Applied Natural Language Processing
Prof. Ramaseshan Ramachandran
Department of Computer Science and Engineering
Chennai Mathematical Institute

Lecture - 51
Recap and Introduction

Hi everyone, today going to start a new session on recurrent neural networks. So, this is an advanced concept and in the neural network as well as in natural language processing. As a human being right, so we will always keep on improving things that we have done earlier. So, in the same fashion, there is a certain application where you cannot use the standard mechanisms that we had studied earlier.

So, we want to see if we can apply a new mechanism to solve a certain problem that we are going to be talking about a little late. So, this is a very interesting topic that has come up in recent years; when we try to improve certain things we make it a little more complex right.

So, this one is a bit complex than what we had studied earlier. The mechanisms of learning are going to be a lot more difficult than what we have seen earlier. The way we construct the neural network is very different from what we have done earlier so on ok.

So, we pay attention to this lecture also understand a lot of relevant material that is out there on the internet alright. You definitely have to understand this particular topic; if you want to develop applications based on the RNN we call it alright. So, before that I will take a little step back and then see what we did so far and then get on to the business of understanding a neural network.

So, right from the beginning right, we started with the pre-processing we started with dividing the corpus into words, tokens or sentences, paragraphs, and so on depending on what we are looking at. And then we wanted to find out whether it's possible to get the meaning out of the word, given the context of words surrounding it.

So, we use some methods using LSI and then try to find out the word embeddings which is not just a representative or an index of the word; but it contains some additional information related to the word in terms of its context surrounding words similar words

and so on. So, in this context this similar is not equal to a synonym; it is equal to the word that could have some neighborhood relationship alright.

So, for example, if you say automobile workshop could be a neighborhood word for the automobile. So, the word vector could have some representation of the dealership or workshop along with the automobile, that is what we meant by word embedding.

So, we are able to get something out of that from the corpus through LSI and we got a set of word vectors for the vocabulary that we are looking at in a given corpus right. And then the accuracy in terms of the prediction increases alright. And then we wanted to find out if these learnings can be made possible through another mechanism through neural networks.

So, we started looking at the perceptron and then started looking at how we can classify words and then how we can really find out a sentiment of a word you know by taking a small example. And then try to construct a multilayer perceptrons. And then later converted the multilayer perceptron into a model where we used a training mechanism to train the network so that the network of the network achieves whatever that we are looking at ok.

And the parameters which we are looking at are the weights as part of the network. And then we also saw that its possible to really construct a neural network using a different mechanism we saw CBOW model, we saw Skip-gram model. And we were able to really construct the word embeddings using those models alright.

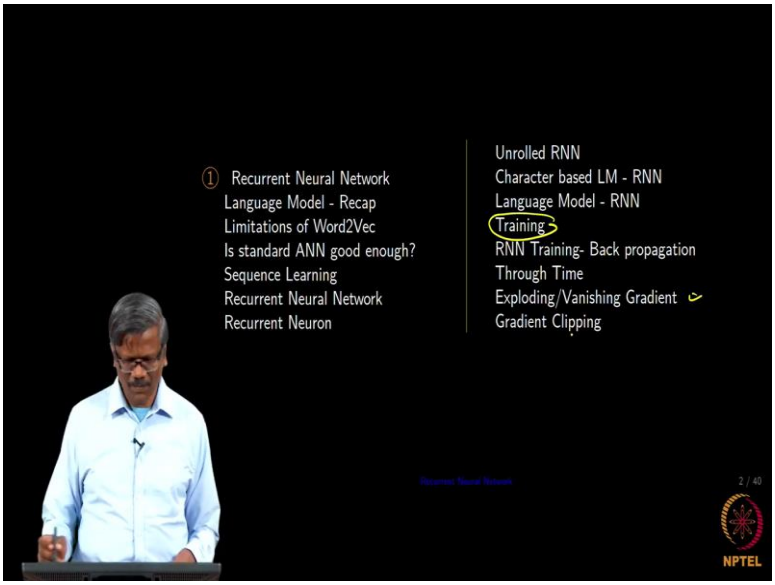
So, now, we want to see whether that can be extended so, that, the next stage in the natural language processing is about understanding a sentence. You know we saw how a word can be understood and the neighborhood relationship can be captured.

And now the next stage is after predicting the next word we should be able to understand the sentence. And then, we should be able to understand the paragraph, we should be able to understand the document and then paraphrase it or summarize that and then later take a paragraph from one language translate it to the other and so on so forth.

So these are all the tasks that are still pending, in the neural network as well as the class is concerned. So, we will see whether we can move to the next level of understanding a

sentence through a recurrent neural network. So, which was not possible with the standard model of the neural network ok.

(Refer Slide Time: 05:49)



So, this is what I am going to be covering in this session ok. So, I am going to be talking about the recurrent neural network. Giving a brief overview of what we have done with respect to the language model and then we will talk about the Word2Vec model and then see what are the limitations.

And then for the next stage of taking the understanding is standard ANN good enough; if not what should we do. And well talk about sequence learning. So, what is sequence learning? All about and then we take a very small example of a recurrent neuron and then explain that. And then try to talk about what is the meaning of unrolling and RNN ok. So, in the recurrent neural network we are going to be talking about the time series ok.

While we are in the other neural network that we spoke about there's no concept of time they are all static networks. So, in this case there is a time involved in this so, we want to see how words can be input into this network and then how the RNN can be unrolled in the time sequence. We will also look at the character-based language model, we have not done that earlier, but we will look at here. And we look at the standard language model through RNN.

And then the most important part is training contains a lot of interesting mass, we will do this towards the end of the session we will not do it in the middle. We will look at the training especially through backpropagation through time.

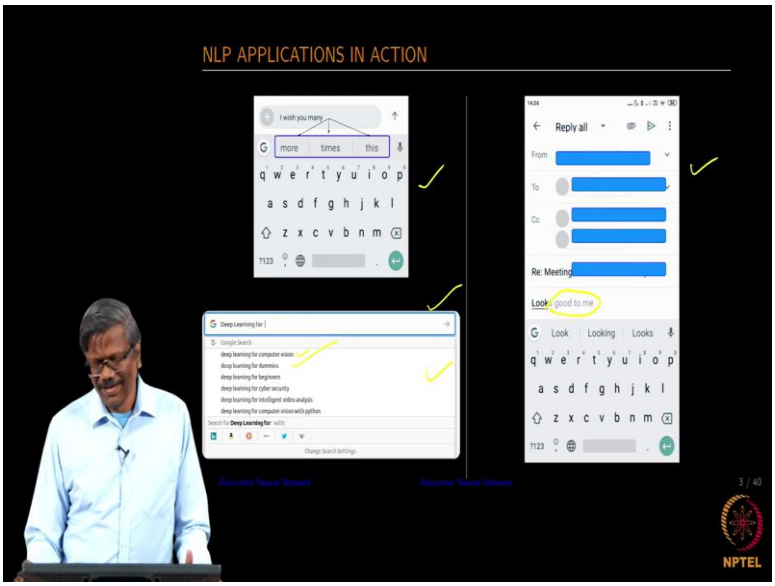
So, when you look at the RNN it is nothing more than a collection of backpropagation network ok. And then how do you really backpropagate through time? Since time is involved. So, we need to backpropagate the error through time so that we are able to adjust all the weights or to make the neural network or take the neural network to an equilibrium state.

What happens when you have long time series? So, we will talk about exploding an vanishing gradient what does it mean? When you have a neural network that is unrolled in time so, we will have several layers right. So, we were talking about taking a sentence as part of the input into neural network right the sentence can be of any size it is only limited by the time. So, when you have a sentence of about 20 words or 30 words or you want to process a paragraph of about 200 words.

I should be able to input all the words, one after the other through the networks so that it learns and then does whatever we are asking it to do right. When we do the backpropagation through time and when you come from the topmost layer to the bottom-most layer, the gradients start diminishing in size or it starts becoming bigger so that is when the neural network either explodes or the gradient vanishes. And then there is a mechanism we will just about how we can really clip it so that we can avoid the problem of explosion and vanishing part of the gradient.

So this is going to be the structure of this lecture. I will talk about this training once we finish the gradient clipping part and so on alright. Until then we assume that we are going to be training the neural network we are going to assume certain mathematical equations as true and move forward alright.

(Refer Slide Time: 09:55)



This is something that you should be very familiar with the right, the first one that you see on the screen is a keyboard that predicts what is going to be the next word. So, what does it do? So, when you start typing the words or when you start typing the part of the word it starts showing you some options for you right.

So, you can pick even before you complete your word or a sentence and it saves a lot of time for you right here is a prediction involved in this. And when you want to predict what happens? Behind the scenes; obviously, there is a natural language processing that is running this application correctly.

So, in this case, you know if you look at this, I wish you many I am sure many of you would not have typed the entire sentence of I wish you many more happy returns of the day, you just keep typing few characters and the prediction appears you just touch them so the entire sentence is completely correct? So, you just go and then play with this and then see how far it takes you? And beyond a point this particular small NLP application will start repeating or start throwing some random words for you ok. So, it will never let you complete it sometimes.

So, but that is wonderful; even for us on a small device we are able to do this and there is a small NLP application behind that which is predicting the word that you are going to be typing. The second one again when you start searching for a particular topic in the Google search box, you see lots of options popping down correct.

So, in this case when I type deep learning for, so you can see the first one is coming as deep learning for computer vision. So, it looks like lots of folks, in the vicinity where I live search for computer vision-related topics and deep learning.

Then interesting deep learning for dummy appears as the second option for you. So, again how do they come up? So, based on what you type? As you type deep and you will see a lot of things popping up right in the drop-down box. And when you type deep learning again the options are getting reduced; and you see certain options being populated.

And then when you type for again you see lots of things when you start typing computer you will see certain more options getting restricted, but you will see still see about 10 or 15 options or depending on how it is configured right.

So, again there is a natural language processing behind this scene correct. And then the third one is very interesting even though I do not like it, this application provides you an option to complete the sentence not just the words right. So, in this case, you know I got a meeting request and then we had done some work and then some attachment was sent to me and then I am just responding to this mail thinking that I am going to be typing this attachment that this person had looks good maybe we can make few more modifications and so on so forth ok.

When I start typing ok, I can see that there is a sentence that is being completed as part of the application right. So, this is pretty dangerous; that means, it is really reading what I am doing in one way, but if you look at it from the technology perspective as a technologist I really love this right. So, I am able to really complete the sentence without really typing the whole thing; these kinds of filling the data for you is really cool.

So, this is another one where the NLP is behind the scenes right. So, for all these right if you look at it, the character that I am typing you know when you are going to be typing the word here, the word size is not limited; it could be a 4 letter word, 5 letter word or 2 letters or 6 letter word and so on right.

So, as and when you type it keeps looking at the characters that you have typed, and then based on what you have typed that is your context and it tries to find out the equivalent word, that is available in the dictionary and then fills it correct. So, the size of the input is

not limited here not just 4 words. So, it does not do it just for 4 words it does for 2 words or other 2 characters 4 character words 5 character words and so on so forthright.

So, the input size is not limited, even in this case the input size is not limited it is not limited to 3 words or 4 words. You start typing the character it starts appearing, it gives you the right word for you, and then when you complete the first word and then start typing the second word it starts completing your sentence correctly.

So that means, the input size is variable here too even in this case the input size is variable. So, can the application that we have developed so far or the techniques that we have looked so far using the neural network solve this problem, it is what we are going to be looking at alright.