Appl. Math. Inf. Sci. **11**, No. 6, 1725-1730 (2017)

Applied Mathematics & Information Sciences
An International Journal

http://dx.doi.org/10.18576/amis/110619

Function Approximation with Deep Neural Network for Image Classification in Fuzzy Domain

Sasi Gopalan[1,*], *Linu Pinto*[2] , *Sheela. C.*[3] *and Arun Kumar M. N.* [3]

[1] Cochin University of Science and Technology, Kochi, Kerala, India
[2] Department of Mathematics, Cochin University of Science and Technology, Kochi, Kerala, India
[3] Federal Institute of Science and Technology, Kochi, Kerala, India

Received: 9 Jan. 2017, Revised: 10 Oct. 2017, Accepted: 14 Oct. 2017
Published online: 1 Nov. 2017

Abstract: Image Classification and retrieval of image from a large database has a great relevance in the present Scenario. A lot of work for an efficient method of image retrieval from large database has been made in the recent surveys. Here we propose a mathematical model based on CBIR system that uses the deep neural architecture for classification where the inputs are fuzzy grassland image features. Grassland image features varies according to the varieties of grassland images available through satellite images and hence its classification is a complex process. This paper proposes a new method for classification in which the inputs to the Neural Network are fuzzified and transformed in such a way that it clusters around a pivot vector there by making the classification task less complicated. This classification procedure is established theoretically by developing a mathematical model based on Neural Network approximation with fuzzy inputs. This model brings a transformation from the input image feature space to the output approximation space through the composition of mapping between the hidden transformation spaces that helps to strengthen the function approximation to the desired output. The Graphical representation on Fig(i) throws an insight into the mathematical theory of a CBIR system which unifies the advantages of deep neural architecture and fuzzy approximators. The mathematical concepts such as open balls, metric, limits, continuity etc are incorporated to establish the necessary and sufficient condition in the fuzzy based neural system for better and clear image retrieval.

Keywords: Image Features, Fuzzy inputs, Neural Network, Approximation, Transformation, image classification

1 Introduction

Image classification is a complex process and there are many factors affecting its process. A suitable classification design is essential to improve the classification accuracy in image processing. This work proposes a new mathematical model on fuzzy based Neural Network system to ensure the approximation of classified images. Classification can be done in the feature space by defining a distance measure. Neural network acts as the best classifier and fuzzy mathematics works as the best approximators so that a combined neural fuzzy approximation has been adopted in the proposed work.

For best convergence in the Neuro Fuzzy model, the mathematical properties such as metric space, approximation space, convergence, continuity and open balls are discussed. The relation between these properties with the deep architecture of neural network is discussed in this paper and is mentioned in **Table1** and **Fig** 1.

A Neural Network consists of an input layer, one or more hidden layers and an output layer. Each layer is made up of units. The input to the layers correspond to the attributes measured for each training tuple. The weights in NN act as the activation function to transform the inputs to next hidden layer. The weights are modified while training the NN to get the desired output. Output of NN is evaluated with the help of modified activation functions and the process is repeated again and again until NN output coincide with the desired output. Many strategies have been recently developed to train neural network that effectively reduces the computation time and leads to better result[3,5,6,7,8]. But the most approximate convergence cannot be guaranteed in any of the strategies. Success of classification depends mainly on the presentation of data to the NN architecture and hence

* Corresponding author e-mail: sgcusat@gmail.com

better feature representation implies less error prone results. In this paper a new method is suggested to represent the input data. According to [9], there is a lack of deterministic relationship between input and the solution. Thus the proposed new representation of input data provides a way to establish a deterministic relationship by introducing a measurable space to strengthen the functional relation inside the model. One of the main attractions of this work is the development of such relation by introducing the target function defined in Section 4.1. Layer by layer unsupervised training procedure has been done in Deep Belief Networks by [10, 11]. Thus a layer wise mathematical transformation has been done in this mathematical model to strengthen the functional representation between the measurable and approximation spaces.

The deep architecture of non linear operations of NN helps to approximate any non linear mappings. In [1], it is shown that the capacity of DNN to approximate any measurable function with any degree of accuracy. According to [4] it is clear that there is chances of error in the back propagation technique. This is mainly due to the overfitting nature of the network and the network may converge to local optima instead of global optimum.

Kurt Hornik showed in 1991 that it is not a specific choice of the activation function, but neural network architecture itself that gives it the potential of being universal approximators [2]. Hence studying the architecture of Deep Neural network by analyzing the function approximation mathematically, throws light into the approximation process carried out inside neural networks. This could further lead to the enhancement of the mathematical database of data mining through neural networks providing a way for future research on developing new strategies that could guarantee the most appropriate convergence of NN approximation on classification process.

This paper is organized as follows: Section 1 gives an introduction, Section 2 explains the basic needs for the development of the mathematical model. In Section 3, Deep Neural Network is introduced for the optimization purpose. In Section 4 convergence criteria for classification is represented. In Section 5, a distance measure is introduced to establish the criteria for convergence. In Section 6 the necessary and sufficient condition for convergence is established with the concept of neighborhood, limit, continuity and open balls in the input feature space. Section 7 deals with conclusion and directions to future applications.

2 Basic needs of the Model

Defintion 2.1. Let X_G be the collection of $'n'$ dimensional vectors with each component as a real number belonging to $[0,1]$, then define the collection of images to be classified as :

$$X_G = \{v_{c_i} \in K^n / K = [0,1]\}$$

where v_{c_i} symbolizes the image to be classified into grassland category.

Defintion 2.2. (Fuzzy Model of input features): Each component of v_{c_i} represents a function transformation, μ_{ij} from a set X to $[0,1]$ for the j^{th} component of the i^{th} image in X_G . Here X denotes the set of linguistic variables that represent the image features. Thus the image features represented by the linguistic variables are fuzzified by the transformation μ_{ij} for each component of i^{th} image. This transformation is well defined as for an image degree of uncertainty on a component is uniquely defined. The highest value in the range i.e; '1' represents the nonexistence of indefiniteness and as the value decreases the degree of indefiniteness increases.

The relevance of each component in the classification procedure is determined by the value of μ_{ij} in the neighborhood of 1. This evokes an importance to define a fuzzy measure on distance in neighborhood definition.

Defintion 2.3. (Fuzzy distance Measure in neighborhood definition): The fuzzy distance measure on $[0,1]$ is defined by (X, d_ϵ) where X is the set of linguistic variables $x_j^{'s}$ fuzzified by the transformation μ_{ij} for the i^{th} image and $d_\epsilon\left(x_j, 1\right) = \begin{cases} 0 & \text{if } \mu_i\left(x_j\right) = B_{\epsilon^-}(1) \\ 1 & \text{if } \mu_i\left(x_j\right) = B^c_{\epsilon^-}(1) \end{cases}$ where $B_{\epsilon^-}(1) = \{t \in [0,1] / 1-\epsilon < t \leq 1\}$ is the ϵ - neighborhood of 1 in the space $K = [0,1]$ and as the ϵ value get smaller and smaller the fuzziness decreases for each fuzzy image feature.

Also $B^c_{\epsilon^-}(1) = \{t \in [0,1] / t \notin B_{\epsilon^-}(1)\}$. Hence $B^c_{\epsilon^-}(1)$ characterizes those image features that are unclear or it cannot be fuzzified due to the presence of higher uncertainty because of noises. Most Grassland images are that kind images since they are captured by satellites and contains much noise.

By the definition of metric, (X, d_ϵ) is clearly a metric space.

2.1 Selection of Fuzzy image features

Let $B_{\epsilon^-}(1)$ be a neighborhood consisting of all image features very close to 1 and n_0 represents the number of such feature components in the i^{th} image. For better classification, the dimension of the vector n is selected in such a way that n_0 is sufficiently very large compared to $n - n_0$. Hence the selection of fuzzy variables play a great role in the classification procedure. Most grassland images are captured by satellite images and contain much noise and hence those fuzzy image features should be considered that are less affected by noises and the fuzzification value falls in the neighborhood $B_{\epsilon^-}(1)$. Grassland image features varies according to the varieties of grassland images available through satellite images and hence its classification is a complex process. To bring uniformity in the image features they are fuzzified in the input space and a one to one onto map between the components of the fuzzified image features is done to cluster around the pivot vector. This kind of feature organization is more suitable for the grassland classification and is a better representation of grassland images as it allows to cluster the vectors that represents grassland images around the pivot vector $(1, 1, 1, 1, 1.....0, 0)$, where 1 occurs in the first n_0 components and zeros in the next $n - n_0$ components. This can improve the entire model accuracy even for the untrained data. So good feature engineering enhances the classification procedure.

3 Classification and Optimization by Neural Network

Artificial Neural Network works as a decision making tool in a way just like human brain. Information processing inside neural network directly depends on the architecture of the Network and the way in which the inputs are presented to the network. Inputs presented could be linear or nonlinear depending on the problem and neural network perform a better classification result in many such problems. Since Image features modeled as in 2.1 follows a specific pattern, the classification task for Neural Network becomes less complicated. While Classification is being carried out, optimization is a challenging factor and main aim is to reduce the error in the classification result. For this the entire NN approximation process should be mathematically analyzed and results should be developed that could reduce the errors. In this paper the entire approximation process is mathematically formulated with Feed forward Neural Network with multiple hidden layers that supports transformations of input data to a space where classification could be done with minimal errors. In image feature space classification

Table 1 NN Hidden Layer Approximation Spaces

Hidden Layers (1)	No of Nodes in each layer (2)	Hidden Approximation space(1) and (2)	Transformations
1	l_1	N_{l_1}	Φ_1
2	l_2	N_{l_2}	Φ_2
.	.	.	.
	.	.	.
$m+1$	l_{m+1}	$N_{l_{m+1}}$	Φ_{m+1}

can be mathematically represented by considering the feed forward Neural Network with multiple hidden layers as shown in fig (i). The inner product operation in the multilayered feed forward NN was performed to define the NN classification as a function Φ_{m+1}. For that purpose the following are considered. Let there be $'n'$ number of inputs, $'m'$ number of hidden layers and $'f'$ be the activation function. Then the approximation space of the proposed model is shown in Table 1.

By using this classification following model can be developed in an image feature space.

3.1 Mathematical Representation

Consider the composite mapping for classification as $\Phi_{m+1} = \Phi_m \circ \Phi_{m-1} \circ ...... \circ \Phi_2 \circ \Phi_1$ with representations

$$\Phi_k : K^{l_{k-1}} \to K^{l_k}$$

defined as $\Phi_k = (\Phi_{k,1}, \Phi_{k,2}, ...\Phi_{k,l_k})$ and $\Phi_{k,p}(v_{c_i}) = f\left(\left\langle \Phi_{k-1} \circ ......\Phi_2 \circ \Phi_1(v_{c_i}), \left[{}^{k-1}W_{pj}\right]^T \right\rangle\right)$,
$\forall j = 1, 2, 3, ...l_{k-1}, p = 1, 2....l_k$ for $k = 1, 2, ...m+1$ where $l_0 = n, \Phi_0(v_{c_i}) = v_{c_i}$. Then there exist a function

$$\Phi_{m+1}(v_{c_i}) = f\left(\left\langle \Phi_m \circ ......\Phi_2 \circ \Phi_1(v_{c_i}), \left[{}^{m}W_{1j}\right]^T \right\rangle\right)$$

$\forall j = 1, 2, 3, ...l_m$.

*Remark.*The assigned vectors corresponding to the connections from i^{th} layer to the $(i+1)^{th}$ layer are represented by the column matrix $[{}^iW_{pj}]^T$. Here considered 'n' inputs to NN as fuzzy components of vector v_{c_i} and the activation function squashed the outputs nearer to 0 or 1. For example in the case of grassland images NN outputs are squashed nearer to 1 for grassland images and nearer to 0 for non grassland images

4 Convergence in image classification

It is essential to show the existence of image classification because of the success of NN approximation for the entire development of mathematical model.Let g be the function represents the inner product operation performed in ANN and the approximation is targeted on $g \subseteq \Phi_{m+1}$

4.1 Convergent Criteria-Functional representation

Let $g : K^n \to K$ be a mapping defined as $g(v_{c_i}) = \begin{cases} \alpha & \text{if } \mu_{ij} \in B_{\epsilon^-}(1), \forall j = 1, 2, ...n \\ \beta & \text{if } otherwise \end{cases}$,
$\alpha \in B_{\epsilon^-}(1)$ and $\beta \in B^c_{\epsilon^-}(1)$ for a fixed ϵ (negligibly small). In Particular $g(v^*) = 1 \forall \mu_{ij} = 1, j = 1, 2, 3....n$.
Result Let v_{c_i} and v_{c_k} corresponding to i^{th} and k^{th} image. Then v_{c_i} and v_{c_k} are classified into same class iff $g(v_{c_i}) = g(v_{c_k})$.

For the Neural Network classification to give better approximations, the mapping Φ_{m+1} should converge to the function g as defined in 4.1 and leads to following representation of classes in X_G.

4.2 Neighborhood points in classification

Let X_{g_0} and X_{g_1} be two classes of X_G such that

$$X_{g_0} = \left\{ v_{c_i} / v_{c_i} \in X_G, \Phi_{m+1}(v_{c_i}) \in B^c_{\epsilon^-}(1) \right\}$$

$$X_{g_1} = \{ v_{c_i} / v_{c_i} \in X_G, \Phi_{m+1}(v_{c_i}) \in B_{\epsilon^-}(1) \}$$

Then the vectors in the class X_{g_0} are mapped to the points x in the compliment of ϵ- neighborhood of 1 and the vectors in the class X_{g_1} are mapped to the points x in the ϵ - neighborhood of 1.

5 Classification in metric Space

Now inspired from the definition of metric space a notion of distance has been implemented for the classification of elements in X_G . The classification could be modeled by defining a metric in the input space and the output space as follows:
Let $d_g : X_G \times X_G \to [0, 1]$ be a function defined as,

$$d_g(v_{c_i}, v_{c_k}) = \frac{1}{n} \sum_{j=1}^{n} \left| \mu_{ij} - \mu_{ik} \right|$$

.
Let $d_\epsilon : K \times K \to K$ be a function defined as,

$$d_\epsilon(\Phi_{m+1}(v_{c_i}), \Phi_{m+1}(v_{c_k})) = |\Phi_{m+1}(v_{c_i}) - \Phi_{m+1}(v_{c_k})|$$

Then $(d_g, X_G)\&(d_\epsilon, K)$ satisfies all the axioms of a metric space.
Result: Let X_G be the collection of n-dimensional vectors representing images. Let X_{g_0} and X_{g_1} be two classes of X_G as defined in 4.2. Then $X_{g_0} \cap X_{g_1} = \phi$.
Let $v_{c_m} \in X_{g_0} \cap X_{g_1}$
$\Rightarrow \Phi_{m+1}(v_{c_m}) \in B^c_{\epsilon^-}(1) \& \Phi_{m+1}(v_{c_m}) \in B_{\epsilon^-}(1)$
$\Rightarrow \Phi_{m+1}(v_{c_m}) \in (\epsilon, 1]^c \cap (\epsilon, 1]$ which is a contradiction from 3.1,Φ_{m+1} is well defined.
Hence $X_{g_0} \cap X_{g_1} = \phi$

6 Convergence of Neural Network classification

For the convergence of Neural Network classification to a better approximation, the concept of continuity between the metric spaces is considered.

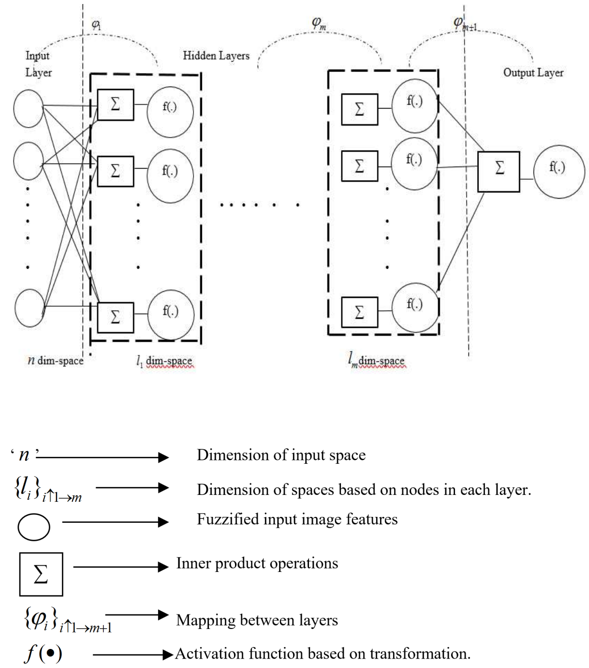

Figure 1 Transformations and Approximations (Geometrical Representation of Input-Output Relationships)

6.1 Necessary Condition

6.1.1 Existence of Limit:

Let $g : K^n \to K$ be a mapping as defined in 4.1 and

$$v^* = (1, 1, 1, 1...., 1, 0, ..0) \in K^n$$

Then $\lim_{v \to v^*} g(v) = 1 \Leftrightarrow \mu_{ij} \in B_{\epsilon^-}(1)$ and

$$\mu_{ij} \to 1, \forall j = 1, 2, ...n_0$$

where $v^* = (\mu_{i1}, \mu_{i2}, \mu_{i3}, ...\mu_{in_0}, 0, 0..0) \in K^n$
With the concept of limit, continuity can be defined on function, mapping the metric space (d_g, X_G) to the metric space (d_ϵ, K)

6.1.2 Existence of Continuity

Let X_G be the collection of n-dimensional vectors representing images and d_g is a metric on X_G, d_ϵ is a metric on $K = [0, 1]$and g be the function as defined in 4.1.
Let mapping $\psi : (d_g, X_G) \to (d_\epsilon, K)$ is continuous at v^* if for a given $\epsilon > 0, \exists \delta > 0$, depending on ϵ and v^* such that

$$d_g(v_{c_i}, v^*) < \delta \Rightarrow d_\epsilon(\Phi_{m+1}(v_{c_i}), g(v^*)) < \epsilon$$

Result: By using 6.1there exist a collection of open balls

$B_{g_m}(\nu^*) \subseteq X_G$ such that

$B_{g_m}(\nu^*) = \{\nu_{c_i} \in X_G/\mu_{ij} \in B_{\delta_m^-}(1)\forall j = 1, 2, ...n_0$ and $\delta_1 > \delta_2 > ...... > \delta_m > ... \to 0\}$
Similarly open balls in K such that:
$B_{\epsilon_m}(g(\nu^*)) = \{\Phi_{m+1}(\nu_{c_i}) \in K/\Phi_{m+1}(\nu_{c_i}) = \alpha \in B_{\delta_m^-}(1)$ and $\epsilon_1 > \epsilon_2 > ...... > \epsilon_m > ... \to 0\}$

6.2 Sufficient Condition

If $\psi : (d_g, X_G) \to (d_\epsilon, K)$ is continuous at ν^* then ψ maps open balls centered at ν^* in X_G to the corresponding open balls centered at $g(\nu^*)$ in K such that $|\psi(\nu^*) - g(\nu^*)| < \epsilon$.

7 Conclusion and Future Directions

In this work an approximation space is developed with deep neural networks and the input fuzzy image features in such a way that it falls in the neighborhood of pivot vector $\nu^* = (1, 1, 1, ...1, 0, ..0) \in K^n$ .The features around the pivot vector strengthens the approximation process of the neural network. A mathematical framework corresponding to the neighborhood helps to collect maximum targeted data points in the image retrieval process. Then the transformation of these data points gives convergence criteria of the model. This shows the existence of the mathematical database in the fuzzy based NN approximations. As a future work pattern recognition can be made by defining a one to one onto map between the image features using the theory of permutation group from algebra so that suitable data points can be identified by Neural Network.

References

[1] Balazs Csanad Csaji, Approximation with Artificial Neural Networks, Faculty of Sciences, Eotvos, Lorand University, Hungary (2001).

[2] Kurt Hornik (1991) "Approximation Capabilities of Multilayer Feedforward Networks", Neural Networks, 4(2), 251-257.

[3] Deng, L.Yu, D. (2014). "Deep Learning: Methods and Applications"(PDF). Foundations and Trends in Signal Processing 7: 3-4.

[4] Bengio, Yoshua (2009). "Learning Deep Architectures for AI" (PDF).Foundations and Trends in Machine Learning 2 (1): 1-127

[5] Schmidhuber, J. (2015). "Deep Learning in Neural Networks: An Overview".Neural Networks 61: 85-117

[6] Y. Bengio et al.., "Advances in optimizing recurrent networks," ICASSP, 2013.

[7] G. Dahl et al.., "Improving DNNs for LVCSR using rectified linear units and dropout,"ICASSP, 2013.

[8] G. E. Hinton., "A Practical Guide to Training Restricted Boltzmann Machines," Tech. Rep. UTML TR 2010-003, Dept. CS., Univ. of Toronto, 2010.

[9] Sivanandam , Paulraj, Manu Sivanandam, Introduction to Artificial Neural Networks,Vikas Publishing House Pvt Ltd, 01 Nov 2009, Neural networks (Computer science),page 143.

[10] Hinton, G. (2009). "Deep belief networks". Scholarpedia 4 (5): 5947.

[11] Hinton, G. E., Osindero, S., Teh, Y. W. (2006). "A Fast Learning Algorithm for Deep Belief Nets" . Neural Computation 18 (7): 1527-1554.

Sasi Gopalan is Associate Professor of Mathematics at Division of Applied science and Humanities, School of Engineering, Cochin University of Science and Technology, Kochi, Kerala, India. He received the PhD degree in Mathematics from Mahatma Gandhi University, Kottayam, Kerala, India. His research interests are in the areas of Applied Mathematics, Fuzzy Mathematics, Approximation Theory, Pattern recognition with Intelligent systems, Mathematical Models in Image Processing.

Linu Pinto is Research Scholar at the Department of Mathematics, Cochin University of Science and Technology, Kochi, Kerala, India. Her main area of interests are Pure and Applied Mathematics, Mathematical modeling in Data mining, Optimization theory and Function approximations with Intelligent Systems.

Scheela C. is currently working as a Professor and Vice Principal in Federal Institute of Science and Technology. She received the Ph.D degree in Fuzzy Topology from Cochin University of Science and Technology. She has received CSIR-SRF Award. She has presented papers in many national/international journals/conferences.She is an active member of Kerala Mathematical Association and is a life member of Indian Society for Technical Education (ISTE).

Arun Kumar M. N. is currently working as an Associate Professor in the Department of Computer Science and Engineering, Federal Institute of Science and Technology. He received the Ph.D degree from University of Mysore in 2013. His research work includes image processing, datamining and machine learning.