Keywords:
Subsumption Neural Networks Cognition Evolution Cortex Learning Memory NuTank Thinking

Artificial Cognition

Richard Keene
Dick.Keene@Central.SUN.COM

Sun Microsystems
5251 DTC Parkway
Englewood, CO 80111

ABSTRACT

This paper introduces a unified theory of how the brain works and proposes a method for duplicating such a system in a computer. The proposed system is a hybrid of a subsumptive neural architecture and a classical neural network. The system self-oscillates to cause what we call thought. The paper also presents a discussion on why the brain is designed as it is and how this line of reasoning produces the proposed architecture.

1 Introduction

One of the great goals of the science of Artificial Intelligence is to duplicate the cognitive processes of humans and other organisms. This paper presents a unified method for doing the above. This paper also presents a unified theory of the mind.

The purpose of natural cognitive systems appears to be to allow the organism to respond more rapidly and proactively to it's environment and thus survive. This is achieved in two ways. First there are reflexes, the built-in actions an organism will take to stimulus. The second is the prediction of what the future state of the immediate environment will be and thus avoid harmful situations. The greater the intelligence of a NC system, the further into the future it predicts. Humans often plan many years into the future.

The goals of Artificial Intelligence systems are often very different from the above goals.

1.1 Classification - UNCLASSIFIED

2 A proposed architecture

This paper proposes the combination of two types of neural architectures to create what is commonly called cognition.

2.0.1 Subsumptive Systems and Subsumptive Loops

The first architecture is a **subsumptive neural architecture**. In a subsumptive architecture lower level abstractions of input sensors are converted to more abstract representations of the environment, and high level reflexes are converted to specific output stimulations.

Figure 2.1 shows a subsumptive system. The high level reflexes are hardwired as responses to high level abstractions and inhibit or augment low level reflexes. These curved paths I have named **subsumptive loops**. Such a design allows a cognitive system to respond to it's environment in predetermined ways.

Such a system can be incrementally evolved since adding new layers of greater abstraction does not require changes to the lower level functions. For such a system to work correctly (except in the trivial case) it must be based on connections that can be made to disappear when not stimulated. Neurons are ideal for this since a neuron that is not pulsing appears to not exist. The requirement of **neural hiding** becomes evident when one actually implements such a system. One first designs simple low level reflexes and gets them to work. Then when one designs higher level abstractions and reflexes the need for neural hiding is obvious and becomes a powerful design feature. This ease of adding new reflexes has already been demonstrated with a program called NuTank.

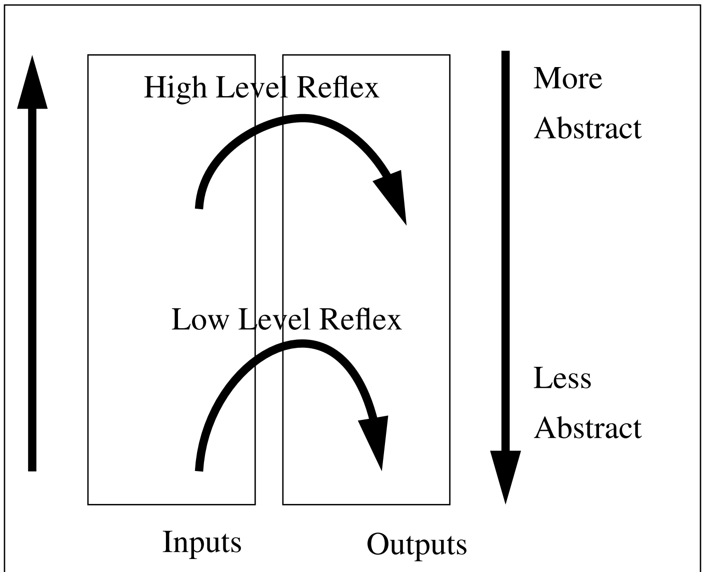

Figure 2.1. A Subsumptive System.

2.0.2 Regular Neural Networks

The second architecture of the system is a regular neural network that can learn. This part of the system is a traditional neural network with some learning function.

Each neuron in the regular neural network (Hereafter referred to as the **cortex**) has a target cell in the subsumptive system, who's behavior the teachable neuron is trying to predict. The cortex neurons have target cells that are mostly in the upper, more abstract, layers of the subsumptive system.

Each neuron in the cortex has a large number of inputs from both the subsumptive system and the cortex. Most connections to the inputs of a given cortex neuron come from nearby neurons. Some come from distant neurons. This **probability of connection** decreases with distance from the cortex neuron. It appears that a typical neuron in the human cortex has 20,000 connections. This set-up results in a traditional neural network but with an indeterminate number of hidden layers. In fact the layers of the cortex would not be hidden layers in the traditional sense.

A given cortex neuron tries to find a pattern in it's inputs and learn what state the target neuron is in when that pattern is matched. Then, at some future time, it stimulates (or inhibits) the target neuron when there is a pattern match. The amount of stimulation varies depending on how close of a match there is.

Figure 2.2 shows the resulting system and a magnified view of a single cortex neuron. Only a single cortex neuron and it's target are shown.

This model of "a cortex neuron with a subsumptive target" may or may not have an *obvious* analog in the brain. The model still is valid as one looks at the predicted effect such an architecture has. Perhaps the actual structure in the human brain is somewhat different, but the operation of the system should still be the same.

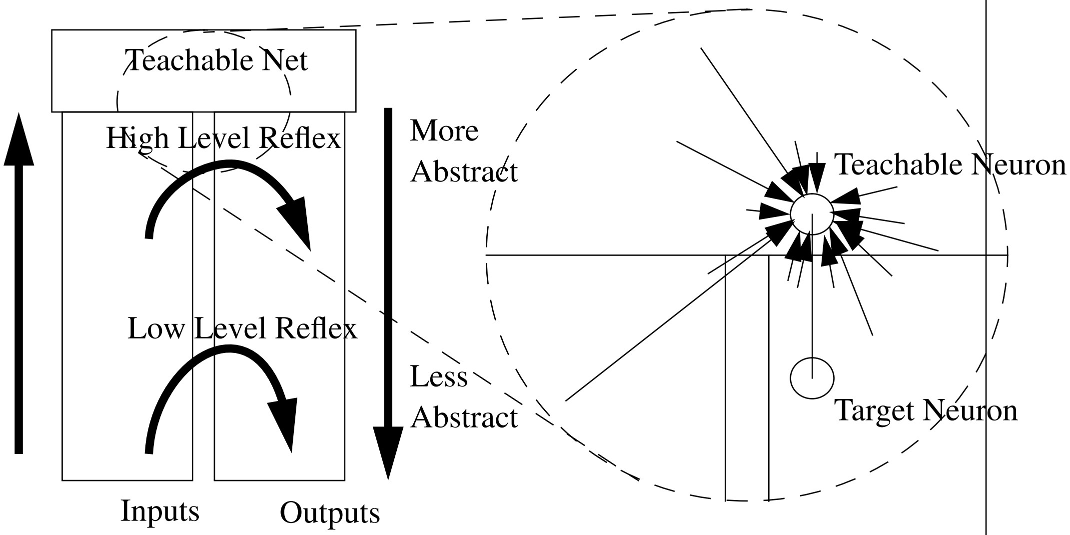

Figure 2.2, The Combined System and Cortex neuron

2.1 The Combined System Function

This combined system I will call a **subsumptive-regular system** or **SRS**.

This combined system of a subsumptive neural network and a regular neural network should have an interesting behavior.

The subsumptive system is constantly converting the raw inputs into abstractions of the environment and reacting to it. The cortex learns to match the abstract patterns and what the next state of the subsumptive system will be. If the inputs result in a recognizable pattern match, then the upper layers of the subsumptive system will be stimulated and inhibited into that next state even before the actual environment causes the next state. Now the subsumptive layers process that state, and the cortex does a pattern match on the new state. This cycle repeats.

In such a system there will be some fuzzy boundary between where the subsumptive system is processing and reacting to the actual environment, and higher levels that are reacting to the "imagined" state the cortex has matched. If there is a sudden and strong change in the environment then this boundary will move up. If the environment is uninteresting then this boundary will move down.

A more concrete example will help illustrate this. Imagine a rabbit sitting under a bush. The rabbit hears a twig snap, smells a fox, and sees a bit of orange fur. Moments later a fox attacks causing the rabbit to flee. (Lets also assume the rabbit gets away.) At some future time the rabbit smells a similar smell, hears a similar twig snap, but does not see red fur. The cortex matches this pattern and stimulates the lower regions of the rabbit's brain into a fleeing pattern (fear), before the fox attacks. This has now given the rabbit a very pro-survival behavior. The rabbit can now react to a predicted future.

In very complex systems, like the human brain, the chain of pattern match - subsumptive process - pattern match - subsumptive process... can get very involved. A person sitting thinking about some subject can loose track of actual environment and time for very long times. This can lead to very complex predictions of the future.

This design does not just increase the speed of pre-programmed reflexes. It results in a system that can setup synthetic situations, evaluate the outcome of the situations, and choose the best one.

2.1.1 Summary of the combined system

The central theory of this paper is that a subsumptive system that converts all the sensory inputs of the body into successively more abstract representation of the environment, and converts abstract commands into successively more exact motor outputs, when coupled with a constantly learning neural net results in a self-oscillating system that exhibits all the behavior of the brain.

2.1.2 The relationship to biological systems

Actual biological systems have several factors that hide such a simplistic architecture.

1.) Biological systems have evolved causing the brain to be made up of several SRSs nested one inside the other. The most obvious example of this is the two layer nature of the brain, the Cerebrum, and the Cerebellum. Also the subsumptive parts of biological brains have coevolved with the cortex to the point that the subsumptive system has few stand-alone high level reflexes. Thus if one removes the cortex (a lobotomy?) the organism becomes inactive. The conversion of inputs to abstract representations is most of what the subsumptive layers of the brain do. The conversion of high level commands to low level outputs is much less important.

2.) Biological systems must have redundancy in all systems to ensure correct operation with cell damage. It is also possible that the redundancy in the brain is necessary to achieve a fine degree of control, such that groups of neurons represent an abstraction instead of a single neuron. An SRS design with redundant neurons would degrade with damage but not fail catastrophically.

3.) The brain does not operate in one synchronous cycle. All parts are free running asynchronously.

2.1.3 Speed of Learning

Such a system has another interesting feature related to learning rates. In a system with 100 million neurons in the cortex the learning rate can be very low and still result in immediate learning. If 1% of the 100 million neurons change by 0.01% when a significant external event occurs, this is equivalent to 100 neural connections changing by 100%. This is a significant change. Such a distributed change would result in immediate pattern matches on similar situations. Repeated similar stimulus would result in a large permanent change.

Such a change probably consists of a slight chemical change that is later followed by slight cell growth. As the chemical change fades the cell growth would make up for it and result in sustained memory over a time range from seconds to years. This change might be seen as a change in the expression of proteins in the cells.

Memory in the brain probably consists of two kinds of memory. First there is the previously described changes in neural connections. The second type would be “ringing” in the subsumptive system. This ringing would be due to neurons in the subsumptive system with fairly long time constants. (By time constant I mean that the neuron can integrate over fairly long time periods with out loosing track of previous pulses.)

2.1.4 Memory in the SRS

Memory is stored globally in the system. With computer one recalls memory by moving a pattern from storage to the central processor. In a SRS recalling a memory means that subsumptive system is thrown into some state that is very similar to some previous state. When one recalls a past experience the following chain of events occurs. First something in the present gets some part of the subsumptive system into a similar state as it was at the previous time. The cortex then recognizes this state and stimulates the rest of the subsumptive system to match. For example:

1.) You read the word “Rose” written here.

2.) The auditory processing sections of the brain are put into the state caused by the word rose.

3.) The cortex matches that pattern to a time when you had a rose in your hand. (The first match has occurred.)

4.) Now the subsumptive parts of the brain related to auditory and visual processing are in the states associated with experiencing a rose.

5.) Another match occurs, from when you once smelled a rose. (The second match)

6.) Now the auditory, visual and olfactory processing systems are in the state associated with roses.

7.) The process continues oscillating until some external stimulus overrides the thought train.

2.2 Proof of the Architecture

Such a SRS has several proofs.

The best proof of this design would be to build a system that is sufficiently complex to talk about it's self. This may seem impossible, but there are several factors that bring this into the realm of reality.

1.) The brain creates what we call thought using neurons, therefore it can be done.

2.) A artificial system need not survive damage. It needs no redundancy. This may result in a 10 times reduction in complexity.

3.) A robot that the system connects to need only have a single arm, single eye, single ear, and a synthesizer output. This cuts the complexity by at least half.

4.) The ability to incrementally evolve the system means much can be designed now and as the system grows the computer technology will grow with it.

5.) The system need not reproduce. This eliminates all the complex behavior related to reproduction.