Informeta, L.L.C.

Markov Chains and Mentys

R. Coleman, PhD.
ron.coleman@informeta.net

This document is a comparative description of Mentys which uses Markov chains. However to put this approach into perspective, we contrast Markov chains with artificial neural networks. We have deliberately left out mathematical formulae and pseudo-code in an effort to enhance the clarity of the points we hope to make.

Artificial neural networks

As a working definition, artificial intelligence is a large body of work whose aims are to automate tasks believed to require human or "natural" intelligence. The distinction between A.I. and statistics, in particular statistical applications called management science, is very narrow in places. Nevertheless, a very popular approach to A.I. that has a close, if not competing, relationship with statistics is artificial neural networks.

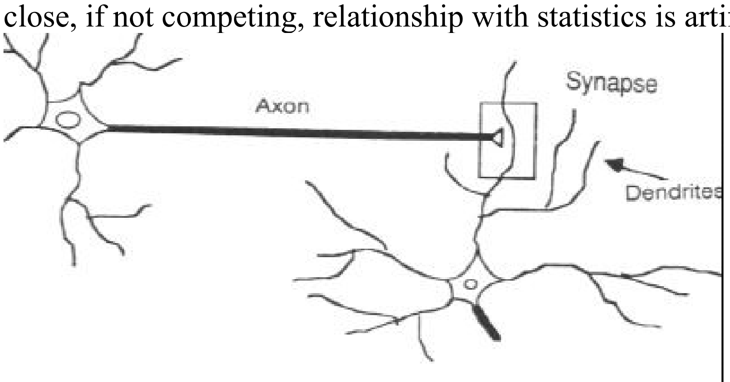

Figure 1. A biological neural network.

ANNs attempt to model in computer software networks of neurons in the brain. Like their biologically-inspired counterparts, ANNs learn from experience in a supervised or unsupervised manner. In the supervised case, the net is presented with data stimuli and responses and the training goal is to recall the correct response given a specific stimulus. In the unsupervised case, the training goal is to discover patterns of relevant associations. In both cases, during training the ANN learns by adjusting the strength or weight of connections in the "synaptic" junction that relate neuron functions for processing input data.

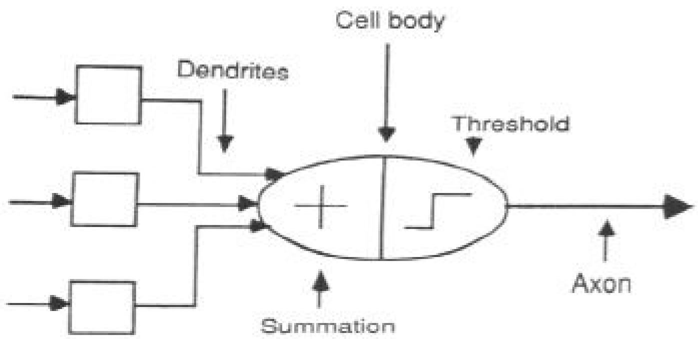

Figure 2. Computer model of a neural network.

Although ANNs had been under study since at least the late 40s, they found many interesting and novel pattern recognition applications only in the 80s when affordable and accessible computer processing made implementations feasible by academicians and independent researchers.

The key technical points to note are:

(1) ANN structures are gross simplifications of real nets because the natural ones work in massive parallelism and furthermore, brains have a hundred trillion or more neurons with extremely dense connections which computers are presently unable to model.
(2) ANN structures are mathematical functions that not correspond directly to any feature or variable in the real world. For instance, you would not find a neuron for INTC in a neural net. Thus, what any one neuron is doing is usually meaningless compared to what INTC is doing.
(3) Training ANN is usually very involved and difficult.
(4) By their nature, when presented with input they have not been trained on, ANNs may be unpredictable or give senseless answers, the invalidity of which may be difficult to verify.

Markov chains

Mentys employs a radically different approach to pattern recognition based on Markov chains. Firstly, the components of a Markov chain are observable features of the real world. For instance, in a Markov chain there is a node called INTC which represents facts about Intel share prices. Markov chain features are also assumed to change in discrete states, namely, jumps, gaps, or spikes, not smoothly as in ANN. One can see from this difference why Markov chains might more naturally represent some types of financial transactions in the real world

The past and present states are discrete forms of knowledge Markov chains capture. The future is just another state beyond the present. For instance, we can ask what the current state of INTC is: Is the price low? What does the current state have to say about the next state?

A collection of these states makes a knowledge base. This is another idea that has no meaning for ANNs since the synaptic weights relate only to neurons that don't correspond to anything in the world.

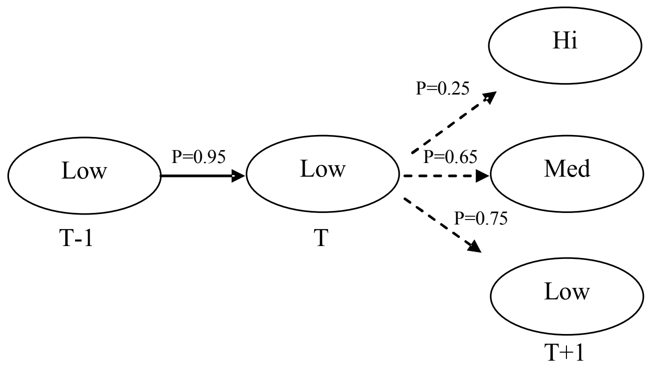

Figure 3. Markov chain of INTC.

The above figure represents how the state of INTC is changing in time. Notice the knowledge base does not allow transitions between states. In other words, in this example, there is no such thing as medium high. INTC is allowed to jump between low and medium or even low and high. When the path to the next state at T+1 depends only on the current state at T, the Markov chain is said to be *memoryless*. How the chain reaches the current state (T) is not important; only the probability of being in the current state and where it will probably go from the current state are important.

In the above example, indeed, the knowledge base allows for three possibilities but they are not equally likely. Mentys always tracks the most likely path and this process gives rise to the reliability estimate—another feature not supported by ANNs. Recall when an ANN identifies a pattern, the plausibility is unspecified which means there's no way to know how good the pattern is.

Training the Markov chain implies creating a knowledge base from historical cases. In effect, we simply discretize the world into states. (The statistical basis for this discretization is founded on simple thermodynamic principles of cluster analysis and it generally does not require input from the user.) Depending on the size and complexity of the data, this process can be extremely fast, even for tens of thousands of features. In any event, this level of complexity and efficiency is well outside the usual range of ANNs by two or three orders of magnitude.

Inferences, that is, identifying a pattern in ANN means supplying an input, running through the mathematical formulae to compute an association. Inferences using Markov chains involves heuristic search. Specifically, we look up the state of the world that best represents the current state and from that point, select a new state that represents the most probable transition. This two-step search conforms precisely to Bayesian inference: namely, the posterior (i.e., the reliability estimate) is proportional to the prior (step 1) times the likelihood (step 2).

Finally, since the Markov chain corresponds to observable happenings in the real world, it is possible to apply prior knowledge in the form of constraints from experts. These constraints redirect or deflect the (step 1) search into states believed from experience or commonsense to be more plausible. In the INTC example, prior knowledge may deliberately rule out transitions from low at T-1 to hi at T because the last price change must always be between bid and ask prices.

The history of Markov chains is very rich for solving simulation-type problems going back to the 50s. A special case, hidden Markov models, are used today in various applications including speech recognition, natural language processing, genomics—curiously the same areas where ANNs have been applied. The use of Markov chains as found in Mentys to identify anomalies, repair data, and assess data integrity, to our knowledge, is unique.

The key technical points to note are:

(1) The knowledge base is represents to the state of the world.
(2) While updating the knowledgebase is fast and efficient, the knowledgebase. On the one hand this is a limitation. On the other hand it errs on the conservative side and won't yield impossible results.
(3) The reliability estimate always provides a measure of how good the prediction is.
(4) The incorporation of prior knowledge
(5) We know of no comparable approach which can handle of thousands of features commonly found in financial type data.