Appendix 1

Equations and Architecture of the Auto Contractive Map Neural Network

Networks Topology

The Auto Contractive Map(Auto-CM) neural network was designed by M Buscema in 1999 and its learning law was improved up to 2018.

Auto-CMhas an architecture based on three layers of nodes (Figure 1): an input layer that captures the signal from the environment, a hidden layer which modulates the signal within the network, and an output layer which returns a response to the environment on the basis of the processing that occurred. The three layers have the same number N of nodes. The connections between the input layer and the hidden one are mono-dedicated, whereas those between this hidden layer and the output layer are completely connected. Each connection is assigned a weight: v_i for connections between the i^{th} input node and the corresponding hidden node, $w_{i,j}$ for those between the generic j^{th} node of the hidden layer and the i^{th} node of the output layer.

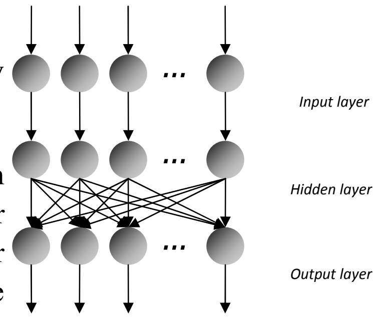

Figure 1 Architecture of the Auto-CM

Equations for Learning

For the training, datasets are scaled between zero and one and all weights initialized beforehand to the same positive value close to zero. Then the network must undergo a series of epochs. In each of them, all the input patterns must be presented one after another to the network, and a calculation made for the appropriate equations with the corresponding output value and a measure of error with respect to the desired value. In accordance with the principle of batch update, the corrections accumulated for an epoch must be applied at the end. The equations of training of the network make reference to the quantities shown below (Table 1).

Symbol	*Meaning*
x_i^p	*i^{th}input node of the p^{th} pattern*
$h_i^p(n)$	*i^{th} hidden node of the p^{th} pattern during the n^{th} time*
$y_i^p(n)$	*i^{th}node in the output of the p^{th} pattern during the n^{th} epoch*
$v_i(n)$	*weight of the connection between the i^{th}input node and in the i^{th}hidden node during the n^{th} epoch*
$w_{i,j}(n)$	*weight of the connection between the j^{th} hidden node and the i^{th}output node during the n^{th} epoch*
N	*the number of nodes per layer*
M	*the number of patterns*
α	*constant learning rate*
C	*constant greater than one, typically* $C = \sqrt{N}$

Table 1 Notation for Auto-CM neural network

At the n^{th} epoch of training, out of each input pattern a value is calculated for the hidden layer, through a contraction, that reduces the input value in proportion to the mono-dedicated weight.

$$h_i^{[p]}(n) = x_i^{[p]} \cdot \left(1 - \frac{v_i(n)}{C}\right) \tag{1}$$

The algorithm then calculates the value on the output layer through a "double conceptual passage." For each output node, an initial operation saves the net input calculation, that is to say, the reduction (contraction) of all the hidden nodes through the weights between the hidden layer and output layer (Equation 2).

$$Net_i^{[p]}(n) = \sum_{j=1}^{N} h_j^{[p]}(n) \cdot \left(1 - \frac{w_{i,j}(n)}{C}\right) \tag{2}$$

A second operation calculates the output value by further contracting the corresponding value of the hidden node thorough the previously calculated net input for the output node:

$$y_i^{[p]}(n) = h_i^{[p]}(n) \cdot \left(1 - \frac{Net_i^{[p]}(n)}{C^2}\right) \tag{3}$$

During the training that occurs in every epoch, in addition to the calculation of the output values (3), for each pattern presented in input the algorithm calculates the correction quantity of the weights, summed and applied at the end of the epoch. For the N-mono dedicated layers between the input and hidden layers, the algorithm considers the contraction, based on the weight being examined, of the difference between the values of the corresponding input and hidden nodes, further modulated for the input node itself.

$$\Delta v_i(n) = \sum_{p=1}^{M} \left(x_i^{[p]} - h_i^{[p]}(n)\right) \cdot \left(1 - \frac{v_i(n)}{C}\right) \cdot x_i^{[p]} \tag{4}$$

$$v_i(n+1) = v_i(n) + \alpha \cdot \Delta v_i(n) \tag{5}$$

Similarly, for N^2 weights between the hidden and output layers the algorithm calculates the contraction, based on the weight being considered, between the corresponding hidden and output nodes. The Learning coefficient (α)is updated according to the average of the errors of each weights (δ) at each epoch (*n*)

$$\Delta w_{i,j}(n) = \sum_{p=1}^{M} \left(h_i^{[p]}(n) - y_i^{[p]}(n)\right) \cdot \left(1 - \frac{w_{i,j}(n)}{C}\right) \cdot h_j^{[p]}(n) \tag{6}$$

$$w_{i,j}(n+1) = w_{i,j}(n) + \alpha \cdot \Delta w_{i,j}(n) \tag{7}$$

$$\delta_{(n)} = \frac{1}{N^2} \sum_{i}^{N} \sum_{j}^{N} \Delta w_{i,j_{(n)}}; \tag{8}$$

$$\alpha_{(n+1)} = \alpha_{(n)} + \frac{e^{\delta_{(n)}}}{n}. \tag{9}$$

Cost Function

The quantity, E, to minimize during the learning process is the following:

$$E = \sum_{p=1}^{M} \sum_{i=1}^{N} \left(y_i^{[p]} - h_i^{[p]} \right). \tag{8}$$

From the equations one can immediately observe how the contractions establish a relationship of order between the layers:

$$x_i^{[p]} \geq h_i^{[p]}(n) \geq y_i^{[p]}(n) \tag{9}$$

One can easily see during training, the mono-dedicated weights v_i grow monotonically, and with different speeds asymptotically towards the constant C:

$$\lim_{n\to\infty} \Delta v_i(n) = 0 \tag{10}$$

$$\lim_{n\to\infty} v_i(n) = C \tag{11}$$

just like the values of hidden nodes tend to cancel themselves out:

$$\lim_{n\to\infty} h_i^{[p]}(n) = 0 \tag{12}$$

along with those of the output units:

$$\lim_{n\to\infty} y_i^{[p]}(n) = 0 \tag{13}$$

while the corrections of the full set of weights diminish:

$$\lim_{n\to\infty} \Delta w_{j,i}(n) = 0 \tag{14}$$

The process of canceling the above quantity occurs with speed modulated by the input patterns and leaves its specific sign in the matrix between the hidden and the output layer.

References:

[1] Buscema M, A novel adapting mapping method for emergent properties discovery in data bases:experience in medical field, in “2007 IEEE International Conference on Systems, Man and Cybernetics (SMC 2007)”. Montreal, Canada, 7-10 Ottobre 2007.

[2] Buscema M., Grossi E., The Semantic Connectivity Map: an adapting self-organizing knowledge discovery method in data bases. Experience in Gastro-oesophageal reflux disease, Int. J. Data Mining and Bioinformatics, Vol. 2, No. 4, 2008.

[3] Buscema M., Grossi E., Snowdon D., Antuono P., Auto-Contractive Maps: an Artificial Adaptive System for Data Mining. An Application to Alzheimer Disease, in Current Alzheimer Research, 2008, 5, 481-498.

[4] Buscema M.(ed), Squashing Theory and Contractive Map Network, Semeion Technical Paper #32, Rome, 2007.

[5] Buscema M, Helgason C, Grossi E, Auto Contractive Maps, H Function and Maximally Regular Graph: Theory and Applications , Special Session on "Artificial Adaptive Systems in Medicine : applications in the real world, NAFIPS 2008 (IEEE), New York, May 19-22, 2008.

[7] Licastro F, Porcellini E, Chiappelli M, Forti P, Buscema M et al., Multivariable network associated with cognitive decline and dementia, int Neurobiology of Aging, Vol. 1, Issue 2, February 2010, 257-269.

[8] M Buscema and E Grossi (eds), Artificial Adaptive Systems in Medicine, Bentham e-books, 2009, 25-47.

[9] M Buscema and PL Sacco, Auto-contractive Maps, the H Function, and the Maximally Regular Graph (MRG): A New Methodology for Data Mining, in V. Capecchi et al. (eds.), Applications of Mathematics in Models, Artificial Neural Networks and Arts, Chapter 11, DOI 10.1007/978-90-481-8581-8_11, Springer Science+Business Media B.V. 2010.

[10] E Grossi, GT Blessi, PL Sacco, M Buscema, The Interaction Between Culture, Health and Psychological Well-Being: Data Mining from the Italian Culture and Well-Being Project, J Happiness Studies, Springer, 2011

[11] F Licastro, E Porcellini, P Forti, M Buscema, I Carbone, G Ravaglia, E Grossi, Multi factorial interactions in the pathogenesis pathway of Alzheimer's disease: a new risk charts for prevention of dementia, Immunity & Ageing 2010, 7(Suppl 1):S4.

[12] M Buscema, F Newman, E Grossi, W Tastle, Application of Adaptive Systems Methodology to Radiotherapy, in NAFIPS 2010, 12-14 July, Toronto, Canada.

[13] C Eller-Vainicher, V V Zhukouskaya,Y V Tolkachev, S S Koritko, E Cairoli, E Grossi, P Beck-Peccoz, I Chiodini, A P Shepelkevich, Low BoneMineral Density and Its Predictors in Type 1 Diabetic Patients Evaluated by the Classic Statistics and Artificial Neural Network Analysis, Diabetes Care , pp 1-6, 2011.

[14] T Gomiero, L Croce, E Grossi, L De Vreese, M Buscema, U Mantesso, E De Bastiani, A Short Version of SIS (Support Intensity Scale): The Utility of the Application of Artificial Adaptive Systems, US-China Education Review A 2 (2011) 196-207.

[15] M Buscema, S Penco and E Grossi, A Novel Mathematical Approach to Define the Genes/SNPs Conferring Risk or Protection in Sporadic Amyotrophic Lateral Sclerosis Based on Auto Contractive Map Neural Networks and Graph Theory, Neurology Research International, Volume 2012, Article ID 478560, 13 pages, doi:10.1155/2012/478560

[16] E Grossi, A Compare, M Buscema, The concept of individual semantic maps in clinical psychology: a feasibility study on a new paradigm, Quality & Quantity International Journal of Methodology , August 04th 2012, ISSN 0033-5177, Qual Quant, DOI 10.1007/s11135-012-9746-8

[17] F Coppedè, E Grossi, M Buscema, L Migliore, Application of Artificial Neural Networks to Investigate One-Carbon Metabolism in Alzheimer's Disease and Healthy Matched Individuals, PLOS ONE, www.plosone.org, August 2013, Volume 8 , Issue 8 , e74012, pp 1-11.

[18] , M.E. Street, M. Buscema, A. Smerieri, L. Montanini, E. Grossi, 2. Artificial Neural Networks, and Evolutionary Algorithms as a systems biology approach to a data-base on fetal growth restriction, in Prog Biophys Mol Biol., 2013, July, pages 1-6.

[19] A. Compare, E. Grossi, M. Buscema, C. Zarbo, X. Mao, F. Faletra, E. Pasotti, T.Moccetti, P.M. C. Mommersteeg, and A. Auricchio, Combining Personality Traits with Traditional Risk Factors for Coronary Stenosis: An Artificial Neural Networks Solution in Patients with Computed Tomography Detected Coronary Artery Disease, Cardiovascular Psychiatry and Neurology, Volume 2013, Article ID 814967, 9 pages, Hindawi Publishing Corporation. http://dx.doi.org/10.1155/2013/814967.

[20] M. Buscema V. Consonni , D. Ballabio , A.Mauri , G. Massini , M. Breda , R. Todeschini , K-CM: A new artificial neural network. Application to supervised pattern recognition, Chemometrics and Intelligent Laboratory Systems 138 (2014) 110–119.

[21] M Buscema,G Massini and G Maurelli, Artificial Neural Networks: An Overview and their Use in the Analysisof the AMPHORA-3 Dataset, Substance Use & Misuse, Early Online:1–14, 2014.

[22] M Gironi, B Borgiani, E Farina, E Mariani, C Cursano, M Alberoni, R Nemni, G Comi, M Buscema, R Furlan, and Enzo Grossi, A Global Immune Deficit in Alzheimer’s Disease and Mild Cognitive Impairment Disclosed by a Novel Data Mining Process, Journal of Alzheimer’s Disease 43 (2015) 1199–1213.

[23] F Drenos, E Grossi, M Buscema, S E Humphries, Networks in Coronary Heart Disease Genetics As a Step towards Systems Epidemiology, PLoS ONE 10(5): May 7 (2015). e0125876. doi:10.1371/journal.pone.0125876.

[24] F Coppedè, E Grossi, A Lopomo, R Spisni, M Buscema & Lucia Migliore, Application of artificial neural networks to link genetic and environmental factors to DNA methylation in colorectal cancer, Epigenomics (2015) 7(2), 175–186.

[25] A Narzisi, F Muratori, M Buscema, S Calderoni, E Grossi, Outcome predictors in autism spectrum disorders preschoolers undergoing treatment as usual: insights from an observational study using artificial neural networks, Neuropsychiatric Disease and Treatment 2015:11 1587–1599.

[26] M Buscema, E Grossi, L Montanini, M E Street, Data Mining of Determinants of Intrauterine Growth Retardation Revisited Using Novel Algorithms Generating Semantic Maps and Prototypical Discriminating Variable Profiles, PLoS ONE 10(7): June 9 (2015) e0126020. doi:10.1371/journal.

[27] PM Buscema, L Gitto, S Russo, A Marcellusi, F Fiori, G Maurelli, G Massini &FS Mennini, The perception of corruption in health: AutoCM methods for an international comparison, Qual Quant DOI 10.1007/s11135-016-0315-4, Springer 2016.

[28] M Buscema, G Ferilli, PL Sacco, What kind of ‘world order’? An artificial neural networks approach to intensive data mining, Technological Forecasting & Social Change 117 (2017) 46–56

[29] G. Ferilli, PL Sacco, E Teti, M Buscema, Top corporate brands and the global structure of country brand positioning: An AutoCM ANN approach, Expert Systems With Applications 66 (2016) 62–75.

[30] M Buscema, PL Sacco, MST Fitness Index and implicit data narratives: A comparative test on alternative unsupervised algorithms, Physica A 461 (2016) 726–746.

[31] FS Mennini, Lara Gitto, S Russo, A Cicchetti, M Ruggeri, S Coretti, G Maurelli, PM Buscema, Does regional belonging explain the similarities in the expenditure determinants of Italian healthcare deliveries? An approach based on Artificial Neural Networks, Economic Analysis and Policy 55 (2017) 47–56.