A Novel Complex-Valued Fuzzy ARTMAP for Sparse Dictionary Learning

Chu Kiong Loo, Ali Memariani, and Wei Shiung Liew

Faculty of Computer Science and Information Technology
University of Malaya
Kuala Lumpur, Malaysia
ckloo.um@um.edu.my, ali_memariani@siswa.um.edu.my,
liew.wei.shiung@gmail.com

Abstract. This work extends the simplified fuzzy ARTMAP (SFAM) to a complex-valued (CV-SFAM) version which is able to work with spatio-temporal data produced in receptive fields of visual cortex. The CV-SFAM's ability for incremental learning distinguishes CV-SFAM from other complex-valued neural networks, which provides the ability to preserve learned data while learning new samples. We considered different scales and orientations of Gabor wavelets to form a dictionary. This work takes advantage of a locally competitive algorithm (LCA) which calculates more regular sparse coefficients by combining the interactions of artificial neurons. Finally, we provide an experimental real application for biological implementation of sparse dictionary learning to recognize objects in both aligned and non-aligned images.

Keywords: complex-valued simplified fuzzy ARTMAP, sparse coding, dictionary learning, genetic optimization, body expression.

1 Introduction

Complex-valued neural networks (CVNN) are a type of neural nets dealing with complex-valued information using complex-valued variables and parameters [3]. They are variety of fields in which the CVNNs provide proper information representations. So far most of the applicable fields are related to wave phenomena, e.g., measurements and communications using waves such as radar image processing, quantum computation, learning electron-wave devices, ultrasonic imaging and so on. The wavelength-dependent dynamics of optical circuit leads to adaptive optical routers in optical wavelength-division-multiplexed communications, variable optical connections, frequency-domain parallel information processing, etc. The carrier-frequency-dependent neural behavior realizes both the adaptability and controllability in neural networks [3].

CVNNs were first introduced by Aizenberg as phasor where amplitude was fixed [5]. Multiple-valued associative memory is developed with memory capacity [6]. Many works attempted to derived complex form of linear models using complex linear systems, e.g. complex steepest descent [7] [8] and backpropagation

M. Lee et al. (Eds.): ICONIP 2013, Part I, LNCS 8226, pp. 360–368, 2013.

learning [9]. Complex-valued activation functions are discussed in both separate and integrated real-imaginary forms [10] [11].

However, the previous work of CVNNs are poor in preserving learned data after new learning arrivals which causes the stability-plasticity dilemma [12]. One approach to face with this is using incremental learning as in fuzzy ARTMAP [18] and fuzzy min-max [14] neural networks.

Simplified fuzzy ARTMAP (SFAM) is a modification of fuzzy ARTMAP [13] as a two-layer neural network which has been successfully used for pattern recognition [15] [16] [17]. Adjustable learning rate and incremental learning capability makes it suitable for real-time applications. With fast learning rates its computational complexity could be compared to a multi-layered perceptron [18]. Furthermore, neural structure of SFAM combined with fuzzy operators, models a human-like behavior [2] [15]. However, using real-valued input features, SFAM is not able to be used in most of the vision applications in which data is intrinsically complex. Complex fuzzy set theory has been proposed in by Ramot et al. [19] [20]. This work applies complex fuzzy operators in order to develop a novel complex-valued SFAM (CV-SFAM).

We applied the proposed model on patterns of different objects in images. Patterns are extracted using a sparse dictionary learning algorithm which is inspired by experimental findings of functional magnetic resonance imaging (fMRI) of mammalian brain. Learning method is tested in presence of noise; since SFAM networks are sensitive to noise [17] learned images are also filtered during the sparse coding. Classification results are compared to the state of the art algorithms. Furthermore, we applied the algorithm to recognize emotions based on body expression data which is inspired by the action based behavior in psychology. Classification results are compared to those of human recognitions.

1.1 Complex Fuzzy Sets

Classical fuzzy logic applies real-valued functions to represent the membership; in order to show the phase term another dimension is added to the membership function [16] [19] [20] [21]. Though the fuzziness remains a real value in range [0,1]. It should be noted that the concept of using complex input numbers into a fuzzy inference system is different from applying fuzzy number into a system of inference. The latter introduced as a fuzzy set with real-valued features as input and complex-valued in the output. However, in most of the computer vision problems input images are preprocessed with complex transforms (Fourier and windowed Fourier families) which provides the complex data as features for machine learning algorithms.

A complex fuzzy set allows membership functions in a set to be specified by a complex number. Membership function μ of any element x, in a complex fuzzy set S, is replaced by complex-valued grade of membership of the general form:

$$\mu(x) = r_s(x).e^{iw\mu(x)} \tag{1}$$

where $r_s(x) \in [0, 1]$. The amplitude term retains the traditional notion of "fuzziness", by the representation of membership of variable x of set S, same as original fuzzy set theory amplitude term ranges between 0 to 1 and with a phase term zero μ will go back to its original fuzzy definition. The phase term denotes the assertion of multiple-valued complex fuzzy set theory and represent the lag of time in a time frequency domain.

Given two complex fuzzy sets A and B on U with complex valued membership function $\mu_A(x)$, $\mu_B(x)$ respectively, complex fuzzy union could be defined as [19] [20] [21]:

$$\mu_{A \vee B}(x) = [r_{A(x) \oplus r_{B(x)}}].e^{iW_{A \vee B}(x)} \tag{2}$$

where $\oplus$ represents the t-conrom and $W_{A \vee B}$ is defined as:

$$\text{Sum} : W_{A \vee B} = W_A + W_B \tag{3}$$

$$\text{Max} : W = max(W_A, W_B) \tag{4}$$

$$\text{Winner Take All} : W_{A \vee B} \begin{cases} W_A, rA > rB \\ W_B, rB > rA \end{cases} \tag{5}$$

1.2 Dictionary Learning by Gabor Wavelets

Decomposition methods extract key information from images and videos in order to create a highly-compressed and simplified representation of the original using only a handful of elementary functions. A set of functions is called a dictionary. Representation is thus achieved by linear combination of elements in the dictionary. To improve the quality of representation, a common approach is to use an orthogonal subset of a large dictionary containing all possible elements.

Textons are developed as a mathematical representation of basic image objects [22]. Images are coded by a dictionary of Gabor and Laplacian of Gaussian elements; Responses to the dictionary elements is combined by transformed component analysis. Furthermore, sparse approximation helps to find a more general object models in terms of scale and posture [23]. Active basis model [1] provides a deformable template using Gabor wavelets as dictionary elements. They also proposed a shared sketch algorithm (SSA) inspired by AdaBoost.

1.3 Sparse Coding Using Artificial Neurons

Response to a dictionary of Gabor wavelets is an overcomplete representation. Sparse coding is the method of selecting a proper subset of responses to represent the image (signal). In addition to biological motivations, sparse coding is necessary to avoid redundant information. Having a fixed number of features, redundancy may cause loss of essential information which is going to be encoded in the lower levels (Figure 1).

Optimum sparse coding minimizes the number of nonzero coefficients, which is an NP-hard optimization problem. We applied a locally competitive algorithm

Fig. 1. Edge detection using Gabor wavelets, A. Original image [1], B. edge detected image with a large number of features without sparsity, C. edge detected image with a small number of features where sparsity is enforced.

(LCA) [4] to enforce local sparsity. Unlike classical sparse coding algorithms, LCA uses a parallel neural structure inspired by biological model. In previous works, there is no real application that has been applied using LCA, although some simulation results are shown. Here an empirical experiment based real application of body expression recognition, is proposed to provide an evidence for the practical utility of Holonomic Brain Model as dictionary learning method by LCA.

Atkinson et al. developed a dataset for both static and dynamic body expressions. The dataset contains 10 subjects (5 female) and covers five emotions (anger, disgust, fear, happiness and sadness)[24]. The bodily expressive action stimulus test (BEAST) [25] provides a dataset for recognizing four types of emotions (anger, fear, happiness, sadness) which is constructed using non-professional actors (15 male, 31 female). Body expressions are validated with a human recognition test.

1.4 Complex-Valued Simplified Fuzzy ARTMAP (CV-SFAM)

CV-SFAM is a two-layer network in which the magnitudes of raw data are scaled to the range of 0 and 1 and further passed through a complement coder. The complement coder normalizes the input vector and stretched to double the size by adding its complement. For an input vector a with length d, the complement coded vector I is [2]:

$$I = (\mu_\nu, \mu_{\bar{\nu}}) = (\mu_{\nu_1}, ..., \mu_{\nu_d}, \mu_{\bar{\nu}_1}, ..., \mu_{\bar{\nu}_d}) \tag{6}$$

$$|I| = |\mu_\nu, \mu_{\bar{\nu}}| = d \tag{7}$$

where the norm $|.|$ is defined as:

$$|a| = \sum_{i=1}^{d} a_i \tag{8}$$

Activation of output nodes are based on their complex fuzzy t-norm (see appendix). Activation function of output node j in response to input node i is:

$$T_{ij} = \frac{u_{I_i \wedge W_{ij}}(v)}{\alpha + |\mu_{w_{ij}}|} \tag{9}$$

where α is a small value greater than zero, and w_{ij} is the weight between nodes i and j. Output nodes form a category according to their complex fuzzy t-conorm:

$$C = \max_{j}(T_j) \tag{10}$$

Match tracking category C is based on the vigilance parameter ρ. The network is in state of resonance if:

$$\text{Match} = \frac{|\mu_{I \wedge W}(v)|}{|\mu_I(v)|} \geq \rho \tag{11}$$

where W is the weight factor for all the output nodes. Otherwise, a mismatch reset happens and the magnitude of T_j is set to zero, and C will be updated using the complex fuzzy t-conorm. The vigilance parameter ρ determines the minimum match for an input to be assigned to the category node C.

Since the norm of the complete coded vector I is constant and equal to the number of features, matching the function could be rewritten as:

$$\text{Match} = \frac{|\mu_{I \wedge W}(v)|}{d} \tag{12}$$

In the state of resonance, updated weights are calculated based on a linear combination of their match and old weights, given the learning rate β:

$$W_{ij}^{new} = \beta(I_i \wedge W_{ij}) + (1 - \beta)W_{ij} \tag{13}$$

We applied a supervised approach to recognize two types of objects in images; First a pixel-wise approach for aligned objects which combines the learned samples of objects in each class to form a prototype and second a feature based approach for non-aligned objects in which Gabor wavelets are localized to represent a potential match between specific scale and orientation and edges of objects. Both approaches are fed into a synergetic neural network to perform a classification task.

2 Proposed Method

2.1 Pixel-Wised Approach

Images are scaled to have the exact same size. Each image is convolved with all the elements in the dictionary. Then sparse coding is enforced to minimize the representing elements for each pixel. Finally, remaining parts are reconstructed to generate the sparse superposition of the image. For pixel values in the local area LCA has the following steps:

1. Compute the response (convolution) of I_i with all the elements in the dictionary.

$$C_j = \langle GW_j, I_j \rangle \tag{14}$$

(Set $t = 0$ and $u_j(0) = 0$, for $j = 1, ..., n$).

2. Determine the active nodes by activity thresholding.
3. For each pixel, calculate the internal state of element j, $u_j(t)$.

$$u_j = \frac{1}{\tau}[C_j(t) - u_j(t) - \sum_{j \neq k} \Phi_{j,k}.a_j(t)] \tag{15}$$

$$\Phi_{j,k} = \langle GW_j.GW_k \rangle \tag{16}$$

4. Compute sparse coefficients $a_j(t)$ for $u_j(t)$.

$$a_j(t+1) = T_\lambda(u_j(t)) \tag{17}$$

$$T_{(\alpha,\gamma,\lambda)}(u_j) = \frac{u_j - \alpha\lambda}{1 + e^{-\gamma(u_j - \lambda)}} \tag{18}$$

5. If $a_j(t-1) - a_j(t) > \delta$, then $t \leftarrow t+1$ and go to step 2, otherwise finish.

2.2 Detecting Objects with Shape Changes

Original SFAM used pixel-wised features to represent an object which is not robust in case objects are in a variable shapes (e.g. different body emotions of human). In this case, we construct a template model as a collection of Gabor wavelet features included in the dictionary which represents the general characteristics of all body posture classes.

Test images are convolved with the components of the template model. Sparsity is then enforced to catch the best fit over the specific posture. LCA thresholding strategy enables us to remove redundancies effectively (producing sparse coefficients with exactly zero values). Number of output Gabor wavelets are fixed in order to make the comparison with trained prototype of each class. Features are selected based on their highest response to the training images; furthermore, each feature is allowed to perturb slightly in terms of location and orientation. In this aspect our template construction is a modification of shared sketch algorithm [1]. For each image i feature value v_{ij} corresponded to the selected Gabor wavelet j, is determined as the following:

$$v_{ij} = \gamma_i C_{ij} - \log(Z(\gamma_i)) \tag{19}$$

where γ_i is derived by maximum likelihood estimation and Z is the partition function. Therefore, boundaries of object are segmented out with before the result is given to CV-SFAM.

3 Experimental Result

Performance of ARTMAP-based classifiers are sensitive to a number of parameters. To achieve optimum recognition rates, a genetic algorithm was employed to optimize parameters such as training sequence and feature subset selection [26] [27].

3.1 Aligned Images

The proposed algorithm is used to classify four classes of animals . Each image used in classification contains an animal object. Image objects are obtained from different animals in variety of postures, and shapes. Gabor wavelets are generated in a (20, 20) matrix and images are resized to (50, 50).

Result of classification is a tag for each image representing the corresponding class. For each class 10 images are selected randomly to form the test set. Remaining images are used for training. Using cross validation classification is performed over all the images. This result is compared to Active Basis Model (ABM). Learning method can effectively detect the edge patterns and represent the main components of objects. This leads to have more distinguishable object definitions (specifically between classes of Bear and Wolf) rather than shared sketching used in ABM.

Table 1. Classification result of aligned images

	Bear	Cat	Cow	Wolf
CV-SFAM	62%	87%	78%	58%
ABM	87%	100%	76%	60%
# of images	60	70	60	50

3.2 Classification of Emotions Using Body Expressions (Non-aligned Images)

We applied the BEAST data set to classify four classes of basic emotions. Gabor wavelets are generated in a (20, 20) matrix and images are resized to have 500 pixels in row and relatively scaled pixels in column. Images are divided into train and test sets for each class 10 images are selected randomly to form the train data and the rest are included for test. Different scenarios are considered to train the model. Classification accuracies of different trained SFAMs are compared with results of human recognition.

Table 2. Classification result of non-aligned images

	Anger	Fear	Happiness	Sadness
Min	93.33	88.00	88.00	100.00
BEAST(human)	93.60	93.90	85.40	97.80

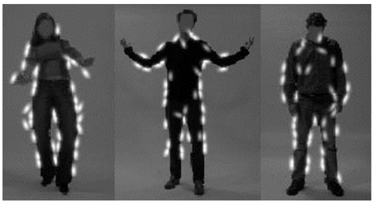

Fig. 2. Features extracted for emotion classes of fear, happiness, and sadness

4 Conclusion

We proposed a biologically-plausible approach for recognition of aligned and non-aligned objects. Our learning algorithm is inspired by the holonomic brain theory. LCA is applied to enforce sparsity on a dictionary of Gabor wavelets. Regarding the parallel structure of the learning method, implementation could be optimized via parallel processing which is essential for real-time applications.

Furthermore, a SFAM neural network is combined with Gabor wavelet features which make it applicable for recognition of non-aligned objects. Gabor features also enhance the model to use images with different size. Effect of background is also removed because of recognition is based on the pattern of edges; Though sparse coding is robust in presence of classical noise since dot noise does not follow any meaningful shape pattern intrinsically.

Experimental results supported the real application of locally competitive sparse coding as a learning method using a biological implementation.

Acknowledgement. This study was funded by UM IPPP research grant FL006-2011 (”Productive Aging Through ICT”), and HIR-MOHE grant H-22001-00-B000010.

References

1. Wu, Y.N., Si, Z., Gong, H., Zhu, S.C.: Learning active basis model for object detection and recognition. International Journal of Computer Vision 90(2), 198–235 (2010)
2. Kasuba, T.: Simplified Fuzzy ARTMAP. AI Experts 8, 18–25 (1993)
3. Hirose, A.: Complex-Valued Neural Networks Fertilize Electronics. In: Hirose, A. (ed.) Complex-Valued Neural Networks, 2nd edn. SCI, vol. 400, pp. 3–8. Springer, Heidelberg (2012)
4. Rozell, C.J., Johnson, D.H., Baraniuk, R.G., Olshausen, B.A.: Sparse coding via thresholding and local competition in neural circuits. Neural Computation 20(10), 2526–2563 (2008)
5. Aizenberg, I., Aizenberg, N.N., Vandewalle, J.P.: Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications. Springer (2000)
6. Noest, A.J.: Discrete-state phasor neural networks. Rapid Communications 38(4), 2196–2199 (1988)
7. Widrow, B., McCool, J., Ball, M.: The complex LMS algorithm. Proceedings of the IEEE 63(4), 719–720 (1975)
8. Hanna, A.I., Mandic, D.P.: A fully adaptive normalized nonlinear gradient descent algorithm for complex-valued nonlinear adaptive filters. IEEE Transactions on Signal Processing 51(10), 2540–2549 (2003)
9. Leung, H., Haykin, S.: The complex backpropagation algorithm. IEEE Transactions on Signal Processing 39(9), 2101–2104 (1991)
10. Kim, M.S., Guest, C.C.: Modification of backpropagation networks for complex-valued signal processing in frequency domain. In: IEEE/INNS International Joint Conference on Neural Networks, pp. 27–31 (1990)
11. Nitta, T.: An extension of the back-propagation algorithm to complex numbers. Neural Networks 10(8), 1391–1415 (1997)

12. Su, M.C., Lee, J., Hsieh, K.L.: A new ARTMAP-based neural network for incremental learning. Neurocomputing 69(16), 2284–2300 (2006)
13. Carpenter, G.A., Grossberg, S., Markuzon, N., Reynolds, J.H., Rosen, D.B.: Fuzzy ARTMAP: A neural network architecture for incremental supervised learning of analog multidimensional maps. IEEE Transactions on Neural Networks 3(5), 698–713 (1992)
14. Simpson, P.K.: Fuzzy min-max neural networks. IEEE Transactions on Neural Networks 3(5), 776–786 (1992)
15. Vakil-Baghmisheh, M.T., Pavesic, N.: A fast simplified fuzzy ARTMAP network. Neural Processing Letters 17(3), 273–316 (2003)
16. Loo, C.K., Rao, M.V.C.: Accurate and reliable diagnosis and classification using probabilistic ensemble simplified fuzzy ARTMAP. IEEE Transactions on Knowledge and Data Engineering 17(11), 1589–1593 (2005)
17. Rajasekaran, S., Pai, G.A.V.: Image recognition using Simplified Fuzzy ARTMAP augmented with a moment based feature extractor. International Journal of Pattern Recognition and Artificial Intelligence 14(8), 1081–1095 (2000)
18. Palaniappan, R., Eswaran, C.: Using genetic algorithm to select the presentation order of training patterns that improves simplified fuzzy ARTMAP classification performance. Applied Soft Computing 9(1), 100–106 (2009)
19. Ramot, D., Milo, R., Friedman, M., Kandel, A.: Complex fuzzy sets. IEEE Transactions on Fuzzy Systems 10(2), 171–186 (2002)
20. Ramot, D., Friedman, M., Langholz, G., Kandel, A.: Complex fuzzy logic. IEEE Transactions on Fuzzy Systems 11(4), 450–461 (2003)
21. Dick, S.: Toward complex fuzzy logic. IEEE Transactions on Fuzzy Systems 13(3), 405–414 (2005)
22. Zhu, S.C., Guo, C.E., Wang, Y., Xu, Z.: What are textons? International Journal of Computer Vision 62(1-2), 121–143 (2005)
23. Figueiredo, M.A.T.: Adaptive sparseness for supervised learning. IEEE Transactions on Pattern Analysis and Machine Intelligence 25(9), 1150–1159 (2003)
24. Atkinson, A.P., Dittrich, W.H., Gemmell, A.J., Young, A.W.: Emotion perception from dynamic and static body expressions in point-light and full-light displays. Perception 33(6), 717–746 (2004)
25. de Gelder, B.V.D., Stock, J.: The Bodily Expressive Action Stimulus Test (BEAST). Construction and Validation of a Stimulus Basis for Measuring Perception of Whole Body Expression of Emotions. Frontiers in Psychology 2(181) (2011), doi:10.3389/fpsyg.2011.00181 (accessed May 7th 2013)
26. Loo, C.K., Liew, W.S., Sayeed, M.S.: Genetic Ensemble Biased ARTMAP Method of ECG-Based Emotion Classification. In: Watanabe, T., Watada, J., Takahashi, N., Howlett, R.J., Jain, L.C. (eds.) Intelligent Interactive Multimedia: Systems & Services. SIST, vol. 14, pp. 299–306. Springer, Heidelberg (2012)
27. Loo, C.K., Cheong, S.F., Seldon, M.A., Mand, A.A., Muthu, K.S., Liew, W.S., Lim, E.: Genetic-Optimized Classifier Ensemble for Cortisol Salivary Measurement Mapping to Electrocardiogram Features for Stress Evaluation. In: Anthony, P., Ishizuka, M., Lukose, D. (eds.) PRICAI 2012. LNCS, vol. 7458, pp. 274–284. Springer, Heidelberg (2012)