Sequence Learning Problem

So, so far we have dealt with two types of networks one is feedforward neural networks and the other is convolutional neural networks and both these networks the input was always of a fixed size.

For example, we fed fixed size 32*32 images to convolutional neural networks for image classification. All your training images, all your test images were always scaled or cropped to this particular size.

Further, each input to the network was independent to the previous or future inputs. The computations, outputs, and decisions for 2 successive images are completely independent of each other.

But in many applications the input is not of a fixed size. So, and also successive inputs may not be independent of each other. So, let us understand this with the example of auto completion that all of us are used to while typing SMSs or Whatsapp or other things.

First, successive inputs are no longer independent. (while predicting “e” you would want to know what the previous input was in addition to the current input.)

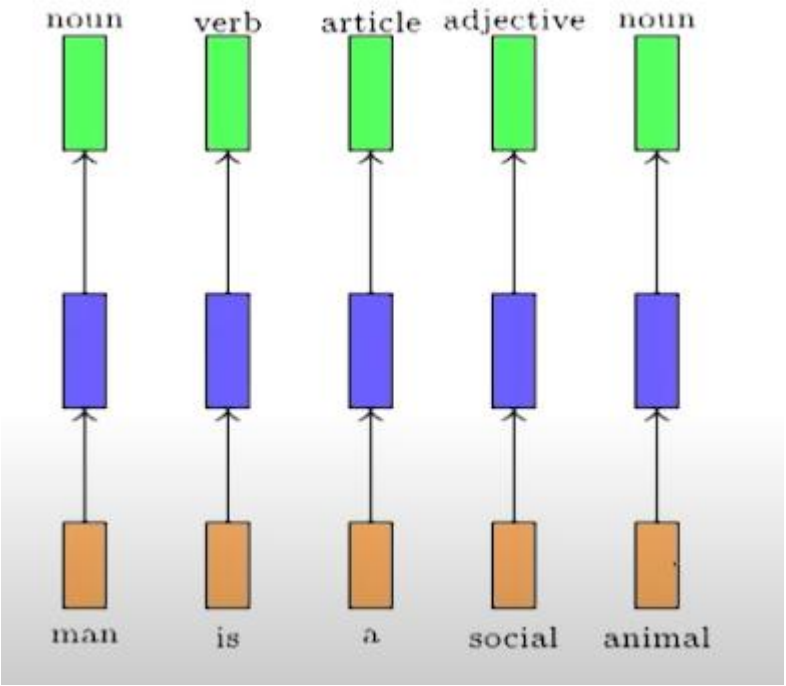
Second, the length of inputs and the number of predictions you need to make is not fixed. (For example, “learn”, “deep”, “machine” have different number of characters.)

These problems are known as sequence learning problems. We need to look at sequence of inputs and produce an ouput/outputs. Each input corresponds to one time step.

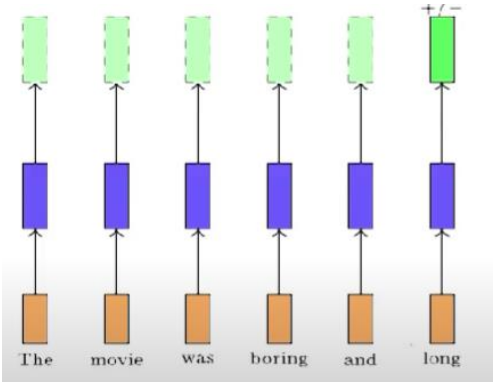
More Examples:

1. Predicting the part of speech tag
    a. Consider the task of speech tag (noun, adverb, adjective) of each word in a sentence
    b. Once we see adjective (eg. social) we are almost sure that the next word should be noun (man)
    c. So, the current output not only depends on the current input, it is also actually depends on the previous input.

d. Furthe the size of input is not fixed.



2. Output not expected at every stage
    a. Sometimes we may not be interested in producing an output at every stage.
    b. Instead we would look at the full sequence and then produce the output.
    c. For example, consider the task of predicting the polarity of a movie review.
    d. One cannot predict about movie by just looking at the last input which is long in this example.



3. Video sequence:
    a. Sequence could be of any thing (not just words)
    b. A video could be treated as a sequence of images.
    c. You have to see entire sequences and not just intermediate sequences.