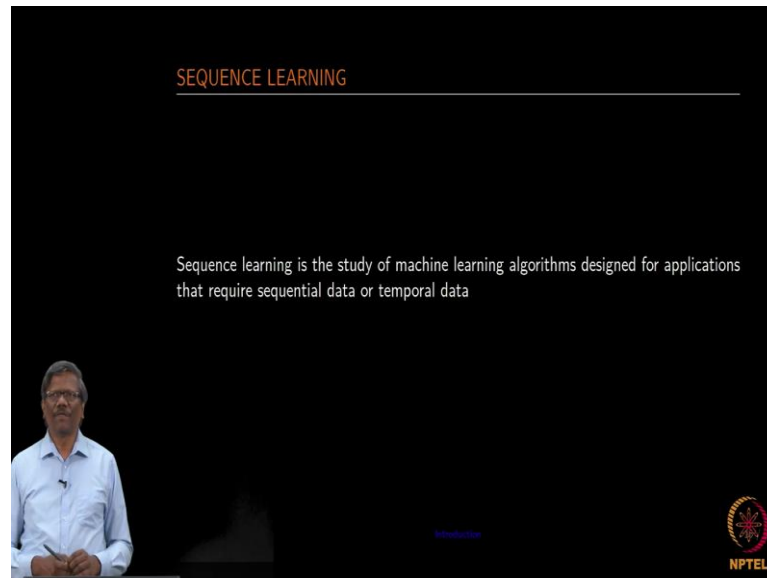**Applied Natural Language Processing**
Prof. Ramaseshan Ramachandran
Department of Computer Science and Engineering
Chennai Mathematical Institute, Madras

Lecture – 05
Sequence Learning

(Refer Slide Time: 00:15)

now we spoke about at a very high level, again I am repeating the counting of the words in the corpus, and then we did some prediction based on what we counted. And then we started learning more about the words in terms of the semantic connectivity or the contextual information using the vector space models as well as a machine-learned model. once the words are understood in certain fashion the next is about understanding a sentence right. it is all about that right. For starting from the number, now some meaning about the words, can I look at this sentence as a whole to understand that, that is where the sequence learning comes into the picture in the natural language processing?

we start doing sequence learning when we have an understanding of the words in certain fashion right. we can define sequence learning as a study of a machine learning algorithm designed for an application that requires sequential data or temporal data ok. that means, I am going to be expanding my horizon in terms of learning what the sentence is all about. starting from the numbers, the words now the sentence right.

see the progression that we are making, so that where we are going to be bringing in the new models. the old models are good for word understanding and so on, but they are not good enough if you want to really represent a word as the sentence as input. And then how do we do this is where the recurrent neural network comes into the picture.

this is one of the key problems. Sequential data prediction is considered as a key problem in machine learning and artificial intelligence. if you are able to understand a sentence, that means, you are gaining more and more knowledge about that language and the words that are there in the sentence right. And language understanding is always about intelligence. we are trying to bring in more and more intelligence into the system by starting to look at the count, starting to look at the prediction part, and then trying to find out how words could be constructed. And then now try to use the knowledge that we have gained and then see if we can really understand the sentence which is nothing but the string of words right.

in the case of the natural language processing which is nothing but this string of text, a combination of sentences, a combination of paragraphs and so on. We read this sequentially to understand the content, it is very important, is not it? we cannot just throw away certain words, and then if your words are not in order it is very difficult to understand that. Unlike the image where we need to look at you know portions and then

we are still able to make it out, whereas in the case of text for only from one word, and then if you ask to construct the entire paragraph, it is going to be very difficult right.

 if you are given some portions of the image, you may be able to extrapolate that. even in the text it is possible, but I am not sure how close we can get to if you are only given certain words to reconstruct the paragraph ok. it is the task that requires a sequential understanding of the content. what work we will be studying in this recurrent neural network, we will talk about the architecture of the recurrent neural network, we talk about the language model, is it possible for me to really predict the next word or say that our whether a sentence is legally right or wrong, I mean in terms of the syntax ok. And then how do I encode a sentence into a fixed exercised vector, given a long sentence of 50 words, can I compress that into a fixed exercised vector? Like we had done in the embedding of words, can I create an embedded or a dense vector for a sentence.

And then we talk about what happens when we try to learn long sentences because there are some inherent problems in the neural network that we would like to address and bring in new models using LSTM and GRU, it is Long Short Term Memory and then Gated Recurrent Unit model. To make sure that we do not get into the trouble of exploding and vanishing gradient, these are all the model constructions and that we will also take a look at those ok. I will be covering in detail all of these in the subsequent week.