International Journal of Instrumentation Control and Automation

Volume 2 | Issue 1 Article 8

April 2012

METACOGNITION REVEALED COMPUTATIONALLY

RAMAKRUSHNA SWAIN
Dept. of CS&E, BIT, Bhubaneswar, India, rkswain24@gmail.com

NARENDRA KUMAR KAMILA
Dept. of CS&E, CVRCE, Bhubaneswar, India, NARENDRAKUMARKAMILA@GMAIL.COM

METACOGNITION REVEALED COMPUTATIONALLY

RAMAKRUSHNA SWAIN [1], NARENDRA KUMAR KAMILA [2]

[1]Dept. of CS&E, BIT, Bhubaneswar, India; [2]Dept. of CS&E, CVRCE, Bhubaneswar, India

[rkswain24@gmail.com]

Abstract- This paper focuses on current progress for the understanding of human cognition. Here different models have been considered such as MLP, FLANN, PNN, MLR, and HSN for recognition of one of the state of mind. It is argued that in addition to other models, PSO occupies a prominent place in the future of cognitive science, and that cognitive scientists should play an active role in the process. Baysian Approach in the same context has also discussed. The special case of predicting harm doing in a particular mental state has been experimented taking different models into account in depicting decision making as a process of probabilistic, knowledge-driven inference.

Keywords *- Hamming Network, Particle Swarm Optimization, Pattern Recognition, Swarm Intelligenc,Bayes Approach*

I. INTRODUCTION

The most unambiguous modern scientific view is that the brain enables the mind — that is, the physical organ gives rise to the hard-to-define collection of mental mechanisms governing our cognitive existence. On this view, the brain is widely believed to be a deterministic system, with millions of interacting parts that produce reliable and automatic responses to environmental challenges. Moral judgments and choices are mental phenomena that fit this general pattern.

In recent years, researchers in brain science have attempted to test these competing claims by examining such concepts as reciprocity, justice, and morality. Starting with the simple observation that humans do react largely the same to many moral challenges, and fail to react the same in other situations, how does the human brain sort this all out? How do moral behavior and thought actually work? The aim of this effort is to analyze different methods evaluating the nature of human decision making behavior computationally. When our brains integrate the myriad information that goes into a decision to act, prior learned rules of behavior are part of that information flow[1].

New approaches analyze and exploit the complex causal structure of physically embodied and environmentally embedded systems, at every level, from molecular to social. These approaches have improved our ability to use computers for more and more robust simulations of intelligent agents-simulations that will increasingly control machines occupying our cognitive niche[2].

A major in Neuropsychology and Cognitive Science will build a scientific understanding of the psychological processes of the individual and the relationship of these processes to brain function. It assumes that cognition can at least in principle be fully revealed by the scientific method, that is, individual components of mental processes can be identified and understood[3].

Given that the core of cognitive science is computational accounts of human cognition, and metacognition the scope of cognitive science is nothing less, in general terms, than the mind, or rather, the functions and processes of the mind, which are also, to a large extent, those of the brain[4][5].

Development of mathematical models of higher level cognition i.e., metacognition and understanding the formal principles that underlie our ability to solve. The computational problems we face in everyday life related to thinking can be done by analyzing these aspects of human cognition by comparing human behavior to optimal or "rational" solutions to the underlying computational problems. For inductive problems, this usually means exploring how ideas from artificial intelligence, machine learning, and statistics connect to human cognition[6].

Section II discuss the conditional probability approach to deal with the uncertainity and Section III deals with the Particle Swarm Optimisation (PSO) processes along with MLR, MLP, FLANN, PNN, HN and HSN. The next sections give the simulation result and conclusion.

II. BAYESIAN APPROACH FOR PREDICTION

In this context, Gott's (1993) Copernican anthropic principle has some applicability which suggests how we might formulate a rational statistical account of our ability to predict the future behaviours. Gott's (1993) delta-t argument does not incorporate the prior knowledge about durations that people bring to the problem of prediction, or the possibility of multiple observations. However, it can be shown that the delta-t argument is equivalent to a simple Bayesian analysis of the problem of predicting the future behavior (Gott, 1994). Bayesian inference naturally combines prior knowledge with information from one or many observations, making it possible to extend Gott's argument to provide a more general account of prediction. Bayes' rule states that

$$P(h|d) = \frac{P(d|h)P(h)}{P(d)},$$

where h is some hypothesis under consideration and d is the observed data. By convention, $P(h|d)$ is referred to as the posterior probability of the hypothesis, P(h) the prior probability, and P(d|h) the likeliihood, giving the probability of the data under the hypothesis. The denominator P(d) can be obtained by summing across P(d|h)P(h) for all hypotheses, giving

$$P(h_i|d) = \frac{P(d|h_i)P(h_i)}{\int_{\mathcal{H}} P(d|h)P(h)dh},$$

Where $\mathcal{H}$ is the set of all hypotheses.

A cognitive architecture featuring an explicit set of goals, and an action selection system that causes it to choose those actions that it rationally calculates will best help it achieve the above goals.

III. PSO Approach

The focus is on the design and implementation of the Particle Swarm Optimization (PSO) algorithms for function optimization problems of real world applications. PSO incorporates swarming behaviors observed in flocks of birds, schools of fish, or swarms of bees, and even human social behavior, from which the idea is emerged [14,7,22]. PSO is a population-based optimization tool, which could be implemented and applied easily to solve various function optimization problems[7]. Human social behavior is more complex than a flock's movement. Social sharing of information among the individuals of a population may provide an evolutionary advantage. This was the core idea behind the development of PSO[11]. A program capable of empathetic decision-making or compassionate social interaction requires some meta-cognition as part of the bounded informatic situation. Namely, the cognition of such an agent includes thinking about thinking, thinking about feeling, and thinking about thoughts and feelings – its own and/or those of other agents[12].

Mental state and their behavioral expressions play an important role in human reasoning, decision making, and communication [9]. Beliefs, intents, desires, pretension, and knowledge are the back bone of the affective states, which may be the reason for human behavior, and hence can be used to predict others behavior[13].

Visualisation of human reasoning is one sort of metacognition. However little has been done to evaluate it computationally. Swarm Intelligence techniques to visualize the human reasoning for prediction of behaviour[14]. However, recent researches can approach the process of the mind scientifically by developing measurement machines and methods using computer which simulates the mind.The statistics of moral reasoner of UCI machine learning repository has been used to validate the work. A number of rule sets have been developed to visualize the predictability of human behavior to a reasonable extent.

The different methods we have considered such as MLP, FLANN, PNN, MLR, and HN for evaluating metacognition computationally. This recognizes one of the state of mind. The pattern for mapping the states of brain are quite complex in nature. A single technique may not be suitable to approximate the input-output patterns representing the states of brain[13].

A. Multiple Linear Regression

Multiple regression simultaneously considers the influence of multiple explanatory variables on a response variable Y

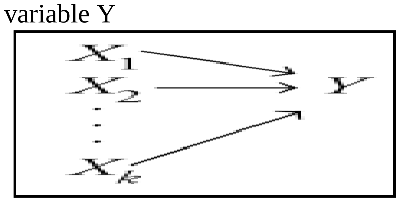

Fig. 1 : Multiple Linear Regression

The intent is to look at the independent effect of each variable while "adjusting out" the influence of potential confounders.

Again, estimates for the *multiple* slope coefficients are derived by minimizing $\sum$residuals^2 to derive this multiple regression model:

$$\hat{y} = a + b_1 x_1 + b_2 x_2$$

Again, the standard error of the regression is based on the $\sum$residuals^2:

$$S_{Y|x} = \sqrt{\sum \text{residuals}^2 / df_{\text{res}}}$$

- Intercept α predicts where the regression *plane* crosses the Y axis
- Slope for variable X_1 (β_1) predicts the change in Y per unit X_1 holding X_2 constant
- The slope for variable X_2 (β_2) predicts the change in Y per unit X_2 holding X_1 constant

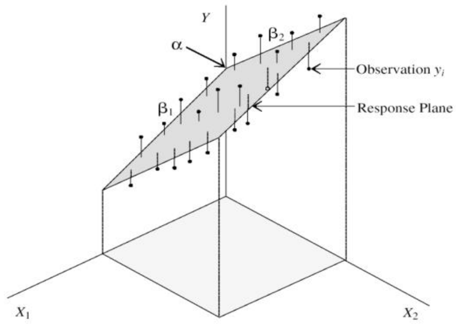

Fig.2 : Three-dimensional response plane

B. *Multi Layer Perceptron*

The most popular class of multilayer feed forward networks is MLP in which each computational unit employs either the threshold function or the sigmoid function. MLP can form arbitrarily complex decision boundaries and represent any Boolean function [15]. For training of the MLP models the back propagation algorithm [16] is used. Back propagation learning uses gradient descent method to minimize the squared error cost function. Fig.1 presents the architecture of the MLP model.

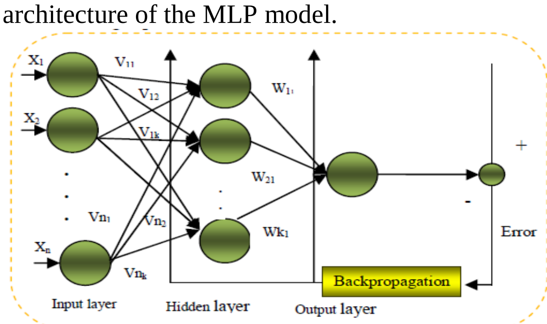

Fig. 3 : Architecture of MLP

C. *Functional Link Artificial Neural Network*

The most common architecture of ANNs is the multilayer feedforward network (MLP). MLP utilize a supervised learning technique called Backpropagation for training the network. However, due to its multi-layered structure, the training speeds are typically much slower as compared to other single layer feedforward networks [3]. Problems such as local minima trapping, overfitting and weight interference also make the network training in MLP become challenging. Hence, Pao has introduce an alternative approach named Functional Link Neural Network (FLNN) in avoiding these problems[17] .

This approach removes the hidden layer from the ANN architecture to help in reducing the neural architectural complexity and provides them with an enhancement representation of input nodes for the network to be able to perfom a non-linear separable classification[18].

Functional Link Neural Network (FLNN) is a class of Higher Order Neural Networks (HONNs) that utilize higher combination of its inputs [20, 21]. It was created by Pao [21] and has been successfully used in many applications such as system identification [23-28], channel equalization [31], classification [29-32], pattern recognition [33, 34] and prediction [35, 36]. In this paper, we would discuss on the FLNN for the classification task. FLNN is much more modest than MLP since it has a single-layer network compared to the MLP but still is able to handle a non-linear separable classification task. The FLNN architecture is basically a flat network without any hidden layer which has make the learning algorithm used in the network less complicated [22]. In FLNN, the input vector is extended with a suitably enhanced representation of the input nodes, thereby artificially increasing the dimension of the input space [20, 21].

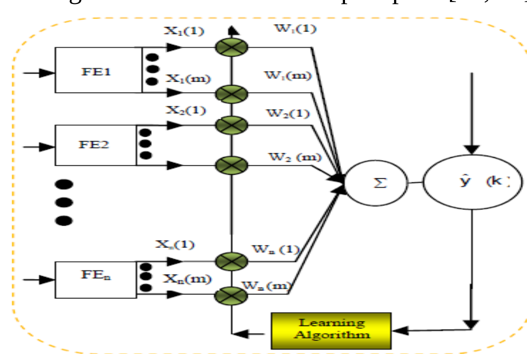

Fig. 4 : Learning Algorithm

D. *Polynomial Neural Network*

PNN - a self-organizing multi-layered iterative algorithm that automatically provides linear and non-linear polynomial regression models. The PNN embodies the advantages of Multiple Linear Regression (MLR) and Artificial Neural Networks (ANNs) into a single entity. It can model both linear and non-linear relationships like ANNs, and it yields a polynomial regression equation like MLR for easy interpretation. This algorithm provides robust results in the presence of correlated and irrelative variables or/and outliers. The results of this algorithm can be easily interpreted[37].

The algorithms are implemented as a GMDH-type Neural Network. First layer generates the models ***y***

$= g(x_i, x_j, x_k)$, where where x_i, x_j, x_k input variables. Next layers generate models as $y = g(w_i, w_j, w_k)$, where w_i, w_j, w_k are the models of previous layers.

The main objective is to create a stable algorithm for nonlinear model reconstruction such as that its results are easily interpreted by users.

The main features of the algorithm can be summarized as follows:

1. **Fast learning.** The transforms with two coefficients only are used, for example $g(w_i, w_j) = a \cdot w_i + b \cdot w_j$ in the linear case. Irrespectively of the power of resulting model and the number of terms the second order matrices are only inverted. This provides fast learning of the algorithm.
2. **Results in the parametric form.** The polynomial structures are coded using vector of simple numbers [1,2] that provides the presentation of the results in the parametric form of nonlinear equation.
3. **Complexity control.** Let us denote vector $(power, c)^T$ as a complexity, power is the *power* of the polynomial and c is the number of terms. The power of the new model is controlled by the condition that if, for example, $g(w_i, w_j, w_l) = a \cdot w_i + b \cdot w_j \cdot w_l$, then $power(g(w_i, w_j, w_l)) = max(power(w_i), power(w_j) + power(w_l))$, where *power()* designates the power of the polynomial. It gives us the possibility to restrict the class of the models under consideration by $power(w_i) < p$ and to search models among the polynomials with power less than p. The maximum complexity is defined by the user or can be automatically selected using a full cross-validation method.
4. **Twice-hierarchical neural net structure.** Twice-hierarchical neural net structure is important feature of PNN. One of the problem is that power of polynomials increases too fast in the traditional GMDH algorithm. At the step r of iteration procedure one can have models of power $r+1$, $W_r \in P^{r+1}$. The control of complexity gives us an opportunity to implement the iteration procedure without an increase of the power of polynomials or/and the number of terms. External iterative procedure controls the complexity, i.e. the number of the terms and the power of the polynomials in the intermediate models. The best models form initial set for the next iterative procedure. This procedure realizes a wide search without the complexity increase. Besides that the twice-hierarchical neural net structure provides the convergence of the coefficients. The models that are calculated as a result of several transformations have the coefficients that are close to the appropriate regression coefficients.
5. **Robust estimation.** To use algorithm in the presence of large errors (outliers) we have developed the PNN algorithm for robust nonlinear (M-regression) model identification. This made possible to improve stability of PNN algorithm[38].

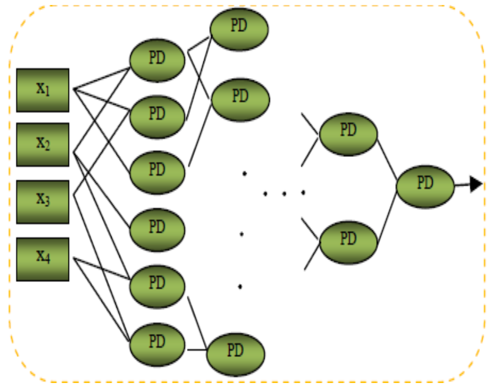

Fig. 5 : Architecture of a typical PNN

E. *Hamming Network*

Hamming network selects stored classes, which are at a maximum distance from the noisy vector, presented as the input [39]. Here the weight vector for the clustering network is termed as exemplar vector or code book vector. The weights for the net are determined by the exemplar vectors. The difference between the total number of components and the Hamming distance between the vectors gives the measure of similarity between the input vector and stored exemplar vectors, where the Hamming distance between the two vectors is the number of components in which the vectors differ.

Consider two bipolar vectors x (input vector) and y (exemplar vector), then the relation obtained is

$$x.y = a - d \quad (6)$$

where a is the number of components in which the vector agree, d the number of components in which the vectors disagree and the value $a - d$ is the Hamming distance existing between two vectors.

Since the total number of components is n, it can be stated that,

$$n = a + d \Rightarrow d = n - a \quad (7)$$

On substituting the value of d from equation (7) in equation (6), it is derived that

$$x.y = a - (n - a) \Rightarrow x.y = 2a - n \Rightarrow 2a = x.y + n \Rightarrow a = \frac{x.y}{2} + \frac{n}{2}$$
(8)

From the above equation, it is clear that the weight should be set one-half of the exemplar vectors and the bias should be set to one-half of the number of bipolar bits in the input pattern.

By calculating the unit with the largest net input, the net is able to locate a particular unit that is closest to the exemplar. The unit with the largest net input is obtained by the Hamming net using the Maxnet as its subnet.

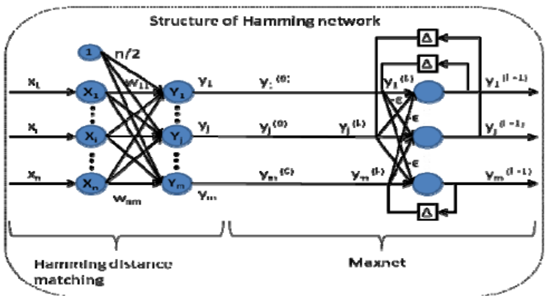

Fig. 6 : Architecture of Hamming Network

F. *Hamming Swarm Net*

It is observed that the Hamming network fails to recognize the pattern of the state of mind. Constant exemplar vectors may be the hindrance in proper mapping. To explore it, the HSN techniques is proposed.

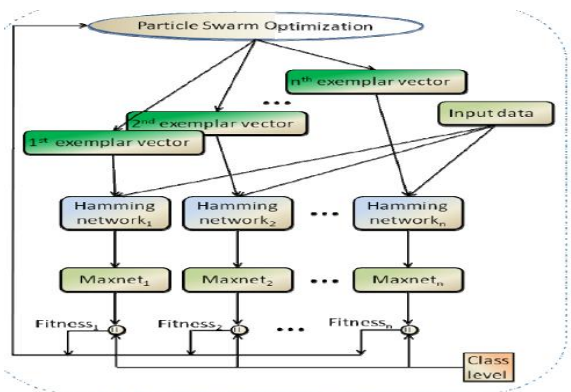

Fig. 7 : Architecture of Hamming Swarm Net

The HSN model deploys a set of n particles for the PSO, which represents the exemplar vectors for n number of HNs. The input data along with the respective exemplar vectors are passed on to the HNs. Each neuron of the HN represents as one class of the problem. The similarity of the patterns with the exemplar vector of each neuron of the HN is passed on to the Maxnet as the output of the HN. The decision of the Maxnet in favor of a particular class is compared with the actual class level. If the predicted class level is same as the actual class level, fitness of the particle is incremented by one otherwise it is decremented by one. In the distributive environment, each HN competes with the other to maximize the fitness. The fitness values are passed on to the PSO to update the personal best and global best values and to modify the values of the exemplar vectors.

IV. SIMULATION

The statistics available in the moral database of UCI machine learning repository [Shultz & Daley, UCI,1994]has been used for the simulation. This database considers a rule-based model that qualitatively simulates moral reasoning. The study was intended to simulate about the prediction of human behaviour particularly of a child making harmful response to the given stimuli. The mental state attribute possessing only binary values in the database are considered.

Table 1 : Parameters of PSO considered for simulation

Parameters	Values
Population Size	40
Maximum Iterations	200
Inertia Weight	0.729844
Cognitive	1.49445
Social Parameter	1.49445

50 simulations results are considered for analysis. For each simulation the database is randomly divided into two sets. The first 70% data is used to train the model and the remaining 30% is used to test the performance of the simulation study.

From average prediction accuracy of 50 simulations by different techniques, HSN model provides the best result followed by MLR and FLANN models.

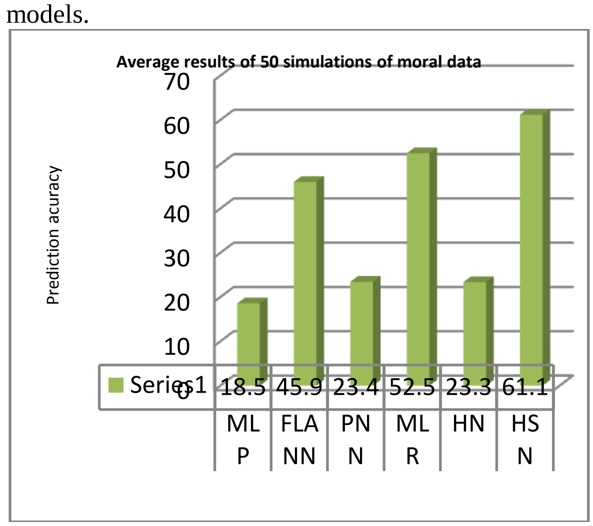

Fig.8:Average results obtained from 50 simulation by diff. models.

V. CONCLUSION

Brains, it has recently been argued, are essentially prediction machines. They are bundles of cells that support perception and action by constantly attempting to match incoming sensory inputs with top-down expectations or predictions. This is achieved using a hierarchical generative model that aims to minimize prediction error.

Predicting the human mental state is a complex problem that can be solved by the component of planning, decision-making, memory, and causal reasoning. This paper presents a simulation study for testing six different models of predicting the accuracy of the mental phenomena from their current state. Such accounts offer a unifying model of prediction for encoding mental states.

REFERENCES

[1] Michael Gazzaniga : Homepage at http://www.templeton.org/reason/Essays/gazzaniga.pdf

[2] Colin Allen "The Future of Moral Machines", Available at http://opinionator.blogs.nytimes.com/2011/12/25/the-future-of-moral-machines/

[3] Fodor, Jerry A. (1983). "Modularity of Mind: An Essay on Faculty Psychology" Cambridge, Mass.: MIT Press.

[4] Patterson, K., and Plaut, D. C. (2009). "Shallow draughts intoxicate the brain",Topics in Cognitive Science, 1, 39-58.

[5] Daniel Andler Dissertation: "Evolutionary modelling and the history of technology", 2003

[6] Griffiths, T.L., Kemp, C., & Tenenbaum, J.B. (2008). Bayesian models of cognition, In Ron Sun (ed.), The Cambidge handbook of computational psychology, Cambridge University Press

[7] Ajith Abraham, He Guo, and Hongbo Liu (Swarm Intelligence: Foundations, Perspectives and Applications Available at http://www.softcomputing.net

[8] S. C. Satapathy, J. V. R. Murty, P. Reddy, B. B. Misra, P. K.Dash, G. Panda, "Particle Swarm Optimized Multiple Regression Linear Model for Data Classification Applied Soft Computing," Volume 9, Issue 2, March 2009, pp. 470-476.

[9] J. P.Sarte, Existentialism and Human Emotions, Citadel Press; Reissue edition 1984

[10] J. A. Muller, F. Lemke, A. G. Ivakhnenko, "GMDH algorithms for complex systems modeling. Math and Computer Modeling of Dynamical Systems" vol. 4, 1998, pp. 275-315.

[11] R. Eberhart, P. Simpson, and R. Dobbins, Computational Intelligence PC Tools, Academic Press, Boston, 1996.

[12] Cindy L. Mason "Human-Level AI Requires Compassionate Intelligence", 2008 AAAI Workshop On Meta-Cognition

[13] R. K. Swain, N. K. Kamila, B. B. Misra, "Recognition of the State of Mind using Hamming Swarm Net", World Congress on Information and Communication Technology 2012.

[14] J. Y. Park, B. M. Gu, D. H. Kang, Y. W. Shin, C. H. Choi, J. M. Lee, and J. S. Kwon, "Integration of Cross-Modal Emotional Information in the Human Brain: an fMRI study," Cortex; a journal devoted to the study of the nervous system and behavior, Vol.46, No.2, pp.161-169, 2010.

[15] M. Minisky and S. Papert, Perceptrons: An introduction to Computation Geometry, MIT Press, Cambridge, Mass., 1969.

[16] M. Morf, and T. Kailath, "Square-root Algorthim for least square estimation," IEEE Trans. on Automatic Control, vol. 20, Issue 4, Aug 1975, pp. 487-497.

[17] M. Kimura, I. Daibo, and M. Yogo, "The Study of Emotional Contagion from the Perspective of Interpersonal Relationships," Social Behavior and Personality, Vol.36, No.1, pp.27-42, 2008.

[18] Available at taskhttp://arxiv.org/ftp/arxiv/papers/1212/1212.6922.pdf

[19] M. Kimura, I. Daibo, and M. Yogo, "The Study of Emotional Contagion from the Perspective of Interpersonal Relationships," Social Behavior and Personality, Vol.36, No.1, pp.27-42, 2008.

[20] P. S. Foster, V. Drago, D. G. Webster, D. W. Harrison, G. P. Crucian, and K. M. Heilman, "Emotion Influences on Spatial Attention," Neuropsychology, Vol.22, No.1, pp.127-135, 2008.

[21] J. Y. Park, B. M. Gu, D. H. Kang, Y. W. Shin, C. H. Choi, J. M. Lee, and J. S. Kwon, "Integration of Cross-Modal Emotional Information in the Human Brain: an fMRI study," Cortex; a journal devoted to the study of the nervous system and behavior, Vol.46, No.2, pp.161-169, 2010.

[22] T. Sobol-Shikler, and P. Robinson, "Classification of Complex Information: Inference of Co-Occurring Affective States from Their Expressions in Speech," IEEE Transactions on PAMI, vol. 32, no. 7, 2010, pp. 1284-1297.

[23] M. V. den Noort, M. P. C. Bosch, and K. Hugdahl, "Understanding the Unconscious Brain: Can Humans Process Emotional Information in a Non-Linear Way?" Proc. Int'l Conf. Cognitive Systems, Dec. 2005.

[24] J. Kennedy, and R. Eberhart, Swarm Intelligence, Morgan Kaufmann Academic Press, 2001.

[25] J. Kennedy, and R. Mendes, "Population Structure and Particle Swarm Performance," Proc. IEEE Congress on Evolutionary Computation, 2002, pp. 1671–1675.

[26] S. C. Satapathy, J. V. R. Murty, P. Reddy, B. B. Misra, P. K.Dash, G. Panda, "Particle Swarm Optimized Multiple Regression Linear Model for Data Classification Applied Soft Computing," Volume 9, Issue 2, March 2009, pp. 470-476.

[27] C. L. Lawson and R. J. Hanson, Solving Least Square Problems, Prentice Hall,1974

[28] M. Morf, and T. Kailath, "Square-root Algorthim for least square estimation," IEEE Trans. on Automatic Control, vol. 20, Issue 4, Aug 1975, pp. 487-497.

[29] M. Minisky and S. Papert, Perceptrons: An introduction to Computation Geometry, MIT Press, Cambridge, Mass., 1969.

[30] D. E. Rumelhart and J. L. McClelland, Parallel Distributed Processing: Exploration in the Microstructure of Cognition, MIT Press, Cambridge, Mass., 1986.

[31] B. B. Misra, S. Dehuri, "Functional Link Artificial Neural Network for Classification Task in Data Mining," J. Computer Sci., vol. 3, issue 12, pp. 948-955, 2007.

[32] B. B. Misra, S. Dehuri, and S.-B. Cho, "Genetic Feature Selection for Optimal Functional Link Artificial neural Network in Classification", 9th International Conference on Intelligent Data Engineering and Automated Learning, Nov 2-5, 2008, Korea, C. Fyfe et al. (Eds.): IDEAL 2008,Lecture Notes in Computer Science (LNCS) 5326, pp. 156–163, © Springer-Verlag Berlin Heidelberg 2008.

[33] H. R.Madala, and A. G. Ivakhnenko, Inductive Learning Algorithm for complex systems Modelling, Boca raton:CRC Inc, 1994.

[34] J. A. Muller, F. Lemke, A. G. Ivakhnenko, "GMDH algorithms for complex systems modeling. Math and Computer Modeling of Dynamical Systems" vol. 4, 1998, pp. 275-315.

[35] S. J. Farlow, The GMDH algorithm, In: S. J. Farlow, editor.Self-organizing methods in modelling:GMDH type algorithm.New York:Marcel Dekker, 1984, pp.1-24.

[36] B. B. Misra, S. C. Satapathy, B. Biswal, P. K. Dash, G. Panda, "Pattern Classification using Polynomial Neural Network" IEEE International Conferences on Cybernetics & Intelligent Systems (CIS) and Robotics, Automation & Mechatronics (RAM) (CIS-RAM 2006),7 - 9 June 2006 in Bangkok, Thailand.

[37] Yurachkovsky, Y.P. 1981. "Restoration of Polynomial Dependencies Using Self-Organization." Soviet Automatic Control, 14, 17-22. (http://pnn.pnnsoft.com/PNN-algorithm2.html

[38] T.I Aksenova, I.V. Tetko, V.V. Volkovich, GMDH-type Neural Network in Quantative-Structure Activity Relationship Studis on the Internet, Modelling and simulation 2001(ESM2001), 15th European Simulation Multiconference, Prague, pp.685-689

[39] R. P. Lippmann, "An introduction to computing with neural nets," IEEE ASSP Magazine, 1987, pp. 4-22.

XXX