Introduction to Data Analytics

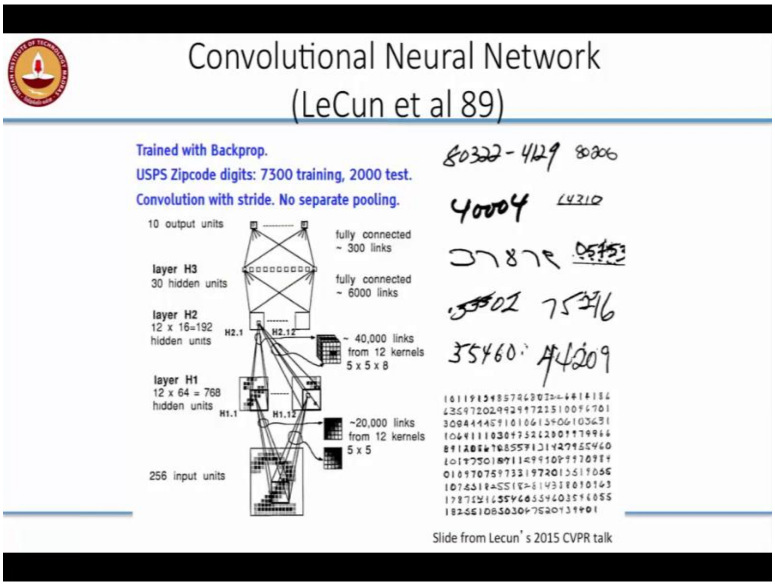
Prof. Nandan Sudarsanam and Prof. B. Ravindran
Department of Management Studies and
Department of Computer Science and Engineering
Indian Institute of Technology, Madras

Module - 06

Lecture – 36

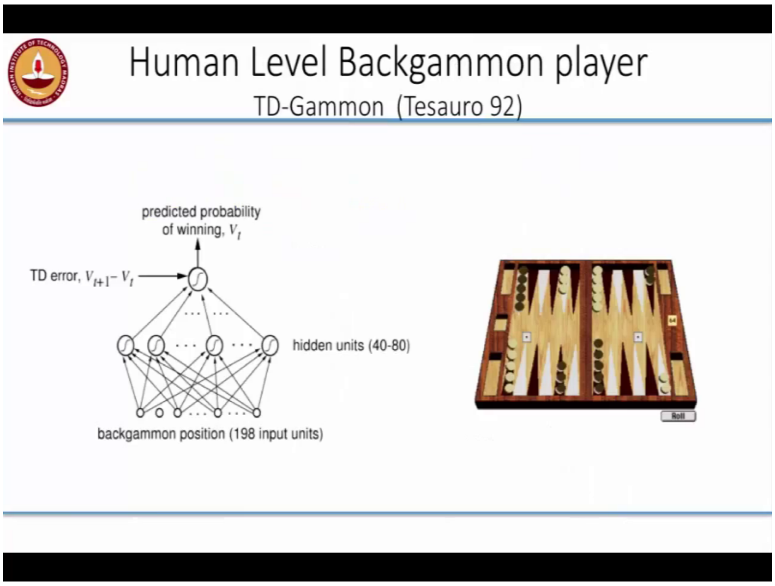
Deep Learning

(Refer Slide Time: 00:10)



So, in the previous module we saw how we can use that propagation in order to find the weights of a neural network; upper three layered neural network. So, one of the earning success stories of back propagation was learning to recognize hand written digits in address; right. So, this is work done by on LeCun back in 89, and essentially they trained a slightly more complex network architecture call the convolutional neural network; in order to recognize hand written digits. You can see the variation in the digits, that their network manage to handle; the fatly well.

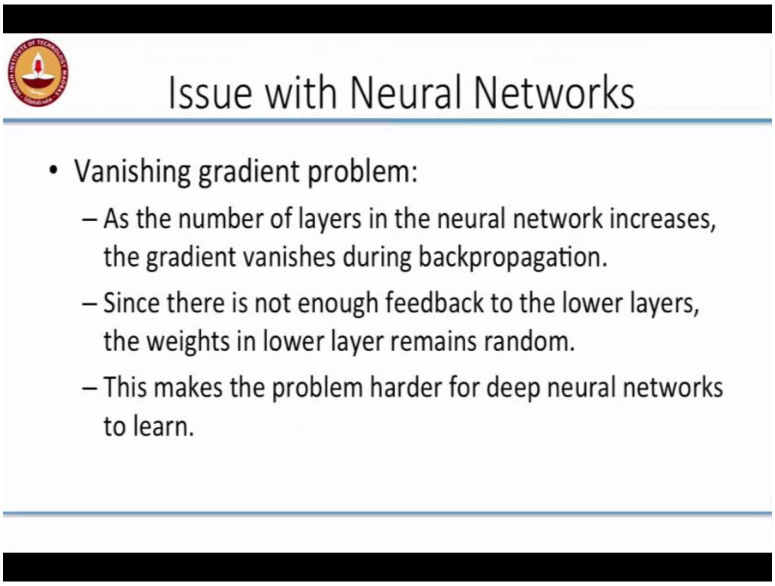
(Refer Slide Time: 00:56)



So, another early success story of neural networks was the Backgammon Player built by Gammon Tesauro from IBM, T. J. Watson labs in 1992. So, he built 2 versions of this backgammon player on called the Neurogammon which came in 89, which was essentially neural network trend using back propagation in supervise learning manner; in order to play a game of backgammon. So, it is like Ludo of people know about it. So, you throw a dice, you throw a dice and depending on that die roll; you move your coins around. So, that is a white and black side. So, the idea is to move all your coins of the board by repeatedly rolling the dice. So, the TD-Gammon player essentially use the 3 layer, standard 3 layer neural architecture and was trying using a specific form of what is called reinforcement learning. We look at reinforcement learning in the later module, but here is reinforcement learning in order to generate the error signals and, but then use back propagation to train the weights of the hidden unit. And you manage to build Backgammon player which was able to beat human Backgammon champions in game play.

So, this was some of the early success and lot of interests was being aspect on looking at neural networks of solving variety of different problems. But then again people discovered a 2nd drawback with neural networks. So, the one thing was that I toss incredibly hard to train a neural network, because you had so many parameters you had to treat them appropriately. So, that they desired to deserved for obtained.

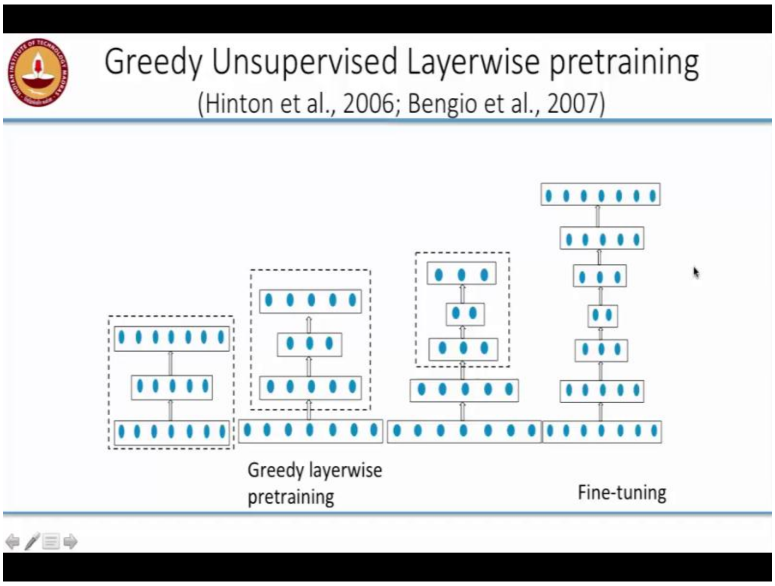
(Refer Slide Time: 02:46)



Further there was this problem called the vanishing gradient problem. So, people quickly figure out that if we have many layers in the neural network; right. It was easier for the network to represent more complex functions; even though 2 layer network or a 3 layer network and depending on the how we contact. The standard 3 layer network was the universal approximator, having more layers allowed more complex representation to be build. But see, number of layers in the neural network increases, the gradient that we compute by doing that propagation are becomes vanishingly small; right. And since there are lot enough feedback for the early layers in the network. The rates in the lower layer remains random wherever universalize them ; right. And so, you are not able to learn any deep networks; right. So, the networks, neural networks it will learn necessarily had to be shallow; because only a few 3 layers could be, 3 or 4 layers could be trained meaningfully using back propagation. So, people had come up with different tricks for training deeper networks.

So, and this discovery of these tricks for training deeper networks, is essentially what is revive the interest in the field. And now we can build networks at have 8 layers, 10 layers and so on so for. And this produce some fantastic performances in problems that we are talk to be very hard to solve for Ain machine learning. And so, there has been reviewed interested looking at neural networks. So, I will talk about one specific mechanism for training multiple layers in the network, and then you can I mean if you are interested in this follow it up with other material.

(Refer Slide Time: 04:32)



So, this is proposed in met 2000 by Hinton 2006 and independently by Ashwa, Bengio and others. In 2007 it is more like a greedy unsupervised layer wise pre training, so what we mean by this? So, you start off with very simple network architecture called an auto encoder. So, you have an input layer and you want to produce the same input at the output. So, the input layer and the output layer need to be identical, but in between the connections will go through a smaller hidden layer of neurons. So, the idea here is that the smaller layer is going to learned some kind of a encoding or some kind of a reduce representation of the input; that is sufficient for you to produce the output that you are looking for.

So, you could train this using back prop or other slightly more advanced gradient techniques. So, if you think about it, that is real supervision that is required here; because as soon as have the input; right. The output is just the same input. So, I am going to set it through a smaller hidden layer. So, that I am learning a reduce representation of the input.

So, once I have this 1st level of reduce representation; right, I am going to it relatively deepen the network. So, what I do? So, once I have the 1st hidden layer of representation, the first hidden layer of representation I am going to add another auto encoder network on top of it. So, this network takes the hidden layer representation for the input and tries to produce the same hidden layer representation at the output, but in the middle layer is going to have pure neurons. So, in the effect I am taking this larger input and reducing it to a smaller hidden representation here. And I can repeat this; right.

So, once I have train this, I remove the outer layer; and then take on another layer of auto encoders.

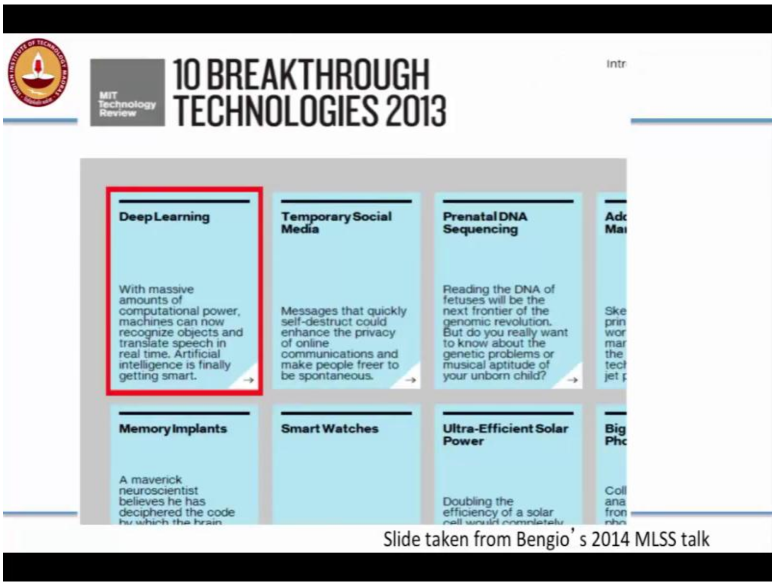
So now, you can see that my training has gone several layers deep. So, in stuff just looking at one layer deep auto encoder which have I was able to train using back prop efficiently. So, I am essentially using the same construction again and again. So, at any point the training happens only on a one layer auto encoder, but then because of this iterative deepening. Now I have actually taken this input representation and progressively reduce it to a much smaller representation at the hidden layer. So, this kind of an observation that you can use this layer wise pre training of the data, led to resurgence in lot of deep architectures. So, why do we call this pre training? So, far I am not talked about any kind of classification task that you have performing. we are just trying to find features in the input space. So, once I have found these features in the input space then I can take this, and I can tack on top of it.

On top of this features that I have learned I can tack on other neural network. And now I have my full fledged deep auto encoder and then this hidden layer representation can now be used as an input any learning task. So, so this part is call the fine tuning part and the layer wise training part is called the pre training part; where I find the hidden representation. And after that I can added to any complex neural network, that can do my classification task or my regression task whatever it is that I am looking at. So, this kind of layer wise pre training allowed people to, drive more complex comprise representations. That very useful in a variety of problem solving and I use.

(Refer Slide Time: 08:25)

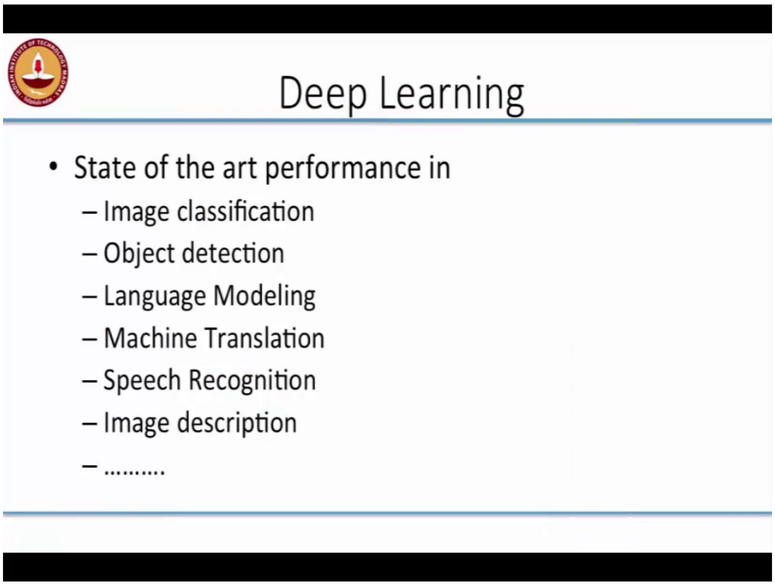
So, this again as review lot of high in neural networks or in deep networks as they are called. So you can see very human similarities to the news items that I showed you from the 1959 thing. So, a stimulated brain and you can teach context to computers, machines at learn without humans, etcetera, etcetera. So, skate what you read in news paper with a pinch of solve, but again there is significant renewed interest in deep networks and neural networks as there was in the 50s.

(Refer Slide Time: 09:04)



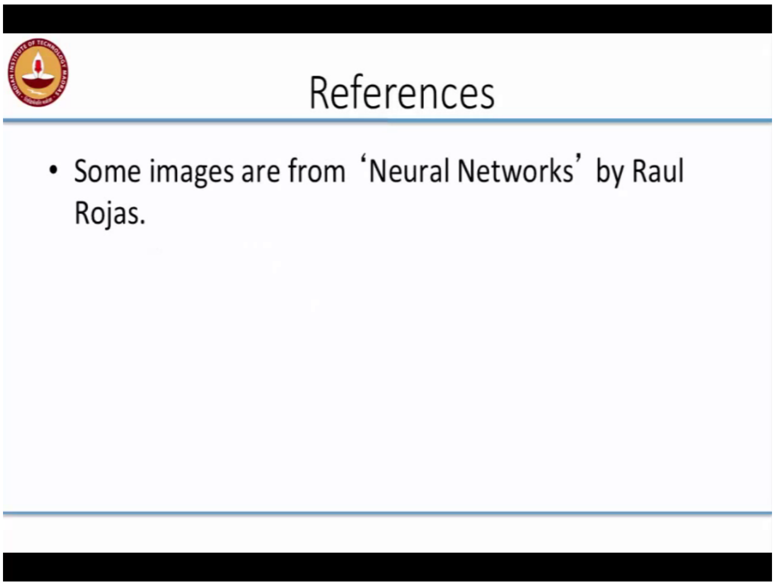
But the overall feeling in the community is that, deep learning is come to stay; because there are lot of nice properties about this generation of neural networks as compare to the earlier generations. So, the things of more stable and things of reproducible; even though significant computing power is needed. So, most of the successful applications of deep learning that you see, would have had significant amounts of computational power. But then the computational power is also cheap; and you are able to solve really complex problems using neural networks.

(Refer Slide Time: 09:41)



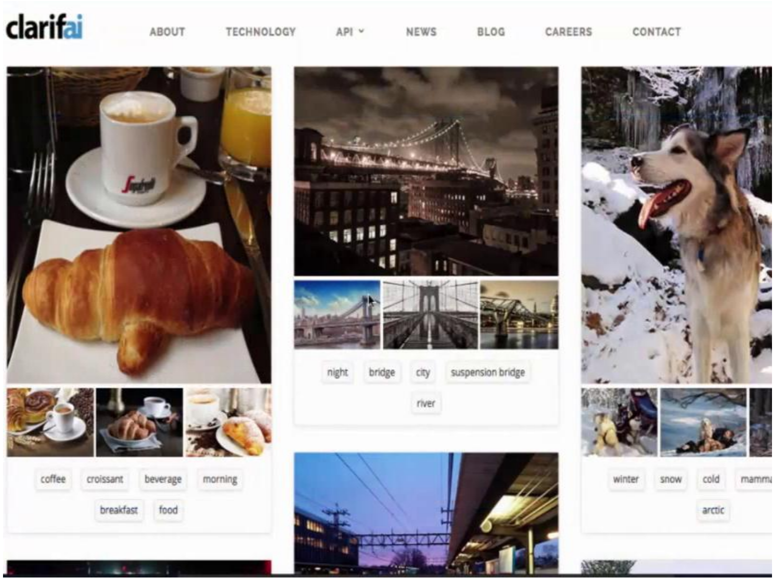
And sp, deep learning collects have yield at state of the art performance in the variety of domains, like image classification, object detection, language modeling, machine translation, speech recognition, image description. These are problem at traditionally considered very hard for machine learning algorithms and deep learning seems to have significant in pattern these areas.

(Refer Slide Time: 10:08)



So, some of the images I used here are taken from the neural networks book by Rahul Rojas.

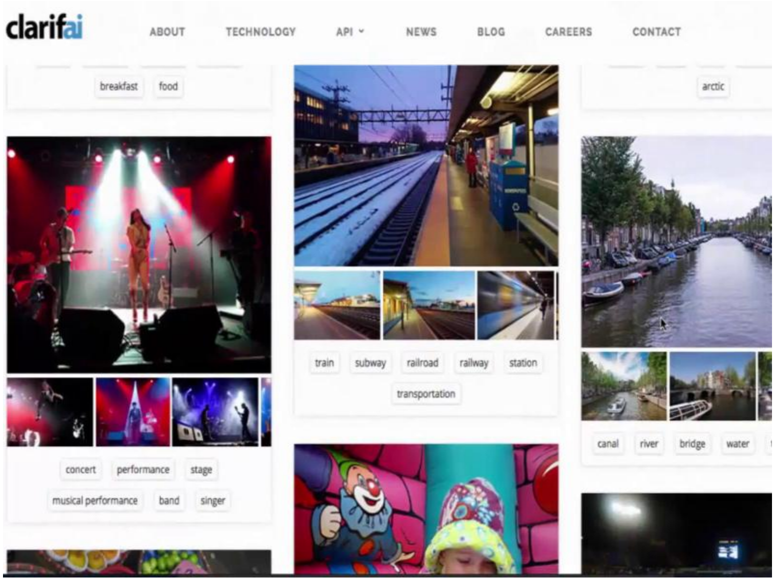
(Refer Slide Time: 10:15)



And so, I like to show you few demos of deep networks and action law. So, here is one algorithm from company call clarify. So, given an image with different kinds of textual labels it says that, here given this image is say that is coffee, at there is croissant, there is a beverage in it, and this is probably breakfast and it is had in the morning and overall hey this looks like food. So, it is able to derive all of this tax just by looking at this image. These are all similar images it does manage to retrieve images, similar to these. You can see that all of these are images of breakfast or continental breakfast and all of these have a cup of coffee in there. And so, even though the variety of this is able to cover is truly expounding.

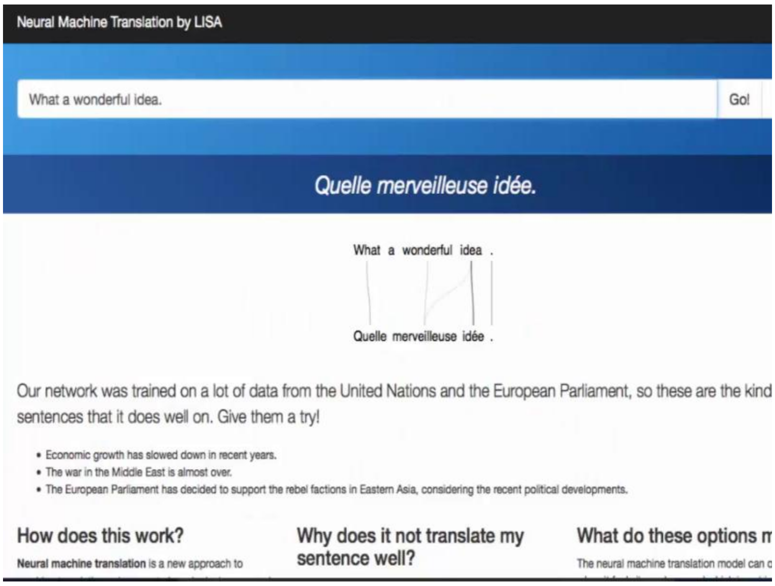
So, like wise look at these picture here, is able to figure out the here suspension bridge here and there is a river here; even though the river is not explicitly visible. And that is it night because it is lighted up and it looks like bridge in the middle of the city. And likewise it has found similar images, which are all suspense bridges in cities on top of rivers.

(Refer Slide Time: 11:40)



And. So, likewise it is able to work on variety of outdoor seems, as well as indoor seems and its able to perform really well. And this is state of the art in terms of image understanding and labeling at correctly.

(Refer Slide Time: 11:51)



And likewise you can look at how well it works in machine translation. So, the both the image tagging and the machine translation tasks use some much more complex neural network architecture that we have seen so far. But then so, here it’s a simple example from here. So, I typed in the sentence what a wonderful idea in English and then it gives me the equivalent in French. And not only does it do that, it tells me which words correspond to which word in re translated language. Which words in the original English language, correspond to which word of the translated language.

So, I do not know French. So, I am not sure that is a good translation, so let us not do this. So, let us stop with the previous… So, it is then this thing actually you can learn translate, can do translations between multiple language is not just in English and French. It first trained on the data from United Nations and European parliament. So, it can do translation between the all the European languages. So, that brings us to the end of this module on deep learning.

Thank you.