Deep Neural Nets Revolutionizing and Automating Healthcare Globally

Aviruk Chakroborty
BE Mechanical Engineering

Abstract: The next five years are going to see an exponential growth of diagnostic support system on the backbone of deep neural nets. The decade of 2010 to 2020 can be projected as the Cambrian explosion in the field of artificial intelligence and machine learning. This paper explores the immediate innovations possible in the field of medical analytics with the help of deep neural nets which will be changed irrevocably with the application of neural networks.

This is possible due to explosion of data availability, the reduction in computing cost, creation of GPU clusters and solution of the vanishing gradient problem in deep neural nets.

Introduction

Deep learning is an emerging field which has come into prominence in the last decade due to sudden exponential rise in data sources (90%) and falling computation power cost with the advent of GPU clusters. Deep learning will revolutionize healthcare in a multi-pronged fashion and will become a standard solution set within the next 3-5 years.

Deep Learning applications in healthcare spans across wide variety of applications starting from analysis of patient data in EMRs to analysis of images starting from x-ray to Citi scans to MRI data, they are also used extensively in the field of pathology to segregate blood samples based on their structure and helps us create personalized medical solutions with the use of genome mapping and its analysis.

Deep learning applications which in other fields are blamed for cutting away jobs are welcome with open arms in the healthcare industry as it essentially increases the patient care quality and reduces diagnostic errors. Thus, making the healthcare industry more efficient.

Deep learning helps in lowering of cost of diagnosis and analysis along with decreasing the analysis turnaround time and thus positively improving the patient experience. No longer do people must wait for days to get their medical analysis and spend that time in anxiety. Today, they receive them instantaneously when deep learning based diagnostic decision support system is used in conjecture with experienced diagnosticians.

Applications of deep learning along with other machine learning applications are used to facilitate various solutions in the field of medical imaging such as computer aided diagnosis, image segmentation, image registration, red flag/anomaly detection in images and using computer vision deep learning models to conduct image guided therapy. They are most typically used in radiology and pathology departments today but using deep nets to identify the emotional state of patients and ranking them on a global 1-10 pain scale will help even general practitioners and internal medicine practitioners in diagnosing patient condition.

Deep neural nets which are used to do medical diagnosis are CNN's i.e. Convolutional Neural Nets in which there are multiple hidden layers and each pixel is analyzed in multiple layers in a recursive fashion. CNNs today are the most cutting edge solutions in the field of Computer Vision. CNNs are used even in the field of automated driving and cars.

The framework of medical image analysis using CNN consists of three major steps:

Step One - Neural network model is created and hosted on a large GPU cluster and then a large volume of past patient data which are labelled are fed to the algorithm as a training set

The algorithm runs on this data set millions of time till the CNN is able to identify and distinguish between different type of images. For e.g. the CNN can easily distinguish between a normal hand x-ray
and a x-ray of a broken arm. This is called supervised learning.

Step 2 - We post training of the neural network feed the users diagnostic images such as x-ray, MRI, pathology images to the neural net and ask them to classify the images in various detailed sub categories and along with a distributed probability score. This image analysis along with probability score is provided to the doctor, radiologist, pathologist for analysis and confirmation and post green light from the medical professional the final analysis is shared with the patient.

Step 3 - Neural nets have to keep learning based on new data sets to become better and hence the results of the neural net along with the final analysis delivered by the doctor is fed to a reinforcement learning based neural network whose unique architecture makes it capable of learning and improving with the introduction of each new positive or negative data point. These optimize waits are then used to retrain the primary neural network.

Privacy can be saved using homomorphic encryption - where encrypted data is run on a CNN and insights are created.

**Literature Review**

Che and others from their paper “Distilling Knowledge from Deep Networks with Applications to Healthcare Domain” taught us a novel knowledge-distillation approach called Interpretable Mimic Learning, to learn interpretable phenotype features for making robust prediction while mimicking the performance of deep learning models. The framework uses Gradient Boosting Trees to learn interpretable features from deep learning models such as Stacked Denoising Autoencoder and Long Short-Term Memory. While from the paper on “Deep learning for healthcare decision making with EMRs” written by Liang and others, we understood that deep models can simulate the thinking procedure of human and combine feature representation and learning in a unified model. A modified version of convolutional deep belief networks can be used as an effective training method for large-scale data sets. Then it is tested on two instances: a dataset on hypertension retrieved from a HIS system, and a dataset on Chinese medical diagnosis and treatment prescription from a manual converted electronic medical record (EMR) database. The experimental results indicate that the proposed deep model is able to reveal previously unknown concepts and performs much better than the conventional shallow models. Hence in the field of unsupervised learning we can use the insights learnt from this paper.

**Methodology**

**Basically, Homomorphic encryption** is a form of encryption which makes possible computations to be carried out on encrypted data i.e. cyphertext, thus generating an encrypted result which, when decrypted, matches the result of operations performed on the plaintext.

Since homomorphic encryption supports only additions and multiplications and the standard ANN formula is a sigmoid function which has division and exponential function we use mathematical/numerical approximation to compute the division and exponential functions.

\[
f(x) = \frac{1}{1 + e^{-x}}
\]

Since the weights are not encrypted, it is possible to use the more efficient plain multiplication operation and in some networks, we can also add a bias term to the result of the weighted sum. Even though the process of encryption and decryption of the ANN takes considerable amount of time but the privacy issue which this process solve makes it a valuable tool in the field of accessing large public data sets for training the neural nets.
1. Faster Diagnosis

GPU enabled CNNs which are used to crunch medical images help diagnosticians reach decisions much faster. Medical images like CT scans, MRIs, and X-rays are one of the most important tools doctors are utilizing in diagnosing conditions ranging from cancer to tuberculosis to spine injuries to heart disease.

But, since analyzing medical images is often a difficult and time-consuming task, researchers and medical experts have already started using GPU-accelerated Deep Learning frameworks which automate analysis and increase the accuracy of diagnosis.

In a proof of concept application of analyzing lung cancer using data from pet scans we were able to reach a very high level of prediction and were mathematically able a path of reaching 97% accuracy if the data sets made available were over 100,000 units. Today even in advanced nations like USA the rate of mis-diagnosis and false prescription is in the range of 10-20 percent.

According to NCPA Nat Centre for policy analysis: An estimated 10 percent to 20 percent of cases are misdiagnosed, which exceeds drug errors and surgery on the wrong patient or body part, both of which receive considerably more attention.

- One report found that 28 percent of 583 diagnostic mistakes were life threatening or had resulted in death or permanent disability.
- Another study estimated that fatal diagnostic errors in U.S. intensive care units equal the number of breast cancer deaths each year ~ 40,500.

According to doctors, misdiagnosis has occurred for quite some time. As far back as 1991, Harvard University found that misdiagnosis accounted for 14 percent of all adverse events and that 75 percent of these errors involved negligence.

With the advent of deep learning based diagnostic support system there will be an exponential drop in such error ratios but for reaching such low error ratios extremely large data sets have to be used for training the neural net.

2. Genomics for Personalized Medicine

Genomics analysis consists of unprecedented quantities of patient gene structure data, giving medical practitioner and researchers the ability to study how genetic factors like mutations lead to disease and those insights create personalized medical solutions for each patient. When the genome was first map the price of sequencing the first human genome was 2.7 billion dollars and needed nearly a decade. Today with the advent of advance algorithms and extremely high computational power at low cost the cost is a few 1000 dollars and time is a few days.

This make personalized medicine possible for the first time in human history.

A single byte (or 8 bits) can represent 4 DNA base pairs. In order to represent the entire diploid human genome in terms of bytes, we can perform the following calculations: 6×10^9 base pairs/diploid genome x 1 byte/4 base pairs = 1.5×10^9 bytes or 1.5 Gigabytes, the only way to read a 1.5 gb file is using machine learning and deep neural nets.

Deep Learning can help in tailoring personalized or "precision" medicine, with treatments that are customized according to a patient’s genomic makeup. Deep Learning can also help in:
3. Computer-Aided Diagnosis (CADx)

One of the major CADx applications is the differentiation of malignancy/benignancy for tumors. Today large data sets availability of labelled tumors has propelled the machine learning community towards creation of extremely efficient tumor classification models. These models when integrated with hospital information system where the input data comes from EMRs will become a standard diagnostic process within the next 5 years.

Deep Learning techniques can potentially change the design paradigm of the CADx framework. When unsupervised neural net applications are run on large sets of medical data it can directly uncover features from the training data and provide an efficient hypothesis which can be either substantiated or refuted by clinical diagnosticians.

Today Deep Learning is transforming the world of medicine. Deep Learning systems can not only allow physicians and other healthcare providers in faster diagnoses but can also help in reducing uncertainty in their decisions thereby avoiding costs and hazards and saving time.

4. Improvement in Patient Experience

Deep neural nets when applied on computer vision is also efficient in improving the patient experience in hospitals. For example CNN based face recognition algorithm helps the front desk customer service executive of each hospital recognize every patient before they even reach the help desk and can be used to greet them by their first name and provide the right directions and recommendations.

5. Touchless Diagnostics

Deep neural net will revolutionize the field of touchless diagnostics where simple video feeds are used to analyze the heartbeat, pulse, emotion and paying scale through non - evasive modalities.

Mathematical Framework:

About ANN, CONVOLUTIONAL DEEP NETS Application in revolutionizing healthcare:

Deep Nets are ANN's with multiple hidden layers executing back propagation

In Artificial Neural network, there is an input layer say \(b_i=1,2,3,...,n\), there comprises hidden layer for different input layers which adds up weighed
gradient say \( w_{kj} \), \((k,j)\) are the number of hidden layers for each ‘k’ nodes. For each hidden layer

\[
a_j^l = \sigma \left( \sum_k w_{kj} a_k^{l-1} + b_j^l \right) \quad \text{(1)}
\]

- Let us take a cost function say

\[
J(\theta) = - \frac{1}{m} \sum_{i=1}^{m} \sum_{k=1}^{K} y_k^{(i)} \log(h_{\theta}(x^{(i)}))_k + \left(1 - y_k^{(i)}\right) \log \left(1 - \left(h_{\theta}(x_k^{(i)})\right) \right) + \frac{\lambda}{2m} \sum_{l=1}^{L} \sum_{k=1}^{s_l} \left[ (\theta_{jl}^{(l)})^2 \right]
\]

- We minimize the cost function \( \min_{\theta} J(\theta) \), minimizing to \( \frac{\partial}{\partial w_{ij}} J(\theta) \) to do so we need to define as below:

- The weights over input and output layer form a linear equation as follow:

\[
Y = x_i w_i + x_j w_j \quad \text{(2)}
\]

- Now as there are hidden layers each weight is being added to each node and  the gradient this forms error \( E = (t-a)^2 \) \( \text{(3)} \)

- And \( a_i = \varphi \left( \text{net } j \right) = \varphi \left( \sum_{k=1}^{n} w_{kj} a_i \right) \)

- Suppose we define, \( \text{Sigmoid function } \varphi \left( z \right) = \frac{1}{1+e^{-z}} \) and further taking derivative we obtain,

\[
\frac{d\varphi(z)}{dz} = \varphi \left( z \right) \left( 1 - \varphi \left( z \right) \right) \quad \text{(4)}
\]

- \( E = \frac{1}{2} \sum_k (t_k - a_k)^2 \)

- Now as we take a step backward, we obtain weighted gradient say \( w_{ij} \),

- So now to reduce difference weight,

\[
\Delta W \propto - \frac{\partial E}{\partial W} \Rightarrow \Delta w_{kj} \propto - \frac{\partial E}{\partial w_{kj}}
\]

\[
\Delta w_{kj} = - \frac{\partial E}{\partial w_{kj}} \frac{\partial a_k}{\partial \text{net}_k} \frac{\partial \text{net}_k}{\partial w_{kj}}
\]

Further, taking derivative with respect to error we obtain,

\[
\frac{\partial E}{\partial a_k} = \frac{\partial (1/2 (t_k - a_k)^2)}{\partial a_k} = - (t_k - a_k) \quad \text{(6)}
\]

\[
\frac{\partial a_k}{\partial \text{net}_k} = \frac{\partial (1+e^{-\text{net}_k})^{-1}}{\partial \text{net}_k} = \frac{e^{-\text{net}_k}}{(1+e^{-\text{net}_k})^2} \quad \text{(by taking partial derivative w.r.t } a_k (1 - a_k)\text{)}
\]

\[
\frac{\partial \text{net}_k}{\partial w_{kj}} = \frac{\partial (w_{kj} a_j)}{\partial w_{kj}} = a_j \quad \text{(7)}
\]

\[
\Delta w_{kj} = \varepsilon \left( t_k - a_k \right) a_k (1 - a_k) a_j = \varepsilon \delta_k a_j \quad \text{where, } \delta_k = \left( t_k - a_k \right) a_k (1 - a_k)
\]

\[
\Delta w_{kj} \propto \left[ \sum_k (t_k - a_k) a_k (1 - a_k) w_{kj} \right] a_j (1 - a_j) a_i
\]

\[
\Delta w_{kj} = \varepsilon \left[ \sum_k \delta_k w_{kj} \right] a_j (1 - a_j) a_i
\]

\[
\Delta w_{kj} = \varepsilon \delta_j a_i \quad \text{where, } \delta_j = \left[ \sum_k \delta_k w_{kj} \right] a_j (1 - a_j) \quad \text{(8)}
\]

Now applying, equation (8) in equation (2), in Newton’s forward differential interpolation equation \( [f_a = f_0 + a \Delta + \frac{1}{2!} a(a-1) \Delta^2 + \frac{1}{3!} a(a-1)(a-2) \Delta^3 + \cdots] \)
we obtain,

\[ a_{kj} = \sigma'(a_k^t x_i) \sum_{k=1}^k \beta_{jk} \delta_{kj} \ldots (9) \]

The continuous learning model using Reinforced learning:

- For above taken set of equations in (1), we apply Markov’s Decision Process for maximization of value of Cost function:
  
  \[ J(\theta) = -\frac{1}{m} \left[ \sum_{i=1}^m \sum_{k=1}^k y_k^{(i)} \log(h_{\theta}(x^{(i)}))_k + (1 - y_k^{(i)}) \log(1 - (h_{\theta}(x^{(i)}))) \right] + \frac{\lambda}{2m} \sum_{i=1}^m \sum_{j=1}^{s_{i+1}}(\theta_{ji}^{(i)})^2 \]

- Further, taking maximum value of the cost function as:
  
  \[ J(\theta) = \arg \max_x \frac{1}{m} \left[ \sum_{i=1}^m \sum_{k=1}^k y_k^{(i)} \log(h_{\theta}(x^{(i)}))_k + (1 - y_k^{(i)}) \log(1 - (h_{\theta}(x^{(i)}))) \right] + \frac{\lambda}{2m} \sum_{i=1}^m \sum_{j=1}^{s_{i+1}}(\theta_{ji}^{(i)})^2 \]

- Now from the equation (9) we minimize it to equation,
  
  \[ J(\theta) = (X_t, R_t) \text{ with respect to time } t' \]

- Applying Markov’s Decision Process,

  - We have to take time, \( t+1 \),
  - So the equation of ‘Q’, will be,
  
  \[ Q(s, a_t) = \alpha[y_{t+1} + \gamma \max_a Q(s_{t+1}, a)] \ldots \ldots (10) \]

- The value of Q at time \( t=0 \), is as follows which are as called Bellman equations:
  
  \[ Q(s, a) = \sum_a T(S, a, s') [ R(S, a, s') + \gamma V(s') ] \ldots \ldots (11) \]

  \[ Q_{t+1} = \sum_a T(S, a, s') [ R(S, a, s') + \gamma \max_a' Q_k (s', a') ] \ldots \ldots (12) \]

- Thus, after combining and equating we can take partial derivative with respect to sum of errors we obtain,

  \[ \frac{\partial q}{\partial (R.S.)} = -\frac{\partial \sum_a T(S, a, s') R(S, a, s') + \gamma V(s')}{\partial (R.S.)} \ldots \ldots (13) \]

- The optimality of the equations will be done as follows:

  \[ v_a(s) = E_{\pi} [ G_t | S_t = s ] \]

  \[ v_a(s) = E_{\pi} [ \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | S_t = s ] \]

  \[ v_a(s) = E_{\pi} [ R_{t+1} \sum_{k=0}^{\infty} \gamma^k R_{t+k+1} | S_t = s ] \ldots \ldots (14) \]

Convolutional Neural Nets: for Image classification

- In Convolutional Nets, on the formula \( y = w_1 x_1 + w_2 x_2 \) where, \( y \) is maximum when \( (x_1, x_2) = (1,0) \) and when \( (w_1, w_2) = (1,-1) : y = x_1 - x_2 \) the input layer is taken.

- Now, the maximum pooling is done with an (N-m+1) * (N-m+1) hidden layers and its dimensions are considered on width, height and depth.

- Where, ReLU is on Sigmoid function := \( \frac{1}{1+e^{-x}} \) and hyperbolic tangent function: \( f(x) = \tanh(bx) \)

- In general form, Convolutional Net formula is:
\[ x^l_j = f \left( \sum_{i \in M_l} x^{l-1}_i \ast a^l_{ij} + b^l_j \right) \] ............(14)

- The process takes place in Forward Propagation:

\[ x^l_{ij} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \frac{\partial x^l_{(i-a)(j-b)}}{\partial y^l_{i-j}} = \omega_{ab} \] ............(15)

- And max pooling layers \( N_k \) and \( y^l_{ij} = \sigma \left( x^l_{ij} \right) \) in Forward Propagation

- And in Backward Propagation, \( \frac{\partial E}{\partial w_{ab}} = \sum_{i=0}^{N-k} \sum_{j=0}^{N-k} \frac{\partial E}{\partial x^l_{i-j}} y^{l-1}(i+a)(j+b) \) ........(16)

- Now, \( \frac{\partial E}{\partial x^l_{ij}} = \sigma' (x^l_{ij}) \) and \( \frac{\partial E}{\partial y^l_{ij}} = \sum_{a=0}^{m-1} \sum_{b=0}^{m-1} \frac{\partial E}{\partial x^l_{i-a}(j-b)} \omega_{ab} \) .......(17)

- In Deep Nets, the stochastic gradient descent: \( \omega_{ij}(t+1) = \omega_{ij}(t) + \eta \frac{\partial C}{\partial w_{ij}} + \epsilon(t) \),

Where, \( \eta = learning\ rate \), \( C = cost\ function \) and \( \epsilon(t) = stochastic\ term \)

- This gradient gives binary output thus to get an activation function is applied say, \( p_j = \frac{\exp(x_j)}{\sum_k \exp(x_k)} \) and cross entropy is applied then after \( C = - \sum_j d_j \log(p_j) \). This is used when there is supervised learning.

**Conclusion**

Deep learning solutions will start becoming ubiquitous only when integration of HIS, EMR and deep learning frameworks occur. The data flow has to be seamless between these three modules and the output of the deep neural net has to be treated as a recommendation to the diagnostician. As stated in the paper the rise of various analysis framework like homomorphic encryption will give rise to more efficient neural net as solving the privacy issues, the data set for training the nets will increase exponentially. We foresee an overall drastic error reduction in mis-diagnosis in hospital due to the amalgamation of medical ANNs in the decision framework which will have tremendous impact not only on quality of patient care but will reshape the entire medical malpractice insurance industry.

**References**