ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Identifying Cancerous Tissues Using Dynamic Neural Network with Reuse

K.Nagi Reddy
Department of CSE,
RRSCET, Medak(Dist),
Andhra Pradesh, India, 502 300.

Dr.B.Ravi Prasad
Department of CSE,
RRSCET, Medak(Dist),
Andhra Pradesh, India, 502 300.

Dr.K.Suvarchala
Professor,
Department of CSE,
St.Mary's Group of Institutions,
Hyderabad, India, 501 502.

Abstract:

In this paper we present a system to identify the cancerous tissues using Hopfield based Dynamic Neural Network. The DNN has a composite structure consisting of several basic nodes. The basic nodes are completely connected Hopfield networks and they are similar to each other. The DNN with relative pruning and reuse is used for the present task as it is having the properties like associative memory with 100% recall, large storage capacity, avoiding spurious states and converging only to user specified states. Since the DNN with reuse operates with binary data while the information in Cancer Database is in decimal form, the decimal digits are mapped a binary string called descriptor. The network is trained by descriptors (binary strings), which are obtained through the steps indexing through thresholding. These descriptors are used as exemplar patterns to store them in the DNN with reuse by training. The associative memory property of the network makes it to recall a closest pattern of the given query pattern among the memorized patterns. The experimental results show that the percentage of accurately identifying a cancerous person depends on the length of descriptor.

1. Introduction:

Associative memory is the idea that relates the association between the stored patterns and the input patterns. The main features of ANN implementation for associative memory is the use of retrieval based on content instead of retrieval based on the address of the data, also may be approximate instead of perfect matching. Hence this research of identifying a cancerous tissue from the large data available in the "Cancer Database".

The DNN with Reuse that is proposed by us will be accurately suitable to the present tasks as it is also having the characteristics like 100% accurate recall, high storage capacity, escaping from spurious states and settling to the user defined states. The need to find an approximate match to a string arises in many practical problems [3]. Identifying cancerous person is such a system where in the documents are stored as a set of decimal digits. We have acquired a large data base of different people consisting of mainly two classes of data i.e., a normal tissue data when it is exposed to parallely polarized light and that of the perpendicularly polarized light. For convenience of DNN recognition we have divided the data into four different classes.

i) Normal Tissue Data exposed to Parallely polarized light. ii) Normal Tissue Data exposed to perpendicularly polarized light. iii) Cancer Tissue Data exposed to Parallely polarized light. iv) Cancer Tissue Data exposed to perpendicularly polarized light each set of data viz., cancer tissue and normal tissue data, when exposed to parallel or perpendicular Polarized light consists of 200 samples. Each set of data available is read by type casting it as character and each set of data is considered as a one dimensional image size of 1x200 pixels. Each set of data consisting of 200 samples is compressed through wavelets and down sampled which results in a data set contains 50 samples and is treated as binary string (descriptor) of size 1x50 pixels. Now, the network is trained by descriptors which are obtained through wavelets and thresholding.

These descriptors (binary strings) are used as exemplar patterns to store them in the DNN with reuse by training. The descriptor of the query is presented to the network as a test pattern. The associative memory property of the network makes it to recall a closest pattern of the given query pattern among the memorized patterns.

2. Dynamic Neural Network (DNN):

The DNN is a massively parallel and distributive network [7]. It is based on the principle of associative memory but unlike other associative memories it does not allow spurious stable states. It demonstrates a novel idea of order sensitive learning which gives preference to chronological order of presentation of the exemplar patterns. The network is dynamic as it changes its architecture during the process. The DNN contains a set of basic nodes which are identical Hopfield networks.

However, during the learning phase, the nodes acquire different synaptic weights. The basic nodes are grouped together in a hierarchical organization. Each group has a designed basic node called the leader. When some pattern is presented to DNN, it is presented to all the basic nodes at the lowest level of the hierarchy of nodes. Each node reaches its own stable state based on the common input and individual synaptic weights. These nodes transmit their stable states to their respective leader.

Here, the DNN adopts a pruning mechanism and retains only the leader nodes. These leader nodes are treated as the basic node of the next level of the hierarchy and they send the resulting states to the leader nodes at the next level of the hierarchy after reaching the stable state.

The process proceeds in this way till the whole network reaches a single stable state. In one cycle, the available basic nodes carry out the state transition function with the given synaptic matrices and in the next cycle these nodes communicates among themselves to change the synaptic weights.

At this stage the network is pruned to retain only the leader nodes of the current level of the hierarchy. In [7] it is shown that by this process the network can accomplish very efficient associate recall without any spurious states or any sort of memory limitations, which are two main drawbacks of the Hopfield model of the neural networks. However, the DNN model overcomes the above drawbacks by retaining back the pruned nodes and reusing them in subsequent iterations [6].

3. Our Approach:

Each set of data viz., cancer tissue and normal tissue data, when exposed to parallel or perpendicular polarized light consists of 200 samples as shown in figure 1. Now each set of data available is read by type casting it as character and each set of data is considered as a one dimensional image size of 1x200 pixels as shown in figure 2.

D_1	D_2	D_3	D_4	…	D_{200}

Figure 2. Data set as image of 1x200 pixels

Each set of data consisting of 200 samples is now read as a one dimensional image and is subjected to compression through wavelets [2]. Here we have used low pass filter of Daubechie's W_6 wavelet for the wavelet transform implementation for compressing the data.

The low pass filter of W_6 has two vanishing movements, which means the transform coefficients will be close to zero for any signal that can be described by a polynomial of degree 2 or less [1].

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

D_1 =1870	D_{41}=1680	D_{81}=1130	D_{121}=510	D_{161}=440
D_2 =1860	D_{42}=1930	D_{82}=1090	D_{122}=720	D_{162}=390
D_3 =1700	D_{43}=1690	D_{83}=1360	D_{123}=520	D_{163}=460
D_4 =1530	D_{44}=2050	D_{84}=1140	D_{124}=670	D_{164}=450
D_5 =1670	D_{45}=1700	D_{85}=1290	D_{125}=640	D_{165}=420
D_6 =1550	D_{46}=1640	D_{86}=1230	D_{126}=680	D_{166}=360
D_7 =1470	D_{47}=2140	D_{87}=1280	D_{127}=740	D_{167}=500
D_8 =1790	D_{48}=1870	D_{88}=1320	D_{128}=610	D_{168}=380
D_9 =1750	D_{49}=1640	D_{89}=1220	D_{129}=770	D_{169}=320
D_{10}=1990	D_{50}=1920	D_{90}=1200	D_{130}=550	D_{170}=360
D_{11}=1860	D_{51}=1700	D_{91}=1100	D_{131}=640	D_{171}=330
D_{12}=2190	D_{52}=1570	D_{92}=1250	D_{132}=590	D_{172}=330
D_{13}=2100	D_{53}=1730	D_{93}=1030	D_{133}=560	D_{173}=410
D_{14}=2020	D_{54}=1610	D_{94}=890	D_{134}=560	D_{174}=330
D_{15}=2180	D_{55}=1400	D_{95}=1000	D_{135}=460	D_{175}=280
D_{16}=2230	D_{56}=1790	D_{96}=1050	D_{136}=460	D_{176}=290
D_{17}=2510	D_{57}=1650	D_{97}=1060	D_{137}=490	D_{177}=320
D_{18}=2400	D_{58}=1840	D_{98}=1050	D_{138}=520	D_{178}=450
D_{19}=2500	D_{59}=1620	D_{99}=1060	D_{139}=470	D_{179}=340
D_{20}=2520	D_{60}=1750	D_{100}=910	D_{140}=550	D_{180}=300
D_{21}=2210	D_{61}=1940	D_{101}=910	D_{141}=610	D_{181}=290
D_{22}=2170	D_{62}=1950	D_{102}=790	D_{142}=500	D_{182}=330

D_{23}=2230	D_{63}=1610	D_{103}=960	D_{143}=470	D_{183}=310
D_{24}=2300	D_{64}=1800	D_{104}=930	D_{144}=400	D_{184}=340
D_{25}=2150	D_{65}=1840	D_{105}=1010	D_{145}=490	D_{185}=380
D_{26}=2240	D_{66}=1760	D_{106}=950	D_{146}=550	D_{186}=290
D_{27}=2770	D_{67}=1990	D_{107}=800	D_{147}=490	D_{187}=360
D_{28}=2510	D_{68}=2040	D_{108}=830	D_{148}=550	D_{188}=300
D_{29}=2150	D_{69}=2150	D_{109}=830	D_{149}=470	D_{189}=280
D_{30}=2210	D_{70}=1870	D_{110}=840	D_{150}=520	D_{190}=270
D_{31}=2210	D_{71}=1610	D_{111}=750	D_{151}=550	D_{191}=420
D_{32}=1880	D_{72}=1540	D_{112}=730	D_{152}=440	D_{192}=360
D_{33}=2030	D_{73}=1450	D_{113}=890	D_{153}=430	D_{193}=270
D_{34}=1910	D_{74}=1540	D_{114}=670	D_{154}=350	D_{194}=350
D_{35}=2010	D_{75}=1490	D_{115}=790	D_{155}=470	D_{195}=290
D_{36}=1990	D_{76}=1560	D_{116}=880	D_{156}=460	D_{196}=330
D_{37}=1890	D_{77}=1210	D_{117}=670	D_{157}=410	D_{197}=410
D_{38}=1860	D_{78}=1540	D_{118}=730	D_{158}=380	D_{198}=340
D_{39}=2020	D_{79}=1360	D_{119}=710	D_{159}=410	D_{199}=300
D_{40}=1770	D_{80}=1480	D_{120}=690	D_{160}=480	D_{200}==330

Figure 1. Cancerous Tissue data of a person with parallely polarized light

The filter coefficients for H (low pass filter) of W_6 are: h(n)= 0.332670552950, 0.806891509311, 0.459877502118, -0.135011020010, -0.085441273882, 0.035226291882.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

Now each data set is compressed once i.e., filtering through the low pass filter and down sampling which gives the data set of the size 100 samples. Repeating the same process of compression, which results in each set of data which contains 50 samples and is treated as an image size of 1x50 pixels as shown in the figure 3.

D_1	D_2	D_3	D_4	…	D_{50}

Figure3.Sampledata set format after compression

3.1. Thresholding:

The dimension of each data is 1x50 pixels. Each pixel having data values $D_{1,}$ D_2, …, D_{50}. Thresholding [4,5] is done for this compressed data file as described below. Taking sum of all the 50 data values, an average value is evaluated. Average Value = Sum of all the 50 Sample Data Values/Number of Samples. Now, each sample value of the compressed data is compared with the average value and is replaced by 1’s and 0’s, if greater and less than the average value, respectively. This is how the compressed data file of each set of data in the database is converted into a binary string (descriptor) of 50 bits. Through this method, each data set in the database is mapped to a descriptor (binary String).

3.2. Associative Memory using DNN with Reuse:

Identifying a cancerous tissue of a person uses a binary string or descriptor with each record in the database. The descriptor or binary sting is a bit encoding of the values used to retrieve the record. The binary string of each record is obtained by compression scheme using wavelet decomposition and thresholding technique. These binary strings of each record in the database are memorized in DNN with reuse by training. A query descriptor or binary string is formed using the same encoding techniques.

S_1 = 11111111110011001111111111100000111000000001111110

S_2 = 11111111101100010011111111111100011111111111001111

S_3 = 00001000000011111111111100000000011100000000000000

S_4 = 11111011100000111001101000100111100111111111100111

S_5 = 11111110001001110111100111111111011111101111010000

S_6 = 11111110000001100101000001101011100001100000000001

S_7 = 01111111111100011111011100111110100010000000000000

S_8 = 00000001100000001111011000101101111110011111100000

S_9 = 11111111100101101000011011110110101011011110000101

S_{10}= 00000110001011010010011111101100001010000100011101

S_{11}= 11101111001111010000000001000000000111100001000011

S_{12}= 00001100000111100010010111011111111000000110000001

S_{13}= 11111111111111111001101100000111001111101101111001

S_{14}= 00000000011111111110111111100001010101111101000100

S_{15}= 11111011011100111100001000010110011101110000011011

S_{16}= 10010000110000000100000001001001001100011000000001

Figure 4.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

And is used as a test pattern. The DNN with Reuse retrieves the descriptor of the record which approximately matches with that of the query. It is possible that a record descriptor or binary string matches a query descriptor but the corresponding record does not satisfy the query. Such occurrence is referred to as false drop. The probability of a false drop can be made arbitrarily small by appropriate size of the descriptor.

4. Experimentation:

The records in the database are divided into four cases as (i) Normal Tissue data exposed to parallely polarized light. (ii)Normal Tissue data exposed to perpendicularly polarized light. (iii)Cancerous Tissue data exposed to parallely polarized light. (iv)Cancerous Tissue data exposed to perpendicularly polarized light Memorizing in DNN with reuse by training for each record in the database, for all the above said cases a descriptor or a binary string is obtained containing 50 bits each. The following example illustrates the network behavior for approximate retrieval in one of the four cases of the database of cancerous or normal person. Let us consider sixteen records, each of which belongs to one case (case (i)), whose descriptors are given as S_i in figure 4.

Let us assume that the user wants to retrieve a record based on the query

X=
11011110110011001111110001100000111001010001 111110.

The computational steps involved in matching the closest pattern to the test pattern through DNN with Reuse is presented as Data matching Algorithm for cancer database(Figure 5). The closest pattern to the test pattern from the cancer database is retrieved after 5 iterations.

The output of the network is
S_1=
11111111100110011111111111000001110000000011 11110
Which is the descriptor of the required data match in the cancer database

Input: C-Cancer database consists of records D_1, D_2, D_3…. D_n, Q – Query

Output: B-Cancer data Corresponding Query Q.

Procedure Cancer (D_1, D_2, …. D_n, Q: in; B: out)

1 for i=1 to n do
2 Filter lx() /* Data will be compressed to the desired level*/
3 Threshold( );
4 S_i = Descriptor (D_1) /*binary string of each of the document in the database*/
5 End do
6 X=descriptor (Q)
7 O = DNN (S_1, S_2……..S_n, X)
8 Look up in cancer database C for a data B whose descriptor matches with O
9 End

Figure.5. Data Matching Algorithm for Cancer Database

5. Experimental Results:

We evaluated the performance of the DNN with reuse for different sizes of descriptor. It is observed that the accuracy of retrieval is dependent on the size of the descriptor and also on the thresholding technique. In the event of occurrence of false drop, it may happen that two distinct records may correspond to the same descriptor. Experiments were carried out for different query inputs to evaluate the performance of DNN with reuse in the context of identifying a cancerous person. The percentage of retrieval is calculated for different sets of input queries. Percentage of retrieval = No. of accurate retrieval/No. of input queries x 100.

ISSN No: 2348-4845
International Journal & Magazine of Engineering, Technology, Management and Research
A Peer Reviewed Open Access International Journal

The percentage of accurately identifying a cancerous person or not is around 71%. It is also observed that with the descriptor (binary string) of length more than 50 bits, the percentage of matching to the correct record has been increased to 75%. As the percentage of retrieval is dependent on the descriptor size, it is concluded that the DNN can perform an efficient approximate search if appropriate technique can be developed to find the descriptor. Since the present study aims at establishing the suitability of neural network for matching, no attempt is made here to investigate the most efficient method for computing the descriptor. However, if proper descriptor computation scheme is employed the retrieval performance is believed to be improved.

6. Conclusion:

In this paper, matching capability in databases of DNN with relative pruning and reuse as an associative memory is discussed. The procedure for obtaining a binary string for each record in the database is presented. Experiments were carried out on the performance of the network by considering the case study, namely cancer data base. It is shown that the neural network based retrieval scheme works efficiently even for very large databases.

References:

[1]Croiser A, Esteban D, Galand C(1976) Perfect channel splitting by use of interpolation decimation tree decomposition techniques. International conference on information science and systems.

[2]Daubechies, I. Orthonormal bases of compactly supported wavelets. Common Pure and Applied Math, 41: 909-996, 1988

[3]Hall, P.A.V. Dowling, G.R. Approximate String Matching. ACM Computing Surveys, 12,4, 381-409

[4]Knuth, D.E. The Art of Computer Programming Vol. 3: Sorting and Searching. Addision Wesley, Reading, Mass, 1973.

[5]Knuth, D.E., Morris, J.H., Pratt, V.R. Fast pattern Matching strings, SIAM Journal of Computing, 6, 2, 323-350, 1977.

[6]Arun K Pujari, C.Dhanunjaya Naidu, B C Jinaga. An Adaptive character recognizer for Telugu Scripts using Multiresolution Analysis and Associative Memory. Proceedings of the Indian Conference on Computer vision, Graphics & Image Processing 235-240, 2002

[7]Rao, M.S., Arun K Pujari and sreenivasan. B. A New Neural Network architecture with associate memory, Pruning and order sensitive learning. International Journal of neural Systems 9, 4, 351-370, 1999

[8]Hopfield, J.J. Neurons with graded response have collective computational abilities like those of two state neurons. Proceedings of National Academy of Sciences, USA, 81, 3088-3092, 1984.