# Mathematical foundation and applications of high dimensional data clustering: similarity regulated processing speed for pattern recognition

Jianhong Wu[*]

Laboratory for Industrial and Applied Mathematics, Faculty of Sciences and Engineering; MITACS Centre for Disease Modeling, York Institute of Health Research, York University, Toronto, Canada, M3J1P3.  wujh@yorku.ca

**Extended abstract:** This talk intends to put a few PhD theses, research publications and projects of the York University's Laboratory for Industrial and Applied Mathematics in a coherent framework about information processing delay, high dimension data clustering, and nonlinear neural dynamics. The objective is to develop both mathematical foundation and effective techniques/tools for pattern recognition in high dimensional data. We refer to the monograph [G. Gan, C. Ma and J. Wu, Data Clustering: Theory, Algorithms and Applications, SIAM, 2007] for other clustering algorithms, and the survey paper [J. Wu, High dimensional data clustering from a dynamical systems point of view, The Fields Institute Communication, AMS, 2006] for a heuristic description of the philosophy that the nonlinear dynamic systems theory may provide some theoretical principles based on recent biological evidences for novel neural network based clustering architectures.

In the papers [Y. Cao & J. Wu, IEEE Trans Neural Networks, 15(2004), 245-260; Neural Networks, 15(2002),105-120] and the thesis [Y. Cao, Neural networks for clustering: theory, architecture, algorithm, and neural dynamics, York University, 2002], we developed a novel neural network architecture and algorithm to detect low dimensional patterns in a high dimensional data set. These patterns are associated with the projective clusters introduced by Aggarwal and his co-workers from the IBM Watson Centre.  The developed projective adaptive resonance theory (PART) has received much attention by data clustering researcher community and industry, and formed the core a Collaborative Research Development project funded by the Natural Science and Engineering Research Council of Canada (NSERC) in collaboration with Generation 5 Mathematical Technologies Inc. The PART algorithm has since been used in a number of applications. For example, it was used to develop a powerful gene filtering and cancer diagnosis method in [H. Takahashi et al, Bioinformatics, 2004; BMC Bioinformatics, 2006], which shows that *"the results have proven that PART was superior for gene screening".* The PART clustering was also used for clustering neural spiking trains [J. Hunter, J. Wu and J. Milton, CDC, 2008], ontology construction [R. Chen and C. Chuang, Expert Systems, 2008], and stock associations [Liu et al., Neurocomputing, 2009]. The PART algorithm has also been extended to deal with categorical data in the thesis [G. Gan, Subspace clustering for high dimensional data clustering, York University, 2003].

The PART architecture is based on the well known ART developed by Carpenter and Grossberg, with a selective output signaling (SOS) mechanism to deal with the inherent sparsity in the full space of the data points in order to focus on dimensions where

---

[*] corresponding author. Jianhong Wu, Professor and Canada Research Chair in Industrial and Applied Mathematics. Tel: 4167365250

information can be found. The key feature of the PART network is a hidden layer of neurons which incorporates SOS to calculate the dissimilarity between the output of a given input neuron with the corresponding component of the template (statistical mean) of a candidate cluster neuron and to allow the signal to be transmitted to the cluster neuron only when the similarity measure is sufficiently large. Recently discovered physiological properties of the nervous system, the adaptability of transmission time delays and the signal losses that necessarily arises in the presence of transmission delay, enabled us to interpret SOS as a plausible mechanism from the self-organized adaptation of transmission delays driven by the aforementioned dissimilarity. The result is a novel clustering network, termed PART–D, with physiological evidence from living neural network and rigorous mathematical proof of exceptional computational performance [J. Wu, H. Zivari-Piran, Hunter and Milton, Neurocomputing, 2011].

Such an adaptation can be regarded as a consequence of the Hebbian learning law, and the dynamic adaptation can be modeled by a nonlinear differential equation. As a result, we obtained a new class of multi-scale systems of delay differential equations with adaptive delay. A key issue then is how to analytically formulate the delay adaptation. This links to another PhD thesis [D. Beamish, 50 years later: a neurodynamic explanation of Fitts' law, York University, 2004], which proposed an alternative neural network formulation of the Fitts' law for the speed-accuracy trade-off of information processing. A number of publications have been resulted from this thesis work, including [D. Beamish et al., Biological Cybernetics, 2008, 2009; J. Royal Soc. London Interface, 2006; Neural Networks, 2006; J. Math. Biology, 2005]. It remains though an open problem how to use this alternative neurodynamical formulation to obtain a precise delay adaption rule of the PART-D neural network architecture for projective clustering.

When the delay adaption rates are in certain ranges, we anticipate nonlinear oscillatory behaviors of the PART-D neural network as the signal processing delay has been recognized as a major mechanism for nonlinear oscillation in the form of Hopf bifurcations, and this oscillation slows down the convergence of the clustering algorithm. How to detect the birth and to describe the global persistence of these nonlinear oscillations is the central subject of the thesis [Q. Hu, Differential equations with state-dependent delay: global Hopf bifurcation and smoothness dependence on parameters, York University, 2008] and the studies [Q. Hu and J. Wu, J. Diff. Eqns., 2010; J. Dyna. Diff. Eqns., 2010].

In summary, there have been increasing physiological evidences to support the idea of projective clustering using neural networks with delay adaption, there has been some preliminary theoretical analysis to show why such a network architecture works well for high dimensional data, and there have been sufficient applications to illustrate our PART network based clustering algorithm is efficient. An interdisciplinary approach for high dimensional data clustering clearly shows the potential to develop a dynamical system framework for pattern recognition in high dimensional data.