do not analyze the parameters of the network directly, but rather try to formulate rules that result in the same input-output associations as the network, effectively using the network as an oracle. The work of Ofer Melnik and Jordan Pollack performs an analysis of neural networks based on first principals. From this, a direct representation of the network's function can be gathered, and an algorithm formulated that extracts the representation from the network. Since this Decision Intersection Boundary Algorithm extracts exact representations, rather than approximations, it allows a thorough analysis of learning and generalizations aspects. One of the major drawbacks of the approach at this point is the high complexity, restricting the size of networks that can be analyzed.

The paper by Jane Neumann (University of Edinburgh, UK) introduces two algorithms that learn holistic transformations for manipulating holographic reduced representations of compositional structures. Unlike the original approach proposed by Plate, these methods are based on machine learning and do not require constructing by hand the transformation vectors. Neumann's algorithm is based on gradient descent and fast-Fourier transform, and can learn in one-shot transformation vectors from a dataset of pairs of structure vectors. Interestingly, the system is shown to have a high degree of systematicity as it generalizes to structures of higher complexity than the training examples, and to structures containing novel elements.

One way of addressing the integration of symbol-oriented and sub-symbolic approaches to the treat-

ment of knowledge is the transformation of the information stored in the configuration of a neural network, most importantly the weights, into rules that can be used by symbol-oriented techniques. In their paper, D. Vogiatzis and A. Stafylopatis apply techniques from reinforcement learning to the extraction of rules from a labeled data set. Their translational approach tries to construct **if...then** rules that assign data into one or more predefined categories, and indicate through a percentage figure the portion of data that belong to that category. This percentage figure gives an indication of the quality of the rule as a whole, and in turn is used as a reinforcement signal in order to assess the quality of the components of that rule. Although reinforcement learning has mostly been applied to different types of problems, with moderate modifications the authors achieve consistently good results on a selection of test data sets including the Iris Data Set, 4 Gaussians, and the Image Segmentation data base.

Despite the association between artificial neural networks and their biological counterparts implied by the name, the computational aspects of artificial neural networks bear little resemblance to the processes found in real neurons. In their paper, Stefan Wermter and Christo Panchev from the University of Sunderland, UK, present a framework that links spiking neural networks, connectionist networks, and symbol-oriented computation. It uses as a starting point the observation that in the domain of language processing, sequential patterns and preferences are very essential methods to deal with the inherent complexity of the problem. Their framework utilizes preference machines, which transform sequential input preferences into sequential output preferences. These preference machines are related to finite-state machines; however, they do not only accept or reject input sequences, but can also produce output sequences. The preferences these machines deal with

express the degree to which certain features are present or absent in a sequence. The framework specifies preference machines at various levels of abstraction: At the lowest level, temporal processing is essential, basically capturing important aspects of the temporal encoding of information in spiking neural networks. At the highest level, symbolic structures, which typically have difficulty with graded preferences, are nevertheless very useful for the construction of computational models of cognitive processes. The layer in between, described through connectionist preference machines in the framework, establishes the link between the temporally oriented processing at the lower level, and the abstract representations at the higher level. In addition to the conceptual discussion of the framework, the authors apply it to a pulsed neural network for auditory processing.

## References

- Frasconi, P., Gori, M., & Sperduti, A. (1998). A general framework for adaptive processing of data structures. *IEEE Trans. Neural Networks*, 9(5), 768–787.
- Giles, C. L., Miller, C. B., Chen, D., Chen, H. H., Sun, G. Z., & Lee, Y. C. (1992). Learning and extracting finite state automata with second-order recurrent neural networks. *Neural Comput.*, 4(3), 393–405.
- Honovar, V., & Uhr, L. (Eds.), (1994). *Artificial intelligence and neural networks: steps toward principled integration*. John Wiley & Sons.
- Pollack, J. B. (1990). Recursive distributed representations. *Artific. Intell.*, 46(1–2), 77–106.
- Sun, R., & Bookman, L. (Eds.), (1994). *Computational architectures integrating neural and symbolic processes: a perspective on the state of the art*. Kluwer.
- Towell, G. G., & Shavlik, J. W. (1993). Extracting refined rules from knowledge-based neural networks. *Mach. Learn.*, 13, 71–101.