An Information Theoretic Approach to Modeling Neural Network Expert Systems

Rodney M. Goodman and John W. Miller
Department of Electrical Engineering
California Institute of Technology 116-81

Padhraic Smyth
Communication Systems Research
Jet Propulsion Laboratories 238-420

In this paper we propose several novel techniques for mapping rule bases, such as are used in rule based expert systems, onto neural network architectures. Our objective in doing this is to achieve a system capable of incremental learning, and distributed probabilistic inference. Such a system would be capable of performing inference many orders of magnitude faster than current serial rule based expert systems, and hence be capable of true real time operation. In addition, the rule based formalism gives the system an explicit knowledge representation, unlike current neural models. We propose an information-theoretic approach to this problem, which really has two aspects: firstly learning the model and, secondly, performing inference using this model. We will show a clear pathway to implementing an expert system starting from raw data, via a learned rule-based model, to a neural network that performs distributed inference.

Consider first the learning problem. We have recently developed an information-theoretic algorithm called ITRULE which extracts an optimal set of probabilistic rules from a given data set [1,2]. We use the following simple model of a rule, i.e., If $Y = y$ then $X = x$ with probability p where **X** and **Y** are two attributes (random variables) with "x" and "y" being values in their respective discrete alphabets. We then define the information which the event y yields about the variable X as

$$j(X;y) = p(x|y)\, log \left(\frac{p(x|y)}{p(x)}\right) + p(\bar{x}|y)\, log \left(\frac{p(\bar{x}|y)}{p(\bar{x})}\right)$$

Recently we have shown that $j(X;y)$ possesses unique properties as a rule information measure [3]. Having learned the rules, we then map these onto the connections of a neural network, using several different information theoretic metrics as the weights. Essentially these connections form a set of lower order constraints on the N-th order joint distribution in the form of probabilistic rules. This is our a priori model. In a typical inference situation we are given some initial conditions (i.e., some nodes are clamped), we are allowed to measure the state of some other nodes (possibly at a cost), and we wish to infer the state or probability of one more goal propositions or nodes form the available evidence. This is the inference problem, determining an a posteriori distribution in the face of incomplete and uncertain information. Our goal then is to perform an approximation to exact Bayesian inference in a robust manner. With this in mind we have developed two particular information-theoretic interpretations of the inference process, which we describe as the hypothesis testing network and the uncertainty network[4]. The former scheme views the neurons in the network as performing a form of distributed Neyman-Pearson hypothesis test based on local information. The uncertainty network views each neuron as measuring the uncertainty of all its inputs and then updating its own activation to correspond to this uncertainty. In both cases a neuron can output an estimate of the posterior probability of node x by recovering $\hat{p}(x)$ from an inverse transformation of the activation function.

References

1. R.M. Goodman and P. Smyth, "An information theoretic model for rule-based expert systems," presented at the 1988 International Symposium on Information Theory, Kobe, Japan.
2. R.M. Goodman and P. Smyth, "Information theoretic rule induction," *Proceedings of the 1988 European Conference on AI*, Pitman Publishing: London.
3. P. Smyth and R.M. Goodman, "The information content of a probabilistic rule," submitted for publication.
4. R.M. Goodman, J.W. Miller, and P. Smyth, "An information-theoretic approach to rule-based connectionist expert systems," presented at the 1988 Neural Information Processing Conference, Denver, 1988: also to appear in *Advances in Neural Information Processing*, Morgan Kaufmann, April 1989 (in press).