On Wavelet based Modeling of Neural Networks using Graph-theoretic Approach

B. Bhosale

S. H. Kelkar College (Affiliated to University of Mumbai), Devgad, Sindhudurg (M.S.), India E-mail: bn.bhosale@rediffmail.com

Abstract: A graph is an abstract representation of complex networks. Many practical problems can be represented by graphs. With graphs, it is possible to model many types of relations and process dynamics in physical, biological, social and information systems. More specifically, the relationships among data in several areas of science and engineering, e.g. computer vision, molecular chemistry, molecular biology, pattern recognition, and data mining, can be represented in terms of graphs. For example, graph analysis has been used in the study of models of neural networks, anatomical connectivity, and functional connectivity based upon functional magnetic resonance imaging (fMRI), electroencephalography (EEG) and magnetoencephalography (MEG).

Of late, many important properties of complex networks have been delineated and a significant progress has been made in understanding the relationship between the structural properties of networks and the nature of dynamics taking place on these networks with the help of graph theoretical approach. These developments in the theory of complex networks have inspired new applications in the upcoming field of neuroscience.

In this work, a novel wavelet based neural network stochastic model that extends existing neural network methods for processing the data represented in a graph domain is proposed. Here, the approach is based on defining random walks on arbitrary infinite graphs representing neural networks. The random walk itself is a stochastic process characterized by some probability distribution. More so, random walks exhibit fractal-like patterns that, in turn, legitimately allow the use of wavelet methods for visualizing and processing. The wavelet transform is constructed for the random walk represented by Gaussian function with the vanishing momenta engendered by Gaussian function as an analyzing function. The wavelet transform so devised is well defined since both the testing function and analyzing function employed in the construction belonged to a class of Gaussian functions. Theory is further extended to discrete format for numerical implementation of the transform. The robustness of the proposed model as against the existing ones has been justified by highlighting the potential applications of neuroscience.

In this work, an attempt is made to combine wavelet transform and graph theory. The work is not example specific but it provides a theoretical framework for analyzing the complex structures of neural networks representing various neuroscientific phenomena.

Keywords: wavelet transform, graph neural networks, random process, neuroscience

B. Bhosale, On Wavelet based modeling of Neural networks using Graph theoretic approach

1. INTRODUCTION

Graph theory pertains to the study of *graphs* which are mathematical structures used to model pairwise relations between objects. Many relationships among data in several areas of science and engineering, e.g., computer vision, molecular chemistry, molecular biology, pattern recognition, data mining etc, can be represented in terms of graphs and the graph theoretic approach can be employed to analyse such complex data.

A network is a directed graph or a digraph with weighted edges. In recent years many important properties of complex networks have been delineated and the relationships between the structural properties, nature of dynamics taking place on these networks have been studied. For instance, the 'synchronizability' of complex networks of coupled oscillators can be determined by graph spectral analysis. These developments in the theory of complex networks have inspired new applications in the field of neuroscience such as study of models of neural networks, anatomical connectivity, and functional connectivity based upon fMRI, EEG and MEG (Stam and Reijneveld, 2007). The survey of the recent literature reports several models used for analyzing complex networks. Mashaghi et al. (2004) studied the structure of the yeast protein complex network in which weighted edges between complexes represent the number of shared proteins.

Recently various approaches have been unified in neural network models called graph neural networks (GNN) (Scarselli et al., 2009), which are used for processing the data represented in graph domains. GNN can be used to process structured data inputs, e.g. acyclic graph, cyclic graph, directed or un-directed graphs. The GNNs are of two kinds viz. Biological neural networks (BNN) and Artificial neural networks (ANN) (Podolak, 1998). In BNN, the neurons are linked as a network in a certain order, e.g. human neural network, whereas, the ANNs are aimed at modeling the organization principles of central neural system, with the hope that the biologically inspired computing capabilities of ANN will allow the cognitive and sensory tasks to be performed more easily and satisfactorily. The ANN models incorporate some structures, characters and functions of BNN by employing electronical and optical technologies (Jim and Zheng, 2002).

Wavelets have been used in analyzing graphs and the complex networks represented by graphs. Hammond et al. (2011) proposed a method for constructing wavelet transforms of functions defined on the vertices of an arbitrary finite weighted graph. Also, Szu et al.(1993) employed wavelet transform to analyze neural networks. Useful wavelet applications in turbulence onset, spectrum analyses, fractal aggregates, and bubble-chamber particle-track pattern-recognition problems are indicated. A comprehensive review of the potential use of has been given by Jasen et al. (2009), Mallat (1998), Smalter et al. (2009); and in particular wavelet statistical methodology, in different areas of molecular biology such as genome sequence, protein structure and microarray data analysis (Lio, 2003). Moreover, the wavelet based analysis of genomic sequences is carried out by defining random process on these sequences (Bhosale et al., 2012).

In the present work, the discussion is confined to neural networks, its representation in terms of graphs, and its mathematical modeling. Methodology includes defining random walks on the digraphs representing neural networks, which themselves are the stochastic processes, and employing wavelet techniques. This approach has not seen exploited in the reported investigations on the various studies on graph neural networks.

2. NOTATIONS AND TERMINOLOGIES

Graphs: In the most common sense of the term, a graph is an ordered pair G = (E, V) comprising a set V of vertices or nodes together with a set E of edges or lines, which are 2-element subsets of V. As an extension, of the simple graph is a weighted graph G = (E, V, w) that consists of a set of vertices, a set of edges E, and a weight function $w : E \rightarrow R$ + which assigns a positive weight to each edge.

Random Walk: A random walk is a mathematical formalization of a path that consists of a succession of random steps. Often, random walks are assumed to be Markov chains or Markov processes. A random walk of length k on a possibly infinite graph G with a root θ is a stochastic process with random variables $X_1, X_2, X_3, ..., X_k$ such that $X_1 = 0$ and $X_{i=1}$ is a vertex chosen uniformly at random from the neighbors of X_i .

Wavelet Transform: The wavelet transform (WT) is a decomposition of a function, f(x), with respect to a basic wavelet $\psi(x)$, given by the convolution of a function with a scaled and translated version of $\psi(x)$

$$W_{\psi}[f(x)](a,b) = |a|^{-1/2} \int f(x)\psi^*\left(\frac{x-b}{a}\right) dx = \langle f, \frac{1}{\sqrt{|a|}}\psi\left(\frac{x-b}{a}\right) \rangle \tag{1}$$

where, $f, \psi \in L^2$, the square integrable functions. and ψ , the basic wavelet or analyzing wavelet, satisfies the admissibility condition $C_{\psi} = \int_{-\infty}^{\infty} \frac{|\widehat{\psi}(\omega)|^2}{|\omega|} d\omega < \infty$. '*a*' is the scale parameter, a > 0, '*b*' is the translation parameter.

3. MODELING NEURAL NETWORKS

In graph neural networks (GNN), each node corresponds to one neuron and the arrows usually denote weighted sums of the values from other neurons. GNNs are used to process structured data inputs. This class of neural networks implements a function $\tau(G, n) \in \mathbb{R}^m$ that maps a graph *G* and one of its nodes *n* onto an *m*-dimensional Euclidean space (Scarselli et al., 2009).

A linear factor analysis model that represents the neural network is expressed in the form (Valpola, 2000),

$$x_{i}(t) = \sum_{j} A_{ij} s_{j}(t) + a_{i} + n_{i}(t)$$
(2)

where *i* indexes different components of the observation vector x_i , representing weighted sums of underlying latent variables, *j* indexes different factors and A_{ij} are the weightings of the factors *s*, also known as factor loadings. The factors *s* and noise *n* are assumed to have zero mean. In this model, the effect of the inaccuracies and other causes is summarised by Gaussian noise *n*. In anticipation of the dynamic model, the observations are indexed by *t* referring to time.

The probability density function for the observations $x_i(t)$ is a normal probability density function given by

$$P(x_{i}(t) \mid s(t), A, a_{i}, \sigma_{i}^{2}) = \frac{1}{\sqrt{2\pi\sigma_{i}^{2}}} exp\left(-\frac{\left[x_{i}(t) - \sum_{j} A_{ij} s_{j}(t) - a_{i}\right]^{2}}{2\sigma_{i}^{2}}\right)$$
(3)

The expression (3) can be simply written as $x(t) \sim N(As + a, \sigma^2)$, where the vector σ^2 contains the variances σ_i^2 , the variances of the Gaussian noise terms $n_i(t)$ as σ_i^2 . The normal probability distribution itself is a Gaussian.

With this ground work, we can define the random walk on the graph, G = (V;E;w) representing the neural network, where G = (V;E;w) is the connected weighted graph (or a tree) with *n* nodes and *m* edges; and $w = A_{ij}$, the weights showing a links between the nodes in the given neural network as follows. We can attach a *state* $x_n \in R$ to each node *n*, that is based on the information contained in the neighborhood of *n* as

$$x_n = f_w(l_n, l_{co}[n], x_{ne}[n], l_{ne}[n])$$
(4)

where l_n , $l_{co}[n]$, $x_{ne}[n]$, $l_{ne}[n]$ are respectively the label of *n*, the labels of its edges, the states and the labels of the nodes in the neighborhood of *n*.

The variable x_n contains a representation of the object denoted by n and can be used to produce an *output*, i.e. a decision about the object. Note that the GNN model captures also the random walks on graphs when choosing f_w as a linear function, where f_w is a parametric function, called *local transition function*, that expresses the dependence of a node n on its neighborhood (Scarselli et al., 2009). In random walks on graphs, the state x_n associated with a node is a real value and is described by

$$x_n = \sum_{u \in Pa[n]} a_{n,i} x_u \tag{5}$$

where Pa[n] is the set of parents of *n*, and $a_{n,i} \in R$, $a_{n,i} \ge 0$ holds for each *n*, *i*. The $a_{n,i}$ are normalized so that $\sum_{i \in Pa[n]} a_{i,n} = 1$.

Here, the Eq. (5) actually represents a random walker traveling on the graph. The values $a_{n,i}$ in Eq (5) represent the probability that the walker, when visiting node *n*, decides to go to node *i*. The state x_n stands for

the probability that the walker is on node *n* in the steady state. When all x_n are stacked into a vector *x*, Eq. (5) becomes x = Ax where $A = \{a_{n,i}\}$ and $a_{n,i}$ is defined as in (5) if $i \in Pa[n]$ and $a_{n,i} = 0$ otherwise. The corresponding random walk of length *k* can be defined appropriately as $S = \{s[j]; j = 1, 2, ..., k\}$, where for any position *l*, we have a cumulative sum of x_n for $1 \le j \le k$ described by $s[l] = \sum_{n=1}^{l} x_n$. This *random walk* of length *k* on graph *G* with a root 0 forms a stochastic process with random variables $X = \{x_n\}$. For a completely random sequence $\{x_n\}$, the path mapping gives a Brownian motion type signal (Bhosale et al., 2012), say, $S(t) = \sum_{n=1}^{t} x_n$

4. WAVELET BASED ANALYSIS

The random walk on graph neural network itself is a stochastic process that exhibits self-similarity pattern. Hence, wavelet decomposition, which is also based on self-similarity, turns out to be an effective tool to extract fractal information from the structure of graph neural network. For a random process, X = X(t, w) defined on the probability space (Ω, B, P) , for any function $\psi: R \to C$ satisfying the admissibility condition, the wavelet transform is given as

$$W_{\psi}[X(u,w)](a,t) = \int_{R} \psi_{a,t}^{*}(u) X(u,w) du, t \in R, a > 0, w \in \Omega.$$
(6)

where $\psi_{a,t}(u) = \frac{1}{\sqrt{a}}\psi\left(\frac{u-t}{a}\right)$, the scaled and translated version of analyzing wavelet ψ , subject to the stipulation

$$E \mid X(t) \mid^{2} = \int_{\Omega} \mid X(t,w) \mid^{2} dP(w) < \infty, \forall t \in R, w \in \Omega,$$

$$\tag{7}$$

where E denotes the mathematical expectation.

Applying this formulation to the complex valued random walk defined on GNN, $S(t) = \sum_{n=1}^{t} x_n$, where x_n is as given in Eq (5), we obtain the wavelet transform of S(t) with respect to the analyzing/generic wavelet ψ as

$$W_{\psi}[\mathcal{S}(t)](a,b) = \int_{R} \frac{1}{\sqrt{a}} \psi^*\left(\frac{t-b}{a}\right) \mathcal{S}(t) dt, b \in R, a > 0,$$
(8)

subject to the stipulation $E | \mathcal{S}(t) |^2 = \int_{\mathbb{R}} | \mathcal{S}(t) |^2 dP(x) < \infty, \forall x \in \mathbb{R},$ (9)

where *E* denotes the mathematical expectation. To have the WT well defined, the signal S(t) should belong to the $L^2(R)$, the class of square integrable functions. Indeed, S(t) belongs to the class of square integrable functions since x_n 's have Gaussian distribution and the Gaussian functions belong to this class of square integrable functions.

The wavelet transform (8), is a convolution of function $S(t) \in L^2(R)$ with certain locally supported function $\psi(t)$ shifted and dilated, called analyzing wavelet. It should be noted that the choice of the analyzing wavelet is of vital importance as it dictates the representation and properties of the wavelet transform. In particular, the choice depends on the factors like the kind of signal to be analyzed. Usually, the analyzing wavelet ψ is chosen from the family of wavelets based on the Gaussian functions, which can be successively differentiated to any arbitrary order n,

$$\varphi^n(t) = (-1)^{n+1} \frac{d^n}{ds^n} exp(-t^2/2)$$
 and whose first moments vanish, that is, $\int t^k \varphi(t) dt = 0$, for $0 \le k \le n$.

For appropriate choice of analyzing wavelet, we can make use of the fact that the random walk formed by the x_n given in Eq. (5); of which each x_i , s as the weighted sums of the latent variables given by Eq. (3) in the neural network having induced normal probability density function,

$$f(x_i;\mu_i,\sigma_i) = \frac{1}{\sqrt{2\pi\sigma_i^2}} e^{-\frac{(x_i-\mu_i)^2}{2\sigma_i^2}}$$
(10)

B. Bhosale, On Wavelet based modeling of Neural networks using Graph theoretic approach

with $\mu_i = \sum_i A_{ij} s_i(t), -\infty < x < \infty, -\infty \le \mu_i \le \infty$ and $\sigma_i > 0$.

Thus, the random walk S(t) defined on GNN above is said to have normal probability distribution induced by normal probability density function represented at Eq (10); which itself is a Gaussian function.

Moreover, by the transformation, $X = \frac{x-\mu}{\sigma}$, we can have the standard normal probability density function, $\psi(X) = \frac{1}{\sqrt{2\pi}}e^{\frac{-x^2}{2}}$, which is again a Gaussian function that can be differentiated successively to any arbitrary order, say, *n*, as $\psi^{(n)}(t) = (-1)^{n+1}\frac{d^n}{dx^n}exp(-x^2/2)$, and whose first momenta vanish. In particular, we can obtain second order moment,

$$\psi_2(X) = \psi''(X) = \frac{(x^2 - 1)}{\sqrt{2\pi}} e^{-\frac{x^2}{2}},\tag{11}$$

which is a "Mexican hat wavelet". The Mexican hat wavelet has been analytically proved wavelet that satisfies the admissibility condition, $C_{\psi_2} = 2 \int_0^\infty \frac{|\widehat{\psi}_2(\omega)|^2}{|\omega|} d\omega = 2 \pi \Gamma(n) < \infty$. Hence, we can safely employ $\psi_2(X)$, as the analyzing wavelet in the transform treatment.

Thus, the wavelet transform of the random walk S(t), (which is rather a Gaussian function) with an analyzing wavelet, ψ_2 will be

$$W_{\psi_2}[\mathcal{S}(t)](a,b) = \int_R \frac{1}{\sqrt{a}} \psi_2\left(\frac{t-b}{a}\right) \mathcal{S}(t) dt$$
(12)

It should be therefore noted that the wavelet transform given by Eq (12) is indeed well defined.

For numerical implementation, to compute the wavelet transform, the Eq (12) needs to be discretized as

$$w_{j,k} = W_{\psi_2} \mathcal{S}(j,k) = \int_R \mathcal{S}(t) \psi_{2j,k}(t) dt$$
(13)

In this discretized form, the signal S(t) is represented as a linear decomposition

$$\mathcal{S}(t) = \sum_{j} \sum_{k} w_{j,k} \psi_{2j,k}(t), j, k \in \mathbb{Z},$$
(14)

where $w_{j,k}$ are the wavelet coefficients of the expansion that constitute the discrete set, and $\psi_{2j,k}$ is a set of wavelet functions of t given by $\psi_{2j,k}(t) = 2^{j/2}\psi_2(2^jt - k)$: $j,k \in \mathbb{Z}$.

Note that the scaling and translations are discrete, and the indexes *j* and *k* are respectively related to these processes. The basis functions $\psi_{2_{j,k}}$ are dilated in a dydatic form (in powers of two), when varying the value of index *j*, and in analogous way translated when varying the index *k*. In this process, translation is associated with time resolution, and dilation provides scaling.

Now, let us turn to inverting the transform so as to reconstruct the signal corresponding to the given set of transform coefficients so obtained. Intuitively, the wavelet coefficient $w_{j,k} = W_{\psi_2} S(j,k)$ provides a measure of "how much of" the wavelet $\psi_{2j,k}(t)$ is present in the signal S. This suggests that the original signal may be recovered by summing the wavelets $\psi_{2j,k}(t)$ multiplied by each wavelet coefficient $w_{j,k}$.

That is,
$$\frac{1}{C_{\psi_2}} \sum_{n=1}^N \int_R w_{j,k} \psi_{2j,k}(t) \frac{dt}{t} = S^{\#}.$$
 (15)

Finally, to calculate correlations between wavelet coefficients $W_{\psi_2}[S(t)](a, b)$ at different scales, define the wavelet covariance of the covariance function for all t,

 $R_{\mathcal{S}}(u,v) = E\psi_2^*(u)\varphi(v), u, v \in \mathbb{R}$, as

$$R_W(a,b,c,d) = EW_a(b)W_b(d) = \int_R \frac{1}{\sqrt{ac}} \psi_2\left(\frac{u-b}{a}\right) \psi_2^*\left(\frac{v-d}{c}\right) R_{\mathcal{S}}(u,v) du dv,$$
(16)

provided that the condition, $E\left\{\int_{R}\left|\frac{1}{\sqrt{a}}\psi_{2}^{*}\left(\frac{t-b}{a}\right)\mathcal{S}(t)dt\right|\right\}^{2} < \infty$, holds good.

That is, $R_W(a, c, b - d) = \langle W_{\psi_2}(a, b), W_{\psi_2}(c, d) \rangle$, where the curly brackets $\langle ., . \rangle$. mean the covariance $cov(W_1, W_2) = E \frac{(W_1 - E(W_1)(W_2 - E(W_2))}{\sqrt{DW_1 \cdot DW_2}}$ where *D* is the dispersion and *E* is the mathematical expectation. This leads to infer that for a random walk, wavelet coefficients correlation function will coincide with that of random signal; if no, the structure of wavelet coefficients correlation function will be different. The foregone theoretical design provides generalized theoretical set-up that can be feasibly applied for analyzing complex neural networks.

5. DISCUSSION

Wavelet analysis is well suited for scale-invariant systems, that is, the systems that exhibit self similar properties at different scales. The Brownian motion is a classical example of scale-invariant physical systems. Random walks are the random processes or the stochastic processes, which in turn, constitute a Brownian motion signal. In our work, we adopted a novel approach to model the neural network in graph theoretic framework, defined a random walk on the graph and suitably employed wavelet transform to analyse the complex network in this representation, characterized by Brownian motion signal. One of the goals of random walk is to determine long-range power law correlation based on measurements of fluctuations. The benefit of using this random walk is that the sequence always evolves along the positive direction on z-axis as a function of sequence index and this facilitates transform analysis. The wavelet transform technique locates the periodicities in the random sequence defined on GNN and also gives fine time index resolution so that these patterns can be located accurately in the walk representing the sequence. Transforming the signal actually yields the discrete set of wavelet coefficients, at a particular scale and translation which tells us how well the signal and the scaled and translated analyzing wavelet match. If the signal is similar to the scaled and translated analyzing wavelet, then the wavelet coefficient will have big magnitude. The wavelet coefficient also represents the degree of correlation between two functions at a particular scale and translation. The prime objective of wavelet analysis is to extract structural information from the complex network structure in the transform domain. In mathematical analysis, it is to obtain the wavelet transform of the random walk; which in turn can be plotted in space-time plane. This will demonstrate the structural pattern visually and help locating periodicities of these patterns in the sequence that represents the network. The actual values can then obtained by plotting $|W_{\psi_2}[S(t)](a,b)|$ over the space-time plane. The peculiar feature of this wavelet model is the choice of Gaussian momenta obtained from the normal probability density function of latent variables of the network as analyzing wavelet. The choice is justified owing to the fact that the random walk is formed by the sequence comprised of weighted sums of the latent variables in the neural network that have probabilities distributed normally; which in turn is a Gaussian function.

6. CONCLUSIONS

Complex neural networks arising in the small world of neuroscinece can be modeled with graphs using analogy. The random process can then be defined on graphs that represent complex neural networks; which, in turn, makes it possible to apply wavelet techniques. The present work robustly models the complex neural network in graph theoretic settings for which the wavelet analysis is employed in randomized/stochastic framework. The technique is novel in itself and relatively easy to implement. This technique will provide a new insight in handling the applications of network theory to neuroscience such as modelling of neural dynamics on complex networks, graph theoretical analysis of neuroanatomical networks, and the studies of functional connectivity with fMRI, EEG and MEG and others from the small world of neuroscience.

B. Bhosale, On Wavelet based modeling of Neural networks using Graph theoretic approach

REFERENCES

- Bhosale, B., Ahmed, B. S., and Biswas, A. (2013), Wavelet Based Analysis in Bio-informatics, Life Sci., 10(2), 853-859.
- Hammond, D. K., Vandergheynstb, P., and Gribonval, R. (2011), Wavelets on graphs via spectral graph theory, Applied and Computational Harmonic Analysis, 30(2), 129-150.
- Jansen, M., Nason, G. P., and Silverman, B. W. (2009), Multiscale methods for data on graphs and irregular multidimensional situations, Journal of the Royal Statistical Society: Series B (Statistical Methodology), 71(1), 97–125.
- Jim, XU, and Zheng, BAO. (2002), Neural networks and Graph theory, Science in China (Series F), 45(1), 1-24.
- Lio, P. (2003), Histidine biosynthetic pathway and genes, Bioinformatics, 19(1), 2-9.
- Lovasz, L. (1993), Random Walks on Graphs: A Survey, Bolyai society, Mathematical studies, Keszthely (Hungary), l(2), 1-46.
- S. Mallat. (1998), A Wavelet Tour of Signal Processing, Academic Press.
- Mashaghi, A. R., Ramezanpour, A., and Karimipour, V. (2004), Investigation of a protein complex network, European Physical Journal B Condensed Matter and Complex Systems, 41(1), 113-121.
- Peng, C., Byldyrev, S., and Stanley, H. (1992), Long-range correlations in nucleotide sequences, Nature, 356, 168-171.
- Podolak IT. (1998), Functional graph model of a neural network, IEEE Trans Syst Man Cybern, 28(6),876-81.
- Scarselli, F., Gori, M., Ah Chung, T., Hagenbuchner, M., and Monfardini, G. (2009), The Graph Neural Network Model, Neural Networks, IEEE Transactions, 20(1), 61-80.
- Smalter, A., Huan, J., and Lushington, G. (2009), Graph wavelet alignment kernels for drug virtual screening, Journal of Bioinformatics and Computational Biology, 7, 473–497.
- Stam, C. J., and Reijneveld, J. C. (2007), Graph theoretical analysis of complex networks in the brain, *Nonlinear Biomedical Physics*, 1:3 doi:10.1186/1753-4631-1-3.
- Szu, H. H., Yang, X. Y., Telfer, B. A., and Sheng, Y. (1993), Neural network and wavelet transform for scale-invariant data classification, Phys. Rev., 48(E), 1497–1501.
- Valpola, H. (2000), Bayesian Ensembel learning for Non-linear factor analysis, Acta Ploytechnica, Scandinavica, Mathematics and Computing Series, 108, 54-64.