Lazy and Competitive Learning

JASON BROWNLEE
Technical Report 070508A
Complex Intelligent Systems Laboratory, Centre for Information Technology Research,
Faculty of Information and Communication Technologies, Swinburne University of Technology
Melbourne, Australia
jbrownlee@ict.swin.edu.au

***Abstract*-Lazy learning is a general learning principle in which models are constructed from a database of cases on an 'as-needed' basis. Methods include instance-based approaches like nearest neighbour and case based reasoning. Another learning paradigm with a lot of commonality is competitive learning, in which populations of units or high-order modules compete for attention to adapt or respond (produce system output). This work reviews these two general learning paradigms and proposes the clonal selection approach as possessing aspects of laziness, whilst being strongly competitive.**

Keywords- Instance Based Learning, Competitive Learning, Lazy Learning, Artificial Immune Systems, Clonal Selection Theory

I. INTRODUCTION

Lazy learning describes case-based learning and reasoning approaches that scan a database of stored patterns and construct a localised model as required to respond to a query. Competitive learning is a paradigm in which a structured or unstructured population of units compete with each other with regard to a stimulus, where the winner (or winners) of the competition may respond and be adapted. Algorithms that implement both lazy and competitive learning are suited to different specific concerns, although they are both generally non-parametric, and suited to the general domains of function approximation, classification, and regression. This work briefly and generally reviews the lazy learning and competitive learning paradigms, and comments as to their relevance to algorithms inspired by the clonal selection theory of acquired immunity.

Section II summarises case-based reasoning and the lazy learning paradigm, including nearest neighbour methods, kernel methods, and related concerns in using such approaches. Section III introduces the competitive learning paradigm with a focus on the self-organizing map, the winner-take-all principle, and the neural gas algorithm. Finally, section IV discusses a general clonal selection approach in the context of lazy and competitive learning, and provides some suggestions for future research.

II. LAZY AND CASE-BASED LEARNING

Instance-based learning (IBL) is a nonparametric and supervised machine-learning paradigm where a model is constructed from domain data instances at query time. The complexity increases with the quantity of data, and the more data available, the more specific the model [12,45]. Instance-based learning is typically referred to as *Lazy Learning*, which is a learning paradigm where generalization occurs as required. Lazy learning may be contrasted with *Eager Learning* in which generalization occurs before a query is received. In addition to the deferral of generalization, lazy learning is also typified by the combination of stored data used to reply to requests, and the fact that the response is discarded after it is delivered [10,11,13].

Defer: Store all training data and defer processing until queries are given that require replies
Demand-Driven: Queries are answered by combining training data, using a local learning (neighbourhood) approach. Instance are (1) identified in as points space, (2) a similarity measure is used to define the neighbourhood, (3) prediction function uses information to answer query
Discard: After answering the query, the answer and any intermediate results are discarded.

Figure 1 - Summary of the characteristics of lazy learning

Instance-based learning may be considered a type of *local learning algorithm*, where reasoning is based on models (kernels) constructed from instances similar to the input instance. A typical concern with such models is the capability of the system to generalize is a trade-off between the *capacity* (size) of the model, and the number of *samples* to consider in the model. Bottou and Vapnik [25] rephrase this local learning trade-off to be between the *capacity* of the model and the *locality* of the samples, where locality is the shape and size of the instance selection region (neighbourhood). This locality property of model construction defines how local an algorithm is, such that an instance-based approach such as k-Nearest Neighbour is highly-local with a low capacity, and a connectionist approach such as a neural network is non-local with high capacity.

Case-based reasoning (CBR) (or memory-based reasoning), is a generalized application of instance-based learning applied to reasoning such as expert systems [9]. Aamodt and Plaza [1] provide a concise description of the various different identities of instance based learning from a case-based reasoning perspective. Their taxonomy highlights the perspective each identity places on this style of learning, as follows:

Exemplar-based Reasoning: A concept is defined by a set of exemplars (or prototypes). Typically used to describe the learning of concept definitions in machine learning. Reasoning is a classification task, where an unclassified exemplar is presented and assigned a class value.
Instance-based Reasoning: A specialization of the exemplar perspective, where instances rather than (potentially generalized) exemplars are maintained. A large number of instances are required to describe a concept. The focus of these approaches in on the automation of learning.

Memory-based Reasoning: These approaches focus on maintaining a large memory of cases, and reasoning as a function of accessing and searching the large memory. **Case-based Reasoning**: The focus is on the cases, where an individual case has a richness or complexity of detail. Cases may be modified or adapted for a specific purpose, and additional domain knowledge is applied during reasoning. **Analogy-based Reasoning**: Typically, a synonym for CBR, focus may be on the application of cases from one domain in another. About generalization and reasoning by analogy across (mapping onto other) domains.

Figure 2 - Summary of the various names for instance-based reasoning

Further, Aamodt and Plaza [1] propose a descriptive framework for CBR, which includes a general algorithmic cycle for the application of the approach, as follows:

1. *RETRIEVE*: the most similar case or cases
2. *REUSE*: the information and knowledge retrieved to solve the problem
3. *REVISE*: the proposed solution
4. *RETAIN*: the parts of the experience that may be useful for future problem solving

Finally, this algorithm may provide a general problem solving methodology in which the solution to future problems are drawn to the solution of past problems [19].

A. *Nearest Neighbour Methods*

Nearest neighbour methods are function approximation methods that search through a database for similar instances (cases) to a given input instance and make a prediction [46]. The similar points to consider in the prediction are referred to as the neighbourhood, and the number of points is typically denoted *k*, thus the scheme may be referred to as *k*-nearest neighbour or *k*-NN. An important consideration for this method is the distance metric used to determine similarity or neighbourhood between cases. One may use Euclidean distance for real-value, hamming for discrete, as well as many other similarity measures, as well as transformations including normalization and standardization techniques for preparing data. This paradigm of learning is about exploiting available data for tasks such as regression (fitting points to a local hyper-plane), and classification (voting on a discrete categorical attribute). Some common algorithms with a fixed k include the IB1 (k=1), IB2 (k=2) IB3 (k=3).

B. *Kernel Methods*

A natural extension to nearest neighbour approaches is that each case generates a local density function or kernel. Predictions may be made by summing together densities functions. The kernel function depends upon the distance function between cases, where the Gaussian is a typical function employed. Supervised learning may be achieved by looking at all training instances and weighting their contributions (density functions) to make a prediction (linear weighted regression). This approach may be combined with k-NN such that neighbourhoods of density functions are considered, rather than all training cases.

Some techniques include distance weighted regression such as Linear Weighted Regression (LWR) methods [5], and a neural-network like approach called Radial Basis Function (RBF) [7,26,28,29].

C. *Discussion*

Aha, et al. [12] (sighting Brieman et al. [21]) highlight the deficiencies of nearest neighbour and related techniques, suggesting that these problems must be addressed before a nearest neighbour algorithm can be applied to a real world problem. These highlighted problems are also applicable to case-based reasoning [9], as follows:

1) They are computationally expensive since they save all training instances **2)** They are intolerant of attribute noise **3)** They are intolerant of irrelevant attributes **4)** They are sensitive to the choice of the algorithms similarity function **5)** There is no natural way to work with nominal-valued attributes or missing attributes **6)** They provide little usable information regarding the structure of the data

Figure 3- Problems with nearest neighbour and descendant techniques

Must of the work on instance based learning algorithms is focused on reduction of storage requirements (thus increasing the efficiency of the approach), and improving the algorithms robustness to with regard to attribute noise and irrelevant attributes. Perhaps the most popular concern of nearest neighbour based approaches is what has been referred to as the '*curse of dimensionality*'. This refers to the fact that in common datasets, the nearest neighbour is approximately the same distance away from a given point as the furthest neighbour. This effect is observed in dimensions as low as 10-15 [20]. The problem is caused by the exponential increase in volume of a space by the adding of additional dimensions.

Aha and Kibler [12] propose to focus on five performance concerns when evaluating instance based learning techniques:

Generality: Concepts which are desirable by the representation and learnable by the algorithm **Accuracy**: The concept descriptions classification accuracy **Learning Rate**: The speed of the increase of classification accuracy during learning **Incorporation Cost**: The costs incurred when updating the descriptions with a single training instance (such as space, time, and classification costs) **Storage Requirement**: The size of the concept description, which is the number of saved instances used for classification

Figure 4 -Summary of some concerns in evaluating instance based learning algorithms

III. COMPETITIVE LEARNING

Intrator and Edelman [30] comment that competition between functional units (competitive learning) is a widespread biological phenomenon, particularly in the vision systems, the brain, and other sensory systems. Competitive learning systems are based on the idea of competition for activation, and lateral (same level) inhibition. Without inhibition, the models result in an averaged or muddied converged state. They define competitive learning as a dynamic redistribution of responsibilities of various units over parts of the representation space and propose a *global*, *local*, and a *hierarchal* perspective. The authors suggest that resource allocation may be addressed through selective inhibition, such that those units or modules that have not been activated for some time are inhibited less. Further, they suggest investigation into separating two types of competitive learning: (1) competitive learning via lateral inhibition, (2) competitive learning via the top-down separation of flows of information.

Global Competition: Competition over the entire representation space, with a strong requirement for inhibition via lateral connections between neurons and the distinction between hard (single winner) and soft (multiple winner) competition
Local Competition: The division of the representation space such that local experts (modules of units) compete (known as mixture of experts). Assumes that different processes generate different parts of the representational space, thus hard competition assume a one-to-one matching between modules and hidden generator processes that they model.
Hierarchal Competition: A mixture of global and local competition, where the representation space is partitions or split into a hierarchal tree structure.

Figure 5 - Three perspectives of competitive learning (taken from [30])

Competitive learning is a connectionist machine-learning paradigm where an input pattern is matched to the node with the most similar input weights, and the weights are adjusted to better resemble the input pattern. This is called the *winner-take-all* (or maximum activation) unsupervised learning method where the input pattern is compared to all nodes based on similarity. The nodes compete for selection (or stimulation) and ultimately adjustment (or learning) [18]. Kohonen distinguishes this connectionist learning paradigm from feed-forward and feed-backward approaches [49,53], as follows:

Signal Transfer Networks: (feed-forward paradigm) Signal transform circuits where the output signals depend on the input signals received by the network. Parametric in that the mapping is defined by a basis function (components of the structure) and fitted using an optimization approach like gradient decent. Examples include the multilayer Perceptron, back propagation, and radial basis function.
State-Transfer Networks: (feed-backward paradigm) Based on relaxation effects where the feedbacks and nonlinearities cause the activity state to quickly converge to one of its stable values (attractor). Input signals provide the initial activity state, and the final state is a result of recurrent feedbacks and computation. Examples include Hopfield network, Boltzmann machine, and bidirectional associative memory (BAM).
Competitive Learning: (self-organizing network paradigm) Networks of cells in simple structures receive identical inputs from which they compete for activation through positive and negative lateral interactions. One cell is the winner, and other cells are inhibited or suppressed. Cells become sensitive to different inputs and act as decoders for the domain. The result is a globally ordered map created via a self-organizing process. Examples include the Self-Organizing Map (SOM), and Learning Vector Quantization (LVQ).

Figure 6 - Comparison of some connectionist paradigms (from [49,53])

Fritzke [2] uses a taxonomy of hard (*winner-take-all*) and soft (*winner-take-most*) competitive learning and further distinguishes soft approaches to those with and without a fixed network topology.

Hard Competitive Learning: Winner-take-all (WTA) learning each input signal results in the adaptation of a single unit of the model. These methods may occur online or offline in batch. Examples include k-means.
Soft Competitive Learning: Winner-take-most (WTM) learning where an input signal results in the adaptation of more than one unit of the model. No fixed model dimensionality or topology is prescribed with these methods. Examples include neural gas.
Soft Competitive Learning with Fixed Structure: Winner-take-most (WTM) learning with a fixed model dimensionality and or topology. Examples include the self organizing map.

Figure 7 - Summary of taxonomy of competitive learning used by Fritzke [2]

A. *Self-Organizing Map (SOM)*

Kohonen self-organizing map [47,48,52] is a seminal achievement in the field of competitive learning (see [53] for a complete treatment). The work is centred on a generalization of the spatial order and organization of brain functions, where a (approximately homogeneous) mass of neurons with a fixed geography (such as planar lattice) self-organizes into a map of specialised functions though global competition. The entire map is exposed to input signals, and winning neurons (those with the maximum activation) responds, is adapted, and suppresses the activity of neighbouring neurons in the local vicinity through lateral inhibition. The result is that the topological properties in the feature space are preserved, and compressed to the geometric space of the neurons (dimensionality reduction).

Winner-Take-All: (WTA) A mechanism of self-organization, where the winner and neurons in the neighbourhood of the winner are adapted toward an input signal. Traditionally implemented uses lateral-feedback circuits where neurons outside the neighbourhood are suppressed from responding to the signal. A problem of WTA is that units that do not compete effectively or at all, thus the resource us underutilised or dead.

Figure 8 - Summary of the WTA function (from [53] pg 178-179)

The Self-Organizing Map (SOM) is an unsupervised learning algorithm (clustering) that embodies the WTA principle and facilitates topological preservation in the face of dimensionality reduction. The Learning Vector Quantization (LVQ) is a supervised learning algorithm (classification) and is a zero-order SOM with no connectivity between nodes. Without the connectivity, there is no geometry or topological neighbourhood effects, although the unstructured collection of exemplars compete for stimulation using the WTA principle, and winners that misclassify during the training process are suppressed.

A problem with the winner-take-all learning principle is that units may win too much, and thus dominate selection and response. The result is that some units may never win and the units are considered dead (resources are underutilized) [8]. A relatively simple solution is to dampen the competitiveness of the principle by keeping track of how often units win competitions and using this information to reduce the chance of wining in the future. This may be achieved by introducing a conscience to the units [6], or a similar method called frequency-sensitive competitive learning [41]. This effect may also be used by employed by using a generalized version of the suppression mechanism of LVQ2 called *rival-penalized competitive learning*, where the second winner (rival) in each competition is suppressed [22].

B. *Adaptive Resonance and Counterpropagation Networks*

Grossberg with his Adaptive Resonance Theory (ART) [38,42,43], along with von der Malsburg [4] were among the first to propose theories of self-organization and competitive learning of neural cells. Grossberg exploited ART as the basis of a network architecture and unsupervised competitive learning model [17,44]. The model addressed the instability of unsupervised competitive learning in the face of unexpected irregularities in the input signals. The ART models (ART1, ART2, etc.) are in effect self-regulating control structures for autonomous learning and recognition.

Hecht-Nielsen combined Kohonen's feature map with Grossberg's outstar network structure calling it a counterpropagation network (CPN). This network produces a mapping (like backpropagation) that functions as a statically optimal self-programmable look-up table [35,36]. Further, the network has a number of layers, which include an unsupervised Kohonen hidden layer for self-organizing patterns, and a fully connected Grossberg output layer for pattern classification. The hidden layer facilitates a topology-preserving clustering of input patterns, the results of which are fed to a supervised output layer that associates winning stimulus from the hidden layer with output states.

C. *Neural Gas*

Competitive Hebbian Learning (CHL) is a combination of competitive learning and Hebbian-style learning in which nodes specialise their response to inputs [33,34,37]. Martinetz employed competitive Hebbian-learning rule to construct topology preserving graphs by inserting edges between nearest-neighbour nodes distributed across feature space based on input signals from the domain [51].

The Neural Gas (NG) algorithm of Martinetz and Schulten, like SOM is another self-organizing and unsupervised competitive learning algorithm [50,54,55]. Further, unlike SOM (and more like LVQ), the nodes are not organized into a lower-dimensional structure, instead the competitive Hebbian-learning like rule is applied to connect, order, and adapt nodes in feature space. Martinetz calls this the '*winner-take-most*' learning rule distinct from Kohonen's '*winner-take-all*' rule. The result of the NG algorithm is a set of points that are distributed across feature space in relatively proportion to the input signal density, with a graph topology that preserves features in the feature space.

Fritzke extends the NG algorithm and exploits the competitive Hebbian-learning for constructing topology preserving graphs to propose a growing neural gas algorithm, that in addition to adding and removing edges, is able to apply a similar process to add and remove nodes of the network [3].

D. *Mixture of Experts*

Input signals may be partitioned in a natural way (different tasks or sources) and specialised models (experts) may be prepared for each partition. A gating network may be used to decide how to partition input signals and ultimately learns which expert to use for each input signal. Jacobs and Jordan propose that local experts may compete with each other to contribute to an output signal in such away that the winners of the competition minimize output error [32]. This mixture of experts (ME) approach is further generalized to a hierarchical mixture of experts (HME) [27]. The author's propose that such supervised mixture of expert architectures provide a middle ground for competitive learning and associative (mapping approaches like backpropagation) learning paradigms.

IV. Discussion

The clonal selection theory of acquired immunity [14-16] may be abstracted to a general algorithmic process where a population of units is repeatedly exposed to an antigen to which a high-affinity clone is raised through a process of affinity maturation via hypermutation [23,24]. This section discusses aspects of a general clonal selection approach in the context of lazy and competitive learning, highlighting some broad observations.

1. *Discrete Units*: Lymphocyte cells represent discrete instances from the perspective of lazy learning or exemplar/neurons from the perspective of competitive learning. A random-based guessing strategy is used to initialise the system, and is used for ongoing (online) guessing. This randomness facilitates a non-parametric basis to the learning achieved by this method.

2. *Lazy Adaptation*: The clonal selection approach does not perform any adaptation until the arrival of an antigen stimulus (deferred adaptation). The reason for this is that the system responds to the threat represented by the antigen, and an aspect of the neutralisation of that thread is a general mechanism for improving future detection. This delayed activity is similar to the stimulation-response property of run-time model construction performed in instance-based learning approaches.

3. *Case-Based Reasoning*: The acquired immune response is based on the idea of responding better in the future to those antigens that were detected and responded to in the past. This is a classic case-based reasoning approach to learning, and fits neatly into the 4-R's definition (retrieve, reuse, revise, retain). In addition, the amount of refinement to cases is proportional to the repeat exposures of an antigen.

4. *Competition*: Lymphocytes compete with each other for binding to an antigen. This process has stochastic aspects (the units are mixed in a diffuse substrate), and deterministic aspects (physics of the chemical bonding). Generally, the population is scanned and the highest affinity receptor wins (is proliferated and adapted). Clonal selection is an application of the winner-take-all principle.

5. *Global-Local Competition*: Lymphocytes are distributed throughout the host, although are clustered in sites of high-probability to exposure, such as secondary lymphoid tissue. It is this secondary lymphoid tissue where it is typical for the clonal selection process to take place (germinal centres). Thus, the specific pool of lymphocyte that an antigen is exposed to has a stochastic element, although once in the pool, the competition between units is relatively global. In addition, lymphocytes are mixed across spatial regions via recirculation and homing.

A. *Clonal Selection as Competitive Learning*

The final point has some interesting implications in the context of the competitive learning approaches discussed. This section highlights some aspects of clonal selection as a competitive learning approach. This discussion considers clonal selection across the whole host in the context of differentiated cell types, tissue types, and cell migration.

Self-Organization of Densities: Both the SOM and NG are globally competitive, although the SOM organizes with regard to an arbitrary lower-dimensional structure, and the NG organizes with regard to the topology of the feature space. One may envisage a clonal selection method organizing based on receptor densities, firstly with regard domination of a discrete repertoire (spatial density), and secondarily to the whole system as the response diffuses (global population densities). A discrete repertoire may be considered to facilitate a sampling of units of the system for an antigen, where only the samples in the repertoire compete for the antigen, rather than all lymphocytes of the system.

Inhibition by Attrition: A second interesting point is that the inhibitory effect of clonal selection is attrition and dominance not only of the competition, but also with regard to resource allocation. Cells have a finite lifetime before they are discarded, and the cell population is continually turned-over. Thus, those receptors that are not used, are not preserved in memory (memory cells) and are lost.

Domain-guided Convergence: The competitive learning approaches discussed employ a fixed convergence schedule (learning rate decreased over a fiexed number of iterations). Innate in the clonal selection approach is the need for the domain to define such a schedule. The system is reflexive in this regard, in that it will do as much adaptation with whatever it is exposed to, when it is exposed to it.

Function Optimization as Approximation: Clonal selection is quintessentially considered as a function optimization problem with a single antigen and a collection of receptors on an affinity landscape. A system perspective generalizes this to *n*-landscapes, where *n* is defined by the antigenic environment. Thus the problem is more naturally considered multi-modal, or multi-objective. Competitive learning is typically applied to function approximation where the features of interest in the representation space are preserved. Thus, one may seek to maintain multiple features, multiple objectives concurrently as an approximation of some higher-dimensional surface.

Clonal selection is a method that is strongly associated with hill-climbing behaviour, optimization, and classifications. It is clear that the method is a natural fit for the competitive learning paradigm, given that the general procedure of antigen-selecting-receptor is inherently competitive. Work remains primarily in (1) phrasing a competitive learning clonal selection algorithm, and (2) investigating related hill-climber based competitive learning algorithms. Two examples of potentially related algorithms include the following:

1) Learning Classifier Systems (LCS) in which a genetic algorithm is used to adapt a population of rules [31].
2) Population-Based Incremental Learning (PBIL) in which a genetic algorithm is reduced to a single probabilistic vector. The approach makes strong claims of similarity to supervised competitive learning like LVQ [39,40].

ACKNOWLEDGMENTS

Tim Hendtlass for his patience and for providing useful feedback on drafts of this paper

REFERENCES

[1] A. Aamodt and E. Plaza, Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches *Artificial Intelligence Communications*, vol. 7, pp. 39-59, 1994.

[2] B. Fritzke, "Some competitive learning methods," Systems Biophysics, Institute for Neural Computation, Ruhr-Universität Bochum, Germany, Apr 1997.

[3] Bernd Fritzke, "A Growing Neural Gas Network Learns Topologies," *Advances in Neural Information Processing Systems 7,* pp. 625-632, 1995.

[4] Christoph von der Malsburg, Self-organization of orientation sensitive cells in the striate cortex *Biological Cybernetics*, vol. 14, pp. 85-100, Dec, 1973.

[5] Christopher G. Atkeson, Andrew W. Moore, and Stefan Schaal, Locally Weighted Learning *Artificial Intelligence Review*, vol. 11, pp. 11-73, Feb, 1997.

[6] D. DeSieno, "Adding a conscience to competitive learning," *IEEE International Conference on Neural Networks,* San Diego, CA, USA, pp. 117-124, 1988.

[7] D. S. Broomhead and David Lowe, "Radial Basis Functions, Multi-Variable Functional Interpolation and Adaptive Networks," Royal Signals and Radar Establishment Malvern, England, ADA196234, 1988.

[8] David E. Rumelhart and David Zipser, Feature discovery by competitive learning *Cognitive Science*, vol. 9, pp. 75-112, 1985.

[9] David W. Aha , "Case-Based Learning Algorithms," *Proceedings of the 1991 DARPA Case-Based Reasoning Workshop,* Los Altos, CA, pp. 147-158, 1991.

[10] David W. Aha, Artificial Intelligence Review *Editorial: Lazy Learning*, vol. 11, pp. 7-10, Feb, 1997.

[11] David W. Aha. Feature Weighting for Lazy Learning Algorithms. In: *Feature Extraction, Construction and Selection: A Data Mining Perspective*, eds. H. Liu and H. Motoda. Norwell MA, USA: Kluwer, 1998.

[12] David W. Aha, Dennis Kibler, and Marc K. Albert, Instance-based learning algorithms *Machine Learning*, vol. 6, pp. 37-66, Jan, 1991.

[13] Dietrich Wettschereck, David W. Aha, and Takao Mohri, A Review and Empirical Evaluation of Feature Weighting Methods for a Class of Lazy Learning Algorithms *Artificial Intelligence Review*, vol. 11, pp. 273-314, Feb, 1997.

[14] Frank Macfarlane Burnet, A modification of Jerne's theory of antibody production using the concept of clonal selection *Australian Journal of Science*, vol. 20, pp. 67-69, 1957.

[15] Frank Macfarlane Burnet. *The clonal selection theory of acquired immunity*, Nashville, Tennessee, U.S.A.: Vanderbilt University Press, 1959.

[16] Frank Macfarlane Burnet, Clonal selection and after *Theoretical Immunology*, vol. pp. 63-85, 1978.

[17] G. A. Carpenter and S. Grossberg, The ART of adaptive pattern recognition by a self-organizing neural network *Computer*, vol. 21, pp. 77-88, Mar, 1988.

[18] George F. Luger and William A. Stubblefield. *Artificial Intelligence: Structures and Strategies for Complex Problem Solving*, USA: Addison Wesley Longman Inc, 1997.

[19] I. Watson, Case-based reasoning is a methodology not a technology *Knowledge-Based Systems* , vol. 12, pp. 303-308, Oct, 1999.

[20] Kevin Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft, "When Is "Nearest Neighbor" Meaningful?," *Proceedings of the 7th International Conference in Database Theory (ICDT'99),* Jerusalem, Israel, pp. 217, 1998.

[21] L. Brieman, J. H. Friedman, R. A. Olshen, and C. J. Stone . *Classification and Regression Trees* , Monterrey CA, USA: Wadsworth & Brooks, 1984.

[22] L. Xu, A. Krzyzak, and E. Oja, Rival penalized competitive learning for clustering analysis, RBFnet, and curve detection *IEEE Transactions on Neural Networks*, vol. 4, pp. 636-649, Jul, 1993.

[23] Leandro N. de Castro and Fernando J. Von Zuben, Learning and optimization using the clonal selection principle *IEEE Transactions on Evolutionary Computation*, vol. 6, pp. 239-251, Jun, 2002.

[24] Leandro N. de Castro and Jon Timmis. *Artificial Immune Systems: A new computational intelligence approach*, Great Britain: Springer-Verlag, 2002.

[25] Léon Bottou and Vladimir Vapnik, Local learning algorithms *Neural Computation*, vol. 4, pp. 888-900, Nov, 1992.

[26] M. D. Buhmann, Radial basis functions *Acta Numerica*, vol. 9, pp. 1-38, 2000.

[27] M. I. Jordan and R. A. Jacobs, "Hierarchical mixtures of experts and the EM algorithm," *Proceedings of 1993 International Joint Conference on Neural Networks (IJCNN '93)*, Nagoya, Japan, pp. 1339-1344, 1993.

[28] M. J. D. Powell. Radial basis functions for multivariable interpolation: a review. In: *Algorithms for approximation*, Anonymous New York, NY, USA: Clarendon Press, 1987.pp. 143-167.

[29] M. J. L. Orr , "Introduction to radial basis function networks," Centre for Cognitive Science, University of Edinburgh, Scotland, Apr 1996.

[30] N. Intrator and S. Edelman, Competitive learning in biological and artificial neural computation *Trends in Cognitive Sciences*, vol. 1, pp. 268-272, Oct, 1997.

[31] Pier Luca Lanzi. *Learning Classifier Systems: From Foundations to Applications*, Berlin / Heidelberg: Springer, 2000.

[32] R. A. Jacobs, M. I. Jordan, S. J. Nowlan, and G. E. Hinton, Adaptive mixtures of local experts *Neural Computation*, vol. 3, pp. 79-87, 1991.

[33] R. H. White, "Competitive Hebbian learning," *Proceedings of the International Joint Conference on Neural Networks (IJCNN-91)*, Seattle, USA, pp. 949 , 1991.

[34] R. H. White , "Recurrent competitive Hebbian learning," *Proceedings of the International Joint Conference on Neural Networks (IJCNN),* Baltimore, MD, USA, 1992.

[35] R. Hecht-Nielsen, Counterpropagation Networks *Applied Optics*, vol. 26, pp. 4979-4984, 1987.

[36] R. Hecht-Nielsen, "Counterpropagation Networks," *Proceedings of the IEEE International Conference on Neural Networks,* New York, USA, 1987.

[37] Ray H. White, Competitive Hebbian learning: algorithm and demonstrations *Neural Networks* , vol. 5, pp. 261-275, 1992.

[38] S. Grossberg, Adaptive pattern classification and universal recoding: I. Parallel development and coding of neural feature detectors *Biological Cybernetics*, vol. 23, pp. 121-134, Sep, 1976.

[39] Shumeet Baluja, "Population-Based Incremental Learning: A Method for Integrating Genetic Search Based Function Optimization and Competitive Learning," Carnegie Mellon University , Pittsburgh, PA, USA, Technical Report: CS-94-163, Jun 1994.

[40] Shumeet Baluja and R. Caruana, "Removing the genetics from the standard genetic algorithm ," *In the Proceedings of the International Conference on Machine Learning,* pp. 38-46, 1995.

[41] Stanley C. Ahalt, Ashok K. Krishnamurthy, Prakoon Chen, and Douglas E. Melton, Competitive learning algorithms for vector quantization *Neural Networks*, vol. 3, pp. 277-290, 1990.

[42] Stephen Grossberg, Neural expectation: cerebellar and retinal analogs of cells fired by learnable or unlearned pattern classes *Biological Cybernetics*, vol. 10, pp. 49-57, Jan, 1972.

[43] Stephen Grossberg, Adaptive pattern classification and universal recoding: II. Feedback, expectation, olfaction, illusions *Biological Cybernetics*, vol. 23, pp. 187-202, Dec, 1976.

[44] Stephen Grossberg, Competitive learning: from interactive activation to adaptive resonance *Cognitive Science*, vol. 11, pp. 23-63, 1988.

[45] Stuart J. Russell and Peter Norvig. *Artificial Intelligence: A Modern Approach*, Prentice Hall, 1995.

[46] T. Cover and P. Hart, Nearest neighbor pattern classification *IEEE Transactions on Information Theory*, vol. 13, pp. 21-27, Jan, 1967.

[47] T. Kohonen, "Automatic formation of topological maps of patterns in a self-organizing system," *Proceedings of the 2nd Scandinavian Conference on Image Analysis,* Espoo, Finland, pp. 214-220, 1981.

[48] T. Kohonen, "Clustering, taxonomy, and topological maps of patterns," *Proceedings of the 6th International Conference on Pattern Recognition,* Munich, Germany, pp. 114-128, 1982.

[49] T. Kohonen, The self-organizing map *Proceedings of the IEEE*, vol. 78, pp. 1464-1480, Sep, 1990.

[50] T. M. Martinetz, S. G. Berkovich, and K. J. Schulten, 'Neural-gas' network for vector quantization and its application totime-series prediction *IEEE Transactions on Neural Networks*, vol. 4, pp. 558-569, Jul, 1993.

[51] T. Martinetz, "Competitive Hebbian learning rule forms perfectly topology preserving maps," *Proceedings of the International Conference on Artificial Neural Networks (ICANN'93),* pp. 427-434, 1993.

[52] Teuvo Kohonen, Self-organized formation of topologically correct feature maps *Biological Cybernetics*, vol. 43, pp. 59-69, Jan, 1982.

[53] Teuvo Kohonen. *Self-Organizing Maps*, Berlin Heidelberg: Springer-Verlag, 2001.

[54] Thomas Martinetz and Klaus Schulten, "A "neural gas" network learns topologies," *Artificial Neural Networks,* pp. 397-402, 1991.

[55] Thomas Martinetz and Klaus Schulten, Topology representing networks *Neural Networks*, vol. 7, pp. 507-522, 1994.