On the Conscious and Subconscious Components of Knowledge Representation in Neural Networks

Robert Kozma
Department of Information Science, University of Otago
P.O.Box 56, Dunedin, New Zealand
E-mail: rkozma@commerce.otago.ac.nz

Abstract

Principles of learning and knowledge representation are studied in complex neural networks with a large number of parameters. Our neural networks incorporate both deep and shallow knowledge representations. In the case of stationary environment, the neural nets can develop a deep understanding of the problem and working in a properly chosen, narrow subspace of the system variables. In this functional mode, a small number of parameters dominate the operation of the network on the surface (conscious components), while the overwhelming part of the network operates at a the subconscious level. The components realizing the conscious operational task are quite stable, nevertheless, they might change in time if the external environment of the analyzed system varies.

1 Introduction

Integration of methods of artificial intelligence (AI) and artificial neural networks (ANNs) became a hot topic of intelligent systems' design during the past few years. There is a growing respect within the AI community towards ANN models which can be used for knowledge generation and representation as well as for reasoning. In advanced applications, external knowledge regarding the system behavior can be incorporated into the ANN architecture. Artificially intelligent ANNs can handle, process and generate knowledge 'tidbits' (Bezdek, 1994). Knowledge monitoring contributes to a deeper understanding of the system behavior, e.g., by identifying certain causal relationships and by processing symbolic data. Nowadays, over 20% of the ANN research is conducted in the field of computational intelligence, i.e., it is associated with knowledge processing (Marks, 1996).

In a popular approach, fuzzy neural networks and hybrid neuro-fuzzy systems are designed, in which ANNs are used for approximating/tuning the membership functions and the rule-base of the fuzzy inference system; see Buckley & Hayashi (1995), Kasabov (1996). Hybrid systems have been applied successfully in a wide range of practical problems of system identification and on-line monitoring and adaptive control; see, e.g., White & Sofge (1992). The incomplete, noisy, and intermittent character of signals obtainable from the instrumentation of complex systems requires a robust monitoring method which, at the same time, must be sensitive to changes in the system state. The adaptability of hybrid systems is based on their neural network components.

In this paper, we introduce a knowledge representation paradigm, called local-global paradigm, which is a way of describing consciousness in the learning process both in biological and artificial systems. The idea is related to 'syncretism' described in Werbos (1992, 1993). Syncretism is based on the notion that a real-time system should have both fast learning and good generalization. As a possible solution, Werbos proposed a hierarchical network which combines a real-time supervised learning network acting as a long-term memory, and a more parsimonious network which is trained both in batch-mode and also in real time to match the changing memory.

A well-defined hierarchy of knowledge representation is a key component of AI approaches and it has been implemented in various neural network applications. Knowledge Based Artificial Neural Network (KBANN) is a hybrid system which combines AI and connectionist methods (Towell and Shavlik, 1994). In building an efficient knowledge representation system in vague domain, Sun (1995) proposed an architecture which includes a concept level and a micro-feature level. In the approach of Van de Merckt & Decaestecker (1995), the knowledge representation has also two levels. One level represents deep knowledge (recognition) while the other is a shallow one and it is related to comprehension. The above examples correspond to models, in
which different knowledge levels are incorporated into the design of the network, and the units at each level complete their tasks according to distinctive functional forms.

An alternative way of generating hierarchical knowledge monitoring in ANNs is structural learning. Structural learning can be either constructive or destructive. Destructive methods are also called pruning and they are quite popular in the literature. For an overview of the various methods, see Reed (1993) and Sankar & Mammane (1993). Examples of destructive methods include optimum brain damage, weight forgetting, lateral inhibition, and entropy learning, etc. (Le Cun et al., 1990; Yasui, 1994; Ishikawa, 1995).

A distinctive feature of various structural learning techniques is that the diverse functionality among the nodes is the result of an internal self-organization in the network. In this respect, structural learning is essentially different from the externally prescribed functional behavior which is typical to hierarchical knowledge-based systems. A key idea of the present paper is that the two aspects of learning can be integrated in a single neural network. The conscious and subconscious operational modes give way to one another as the relationship between the external environment and the internal state of the neural network changes. The practical implementation of this strategy is illustrated using the modified IRIS benchmark test.

2 Structural Learning Algorithm

In the present work, structural learning algorithm with forgetting is used which is modified backpropagation with a sum-of-weights penalty term. For details of the method see Ishikawa (1995), Kozma et al. (1996). Starting with random initialization, the magnitude of most of the weights will diminish. Nodes connected by diminishing weights are called decay nodes. They constitute the subconscious dimensions. A few weights will have magnitudes significantly deviating from zero, which connect the so-called surviving nodes (conscious level). As the training progresses, the neural network approaches a saturation level, in which both the survivor and the decayed weights vary only slightly. We start the training with a relatively large, fully connected network. We choose a large network because it gives the proper freedom for the learning algorithm to select the optimal structure. Previous studies indicate that structural learning with forgetting is a convenient tool to implement the finite accuracy of input-to-output mapping by ANNs (Kozma et al., 1996). By selecting the model parameters properly, the trained neural network will have a saturated structure which corresponds to the desired mapping accuracy. The main goal of the present paper is to show that a saturated network can adapt its structure to dynamic perturbations and that this adaptation is realized via a pulsing regime in which conscious and subconscious representation modes follow each other.

3 Consciousness and System Dimensionality

In this paper we introduce a structural learning algorithm which creates an information hierarchy in feedforward ANNs by projecting the actual system state to a subspace having usually much lower dimension than the original system. This subspace is called the conscious region while the rest of the hyperspace constitutes the subconscious dimensions. The projection of the system state is conducted by using the structural knowledge accumulated in the neural network. This knowledge can be expressed in the form of rules and causality relations and, therefore, it constitutes a high-level, intelligent function. It is to be emphasized that the generation and the adaptation of the NN structure take place automatically without external interference. This self-organizing feature under supervised learning conditions is an important advantage of the introduced method.

In the case of a stationary system, both the network structure and the observation volume remain essentially unchanged. This structure, however, is not frozen and it has some oscillations around its equilibrium. The subconscious dimensions also take part in these oscillations. As the system state changes, the observation perspective varies as well. At first, the conscious region is rapidly expanding; at the same time its knowledge content is continuously fading away. The extent of this expansion depends on the intensity and the speed of the actual changes. In the case of an essential external change, the expanded observation volume might include almost the whole hyperspace. For a weak perturbation, the expansion is limited to the direct neighborhood of the conscious region. At the second stage of the adaptation, the expansion slows down, stops and it turns into a contraction. Finally, a new structure is formed which exhibits small oscillations around its new equilibrium. In the framework of the present structural learning approach, an adaptive learning system is proposed which realizes syncretism at two levels.

1. At the first level, there is a neural net (ANN1) which is capable of adapting its structure in response to changes in the external environment, as it is described previously in this section. ANN1 has two basic operation modes. In the first mode, rules are extracted from the data; during this
mode the network operates in a high-dimensional regime. In the second mode, the operation is low-dimensional and ANN1 implements the rules discovered previously. The conscious/subconscious or local/global actions are embedded into a pulsed temporal dynamics of ANN1.

2. At the second level, a reference network (ANN2) is found. ANN2 is stationary over the time scales of the adaptation of ANN1. However, ANN2 is also adapted according to the following scheme. No adaptation is conducted if ANN1 can absorb the possible changes in the environment in one or several pulsed-mode operational step(s). If, however, ANN1 behaves anomalously the reference network has to be modified. In the context of this reference scheme, ANN2 completes a high-level action, while ANN1 conducts semi real-time adaptation.

There are various ways of implementing the ANN1/ANN2 relationship, but this question goes beyond the scope of the present study. An example of the implementation of short-time and long-time adaptation using structural learning in a neuro-fuzzy system is given by Kozma et al. (1995). In this study we concentrate on the description of the first-level ANN1 based on a structural learning algorithm.

4 Pulsing Mode of Conscious and Sub-conscious Operation of NNs

4.1 Levels of Knowledge Representation

Applications are given using Fisher’s IRIS data. The initial structure of the 3-layer, feed-forward NN used for the IRIS data is 4 input, 40 hidden-, and 3 output nodes: 4-40-3. IRIS data consists of fixed input patterns (50 patterns from each of the 3 classes), we have induced artificial deformations to the input data. It is well known that Class I (Setosa) is linearly separable from the other two classes (Versicolor and Virginica). The latter two classes, however, overlap and the classification problem is not a trivial one. The main goal of the present study is to evaluate the adaptiveness of the learned NN structure. Therefore, we introduce a deformation to the original data set for testing purposes. Namely, Class II is modified by doubling its coordinates in 1000 steps. Note that Class III and modified Class II are easily separable after the 1000-step long deformation process has been completed.

The weights can be divided into surviving and decayed ones. The magnitudes of the surviving nodes exceed that of the decayed nodes by two orders. The distribution of the weights at various stages of training is shown in Fig. 1. Figures 1a to 1c belong to training without forgetting. Although some structural evolution can be detected especially in Fig. 1c, the structural knowledge is very uncertain and it cannot be used for rule extraction. Figures 1d to 1f belong to the initial-, intermediate- and final stage of structural learning. After a random initialization, most of the weights and the hidden nodes decay and constitute the subconscious operational mode. The functioning of the NN is determined by the few surviving nodes (conscious operation).

4.2 Rule Extraction from the NN Structure

The structure of the NN is shown in Fig. 2a and 2b trained with standard BP and with structural learning, respectively. Figure 2a depicts a fully-connected ANN; solid and dashed lines denote positive and negative weights. The skeleton network obtained by forgetting is given in Fig. 2b. This network contains 4 active hidden nodes: H#2, H#3, H#18, and H#31. Based on the distribution of the connection weights, several rules can be inferred from the NN. It is easy to see that H#2, H#3, and H#31 play a similar role. In addition, input nodes T#1 and I#4 have an influence only on output node O#3. A possible set of rules writes as follows:

\[
\text{If } I\#2 \text{ is Large AND } I\#3 \text{ is Small} \\
\text{then Class is O#1 ELSE} \\
\text{If } I\#2 \text{ is Small AND } I\#3 \text{ is Large} \\
\text{then Class is O#2 ELSE}
\]
Structure of the neural network in the IRIS problem: (a) no forgetting, (b) forgetting with $\epsilon = 0.001$. Solid and dashed lines denote positive and negative weights, respectively.

If $I\#1$ is Small AND $I\#4$ is Large AND $I\#2$ is Small AND $I\#3$ is Large
then Class is $O\#3$

It is an interesting question how the extracted rules depend on the initial network size. This question, however, goes beyond the scope of the present work and will be described in a future publication.

4.2 On-line Adaptation of the ANN Structure

In real-life problems, the trained NN will inevitably face conditions that it has not experienced during the training. Neural nets which give reasonable responses in unexperienced situations are highly desirable in the practice. The dynamic variation of the size of the neural network is analyzed under changing input conditions. The network starts the learning from 100% utilization of its weights. The active size decreased quickly due to the applied structural learning and it saturates below 5% after 10,000 iterations. The network increases its size in response to the introduced changes. This increase, however, is delayed and it is clear only after 14,000 iterations. Structural rearrangement takes place in a serial way, i.e., the NN tries to activate different hidden nodes one-by-one. After 3 trials, the proper structure is found and the NN reaches a steady state. The adaptation in the case of the deformed IRIS problem is rather difficult. This is seen in Fig. 3, when the expansion of the dimension occasionally reaches 40 - 50% of the total weight space. The pulsing character of the search is clearly identifiable in Fig. 12 and it is associated with the search for the appropriate hidden nodes. Note that this search is performed by the network itself without external control. Finally, the network can identify the proper internal representation and the dimensionality of the NN contracts to its level before the deformation.

5 Conclusions

The local-global learning paradigm is implemented in the framework of a self-organizing knowledge processing system. The method uses structural leaning in neural networks with supervised training. The structural learning allows the reduction of the complexity of the problem by projecting the high-dimensional problem space into a low-dimensional subspace. This projection is accompanied by the generation of knowledge in the form of rules, causal relationships. Therefore, it is called the conscious operational mode, while the rest of the NN operates in a subconscious regime. The mechanism of the structural adaptation is analyzed in detail. In stationary situation, the magnitudes of the weights in the survivor group exceeds that of the decayed ones by several orders. As soon as changes occur in the environment of the ANN, i.e., the input patterns vary, dynamic interaction takes place between the decayed and survivor weights and nodes. The dimensionality of the neural network model increases. The level of the dimension expansion depends on the nature of the external perturbation and also on the skeletonization level of the neural net. When external conditions cease to vary, a new ANN structure is formed and the dimensionality reduces again. The performance of the introduced monitoring method can be improved further by incorporating fuzzy logic-based decision making in the framework of a multi-level learning scheme. Details of a neuro-fuzzy structural learning method will be introduced in forthcoming studies.
Acknowledgment:

This work is completed in the framework of the CBIS project funded by New Zealand Foundation for Research, Science and Technology, project No. FoRST/U00-606. Part of the studies introduced in this paper have been performed while the author stayed at Tohoku University, Japan. Continued support of Prof. Masaharu Kitamura is greatly appreciated.

References


