A New Mixture Model for Concept Learning From Time Series

William H. Hsu and Sylvian R. Ray

Department of Computer Science, University of Illinois at Urbana-Champaign
1304 West Springfield Avenue, Urbana, IL 61801
{bhsu | ray}@cs.uiuc.edu http://anncbt.ai.uiuc.edu

Specialist-Moderator Networks

Mixture-of-experts models, or mixture models, are a divide-and-conquer learning method derived from the mixture estimation paradigm [DH73] that is heavily studied in artificial neural network research [JJ94]. They reduce complexity by decomposing learning tasks and variance by combining multiple classifiers. Recent research has shown how inductive learning algorithms can be augmented by *aggregation mixtures* such as bootstrap aggregation (or bagging) [Br96] and stacked generalization [Wo92], and by *partitioning mixtures* such as *boosting* [FS96] and *hierarchical mixtures of experts* (or *HME*) [JJ94].

Figure 1 depicts a new hierarchical mixture model called a *specialist-moderator (S-M)* network, which combines classifiers in a bottom-up fashion. Its primary novel contribution is an ability to learn using a hierarchy of inductive generalizers (components) while utilizing *differences among input and output attributes* in each component. These differences allow our network to form *intermediate targets* based on the learning targets of its components, yielding greater resolution capability and higher classification accuracy than a comparable non-modular network. In time series learning, this typically means reduced localization error, such as in multimodal sensor integration [RH98, HR98]. Each component (box) in Figure 1 denotes a self-contained statistical learning model such as a multilayer perceptron, decision tree, or Bayesian network. We choose to experiment with artificial neural networks (ANNs) because our target application is time series classification, and ANNs readily admit extension to time series [El90, PL98]. The terms *specialist network* or *moderator network* may denote arbitrary learning models in the overall "network" (a tree of components), but are assumed to be ANNs here.

An S-M network is constructed from a specification of input and output attributes for each of several modules (the leaves of the network). Training data and test input will be presented to these "specialists" according to this specification. The construction algorithm simply generates new input-output specifications for *moderator* networks. The target output classes of each parent are the Cartesian product (denoted $\times$) of its childrens', and the childrens' outputs *and* the concatenation of their inputs (denoted o) are given as input to the parent.

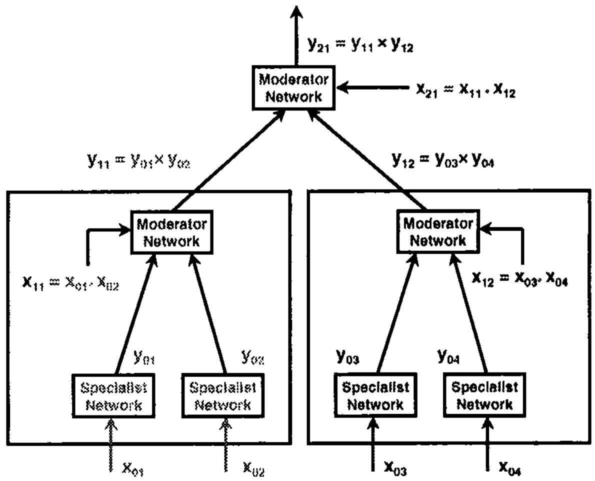

Figure 1. A Specialist-Moderator network

One significant benefit of this abstraction approach is that it exploits factorial structure (i.e., the ability of high-level or abstract learning targets to be factored) in decomposable learning tasks. This results in a reduction in network complexity compared to non-modular or non-hierarchical methods, *whenever this structure can be identified* (using prior knowledge, or more interestingly, through clustering or vector quantization methods). In addition, the bottom-up construction supports natural grouping of input attributes based on *modalities* of perception (e.g., the data *channels* or observable attributes available to each "specialist" via a particular sensor). Finally, we demonstrate that the achievable test error on decomposable time series learned using a specialist-moderator network is lower than that for non-modular feedforward or temporal ANN (given limits on complexity and training time).

Time Series Learning Using Recurrent ANNs

In our experiments, we focused solely on *classification* of time series. Our architecture addresses one of the key shortcomings of many current approaches to time series learning: the need for an explicit, formal model of inputs from different modalities. For example, the specialists at each leaf in our network might represent audio and infrared sensors in a industrial or military monitoring system [RH98, HR98].

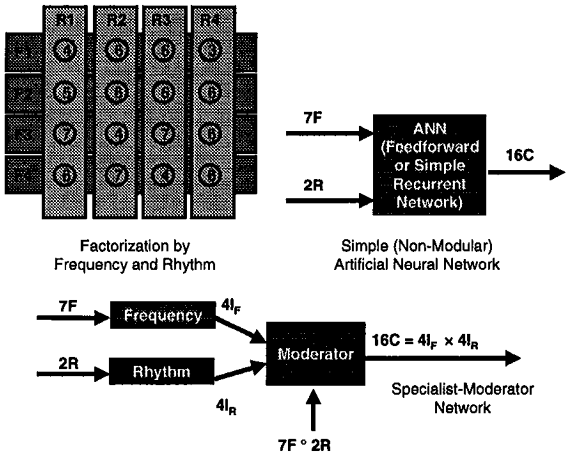

Figure 2. Musical tune classification problem

Figure 2 depicts non-modular and specialist-moderator architectures for learning a musical tune classification database with 89 tunes and 16 target classes. The non-modular network receives 9 channels of input and is trained using a *locally coded* target [KJ97] for the pre-labeled tunes. The first-level (leaf) networks in the specialist-moderator network receive *specialized* inputs: the frequency component only or the rhythm component only. The principle is that only the frequency component is *relevant* to the frequency specialist, and similarly for rhythm. The targets are intermediate attributes I_F and I_R. We used competitive clustering by Gaussian radial-basis functions (RBFs) to demonstrate that I_F and I_R could be formed, by unsupervised learning, for a 4-by-4 factorization, among others [RH98].

Comparison With Non-Modular ANNs

Table 1 shows the performance of the non-modular (simple feedforward, or *FF*, and input recurrent, or *IR* [PL98]) ANNs compared to their specialist-moderator counterparts. The italicized networks have 16 targets; the specialists, 4 each. Prediction accuracy is measured by the number of individual exemplars classified. The results illustrate that input recurrent networks (simple, specialist, and moderator) are more capable of generalizing over the temporally coded music data than are feedforward ANNs.

Network Type	Accuracy	
	Training	Cross Validation
FF, Simple	*344/589 (58.40%)*	*67/128 (52.44%)*
FF, Rhythm	534/589 (90.66%)	104/128 (81.25%)
FF, Frequency	589/589 (100.0%)	128/128 (100.0%)
FF, Moderator	*441/589 (74.87%)*	*77/128 (60.16%)*
IR, *Simple*	*566/589 (96.10%)*	*83/128 (64.84%)*
IR, Rhythm	565/589 (95.93%)	107/128 (83.59%)
IR, Frequency	589/589 (100.0%)	128/128 (100.0%)
IR, *Moderator*	*589/589 (100.0%)*	*104/128 (81.25%)*

Table 1. Modular versus non-modular networks

Comparison with HME

As Table 2 shows, an HME network with 8 leaves outperforms one with 4 and is comparable to the specialist-moderator network of feedforward networks. It is, however, outperformed by the specialist-moderator network of input recurrent networks. This is significant because incorporating recurrence into HME requires nontrivial modifications to the algorithm.

Design	Accuracy	
	Training	[illegible]
HME, 4 leaves	387/589 (65.71%)	58/128 (45.31%)
HME, 8 leaves	468/589 (79.46%)	77/128 (60.16%)
S-M net, FF	441/589 (74.87%)	77/128 (60.16%)
S-M net, IR	589/589 (100.0%)	104/128 (81.25%)

Table 2. Specialist-Moderator network versus HME

Conclusions and Future Work

We have presented an algorithm for combining data from multiple input sources (sensors, specialists with different concentrations, etc.) and a modular, recurrent, artificial neural network for time series learning. Fusion of time series classifiers showcases the strengths of our mixture model because there are many *preprocessing methods* that produce reformulated input. Typical applications are process monitoring, prediction, and control [HR98].

References

[Br96] L. Breiman. Bagging Predictors. *Machine Learning*, 1996.

[DH73] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis.* Wiley, New York, NY, 1973.

[El90] J. L. Elman. Finding Structure in Time. *Cognitive Science*, 14:179-211, 1990.

[FS96] T. Freund and R. E. Schapire. Experiments with a New Boosting Algorithm. In *Proceedings of ICML-96.*

[Hs97] W. H. Hsu. *Spatiotemporal Sequence Learning With Probabilistic Networks.* Unpublished, URL: http://anncbt.ai.uiuc.edu/prelim.doc, 1997.

[HR98] W. H. Hsu and S. R. Ray. A New Mixture Model for Concept Learning From Time Series. Extended abstract, http://anncbt.ai.uiuc.edu/hr98a.ps, 1998.

[JJ94] M. I. Jordan and R. A. Jacobs. Hierarchical Mixtures of Experts and the EM Algorithm. *Neural Computation*, 6:181-214, 1994.

[KJ97] R. Kohavi and G. H. John. Wrappers for Feature Subset Selection. *Artificial Intelligence, Special Issue on Relevance, 97*(1-2):273-324, 1997.

[PL98] J. Principé, C. Lefebvre. *NeuroSolutions v3.01*, NeuroDimension, Gainesville, FL, 1998.

[RH98] S. R. Ray and W. H. Hsu. Modular Classification for Spatiotemporal Sequence Processing. Submitted to *Journal of Intelligent Data Analysis.*

[Wo92] D. H. Wolpert. Stacked Generalization. *Neural Networks*, 5:241-259, 1992.