Comparison of raw state space and coarse grained space separation in coupled map classifier networks

David DeMaris
IBM Microelectronics
demaris@us.ibm.com

Extended Abstract

Several column-like spatiotemporal computing models have been investigated in the last decade and shown to achieve some classification capability without requiring task specific structural microcircuit designs. The liquid state machine (LSM) framework has been introduced recently, with claims that no coding or representation is needed during real time computing; high dimensional projects of the input are read out in lower dimensions by task specific readout units (Maas, Natschlager et al. 2002). Only requirements for separation and approximation must be met for the circuits to function. It was shown that when the number of and average length of random connections in the LSM model are too broadly distributed the spiking dynamics become chaotic, and separation and classification performance is degraded.

Another column-like model, locally coupled map lattices, has been explored as a similar column-like computation framework in which the constituent units represent mixed excitatory inhibitory recurrent neuronal pools (Kaneko 1990; DeMaris and Womack 2001; Freeman, Kozma et al. 2001). These units are intrinsically chaotic (with periodic windows) under the allowed parameter regimes when uncoupled. However, coupling to other units can reduce the effective dimension and eventually suppress chaos via synchronization phenomena, and readout can take place with a reduced dimensionality via the formalism of coarse graining. Coarse grained readout, in neural terms, is interpreted as readout units or assemblies with limited precision response to specific rate distributions integrated over a short time window. Physiological and modeling work has shown that spike timing precision in ensembles can be controlled by fluctuations in noise level of inputs to neural assemblies (Aertsen and Preissl 1991).

The *time series* view of chaotic systems suggests that excessive separation will occur leading to non-robustness in the face of noise.. The CML systems performing computations may exploit the potentially *rapid statistical convergence* properties that come with larger homogenous parameter spatially extended systems (Milton and Mackey 2000; DeMaris 2001), similar to the way that large ensembles of initial conditions presented to a single map or uncoupled set rapidly reach an invariant density (Driebe 1999).

To date I am aware of no systematic studies on the separation or convergence of coarse grained statistics from similar input, similar to those performed by the LSM group.

This paper reports preliminary results of such a study, in the context of understanding differences in performance in a pattern recognition task involving objects rotated in depth.

Separability studies in the liquid state formalism applied input spike trains u and v with known L2 norm distance. The measured L2 norm distances summed over time are compared to a baseline separation derived from applying the same spike train with a different random initial state. In the present author's CML framework, input is applied as an initial condition and iterated for up to 16 cycles; we interpret this as a sample and iterate computational style.

Since readout in the LSM system may occur continually from the time of stimulus application, it is appropriate to measure separability over time. In the CML system readout occurs at a single time step after one or two parameter epochs; the readout time is a learning parameter, and bifurcation and coupling dynamical parameters affecting partial synchronization, synchronization rates and state flows are learned in conjunction with this readout time. Thus the comparison is inexact. With this caveat on methodology differences, some preliminary findings are presented, with distances computed at the readout time of the CML.

Recognition is achieved in this system by applying a target object as an initial condition to an ensemble of classifiers, each of which learns to map different training views of an object near the same point in a space formed by partition cells sampled at readout time after a dynamical evolution. Each such application is termed a *recognition trial*. A *learning trial* consists of a set of target views applied to the classifier ensembles generated by a particular run of the genetic learning, where the history and population encoding of each classifier influences subsequent encodings.

Some idea of the separability of inputs can be obtained by examining the difference between states generated by all classifiers during each recognition trial. The minimum, median, and maximum distances between the output of each classifier in each recognition trial were measured and summed across all presented object views in each such learning trial, and correlations between these measures and recognition rate were computed. These distances and measures were performed for two different spaces – the coarse grained occupancies actually used for recognition, and the "raw" state vectors with spatial units corresponding across classifiers. The computations were performed on a set of 14 learning trials, representing 1960 recognition trials.

The first observation is that no strong correlation ($r>.5$) of recognition rate to any of these measures is found. Since all of the classifiers were trained to generate coarse grained state distributions for a particular object's training views which differed from distributions generated for other objects, and performance varies within a relatively small range (70-84%), this may not be too surprising.

The finding of most interest is that the sign of correlation of recognition rates with distances is opposite between coarse grained space and the "raw CML states". While correlation is weak to moderate, the trends for the dependence of recognition rate on the distance spread among objects are in opposite directions for raw state values (like those measured in LSM) and the coarse grained values used for classification. Specifically, the recognition rate improves with greater L2 Norm distance spread in the coarse grained space (fig 3, slope .02, r=.25), but decreases slightly with distance spread in the directly measured state space.(fig 2, slope -.003, r=.-.28).

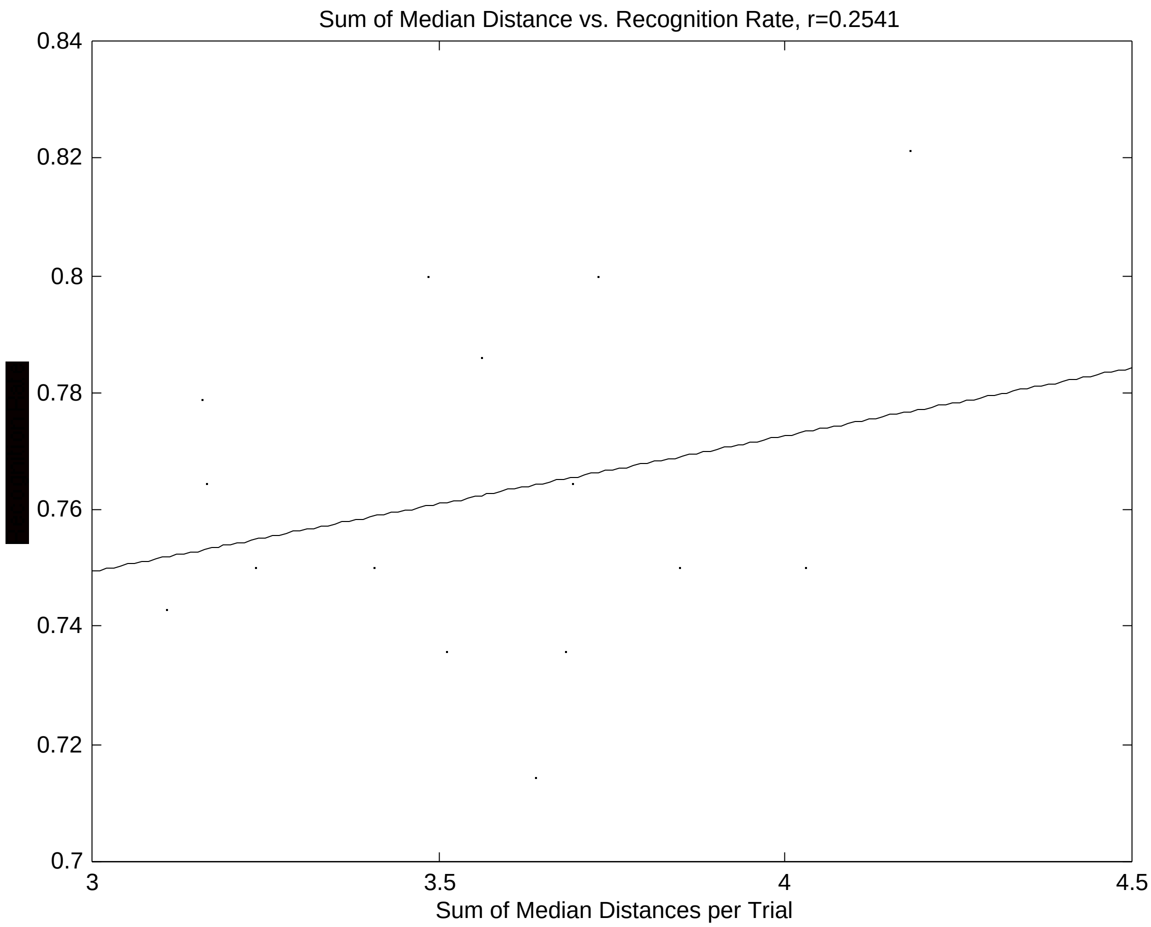

Figure 2 The distance between object-classifier response pairs in recognition trials is summed for each learning trial and plotted vs. the overall recognition rate for that trial. In this case the distances is are computed from coarse grained states measured over the entire lattice. Distance units are abstract space units in the unit cube bounding possible occupancy values.

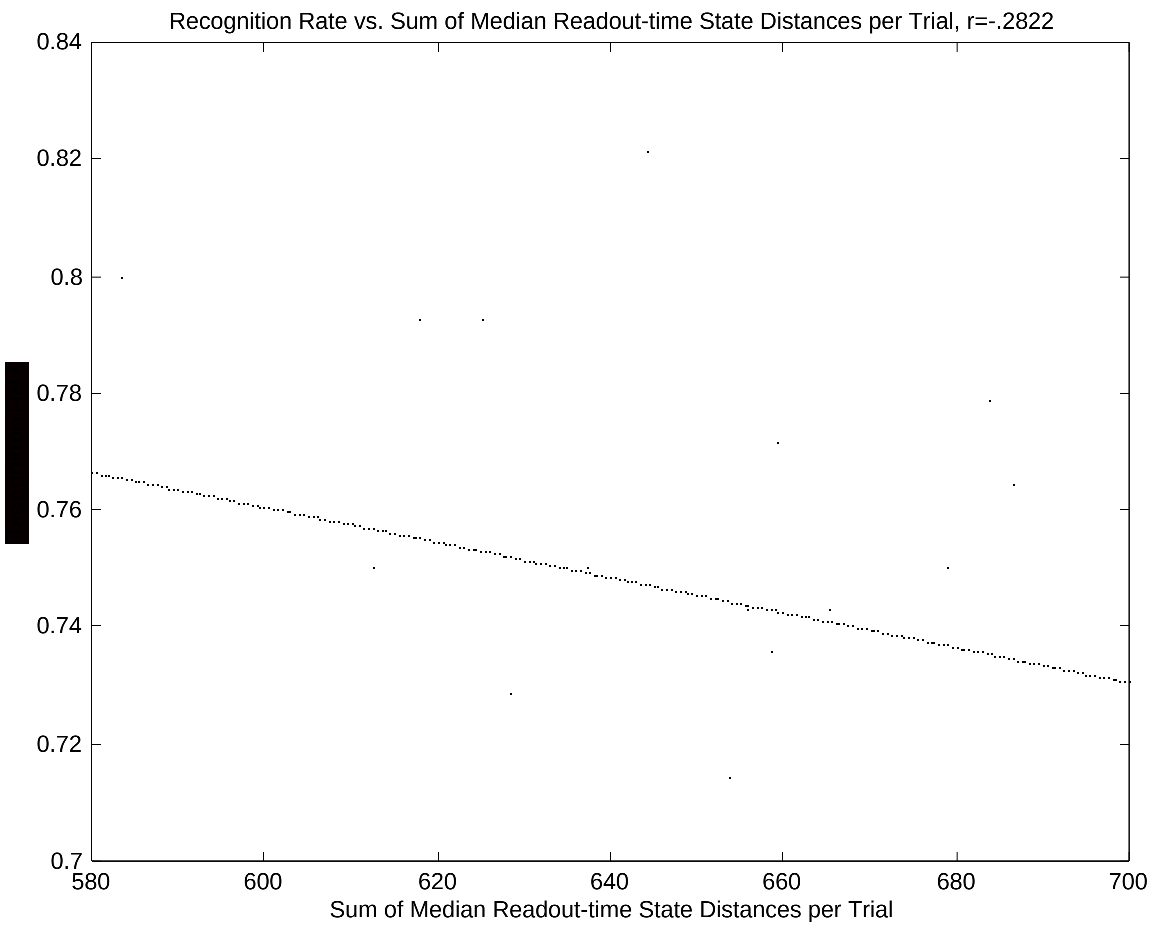

Figure 2 The distance between object-classifier response pairs in recognition trials is summed for each learning trial and plotted vs. the overall recognition rate for that trial. In this case the distance is plotted for “raw states” prior to coarse graining.

Work is now underway to compute these measures on a larger data set, and study the effect of adding a noise term to the logistic map during learning on such measures. It is hypothesized that while the noise will affect time series of individual units, the effect on separability and ultimately performance will be limited. Synchronization effects, broad fan-n to distributions sensitive readout units, and the statistical behavior of coupled chaotic networks with broad spatial convergence may compensate for the effects of local chaotic behavior separation.

The dynamics and learning of the system, which has been described previously in this conference and publications, will briefly be presented.

Aertsen, A. and H. Preissl (1991). Dynamics of activity and connectivity in physiological neuronal networks. Nonlinear Dynamics and Neuronal Networks. S. H. Weinhelm, VHC Verlag**:** 281-301.

DeMaris, D. (2001). Synchronization Opponent Networks: Dynamics, Computation and Coding for Similarity and Object Recognition. Electrical and Computer Engineering. Austin, University of Texas.

DeMaris, D. and B. F. Womack (2001). Recognition of Objects Rotated in Depth Using Partial Synchronization of Chaotic Units. International Joint Conference on Neural Networks, Washington, D.C., IEEE/INNS.

Driebe, D. (1999). Fully chaotic maps and broken time symmetry. Dordrecht, Kluwer Academic.

Freeman, W. J., R. Kozma, et al. (2001). “Biocomplexity: adaptive behavior in complex stochastic dynamical systems.” Biosystems **59**: 109-123.

Kaneko, K. (1990). “Clustering, coding, switching, hierarchical ordering and control in a network of chaotic elements.” Physica D **41**: 137-142.

Maas, W., T. Natschlager, et al. (2002). “Real-Time Computing Without Stable States: A New Framework for Neural Computation Based on Perturbations.” Neural Computation **14**: 2531-2560.

Milton, J. G. and M. C. Mackey (2000). “Neural Ensemble Coding and Statistical Periodicity: Speculations on the operations of the mind's eye.” Journal of Physiology (Paris) **94**: 489-503.