

**Experiment No 8**

Design and develop a model for identifying sentiment of twitter dataset.

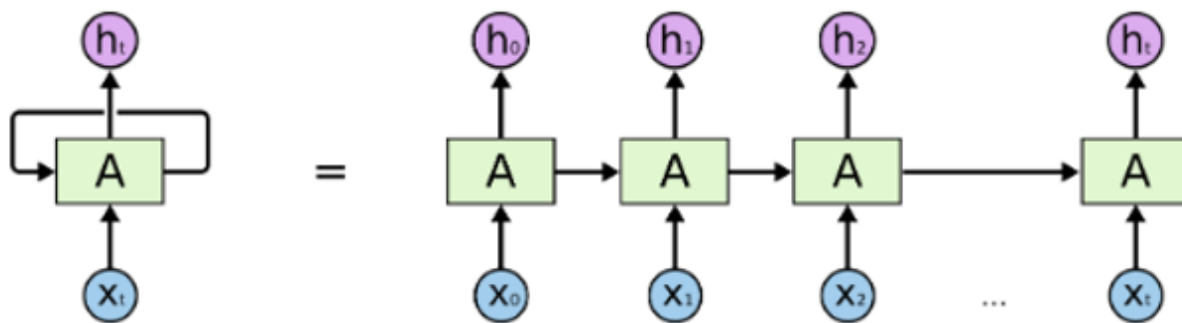
**Objective:**

At the end of this practical session, student will be able to classify reviews in different sentiment classes using sequential model.

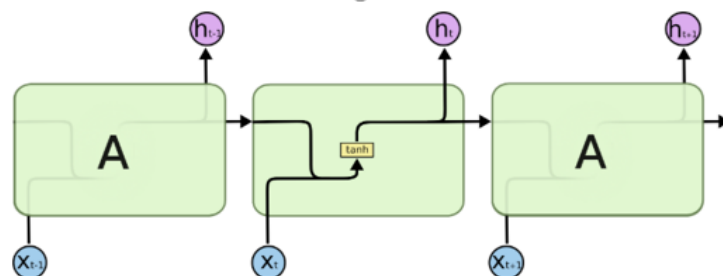
**Theory:**

Text data is the sequential data and can be understood as either a sequence of characters or a sequence of words. Recurrent neural networks or RNNs are a family of neural networks used for processing sequential data. It is specialized for processing a sequence of values  $x_1, x_2, \dots, x_t$  at timestamp  $1, 2, \dots, t$ . For sequential data processing, it shares parameters across different parts of a model. Parameter sharing makes it possible to extend and apply the model to examples of different forms (different lengths) and generalize across them.

It processes sequences by iterating through the sequence elements and maintaining a state containing information relative to what it has seen so far as shown in following figure.



The repeating module have following structure



Deep learning models do not take input as raw text. They only work with numeric tensors. So text need to be converted into vector. Vectorizing text is the process of transforming text into numeric tensors. The different units into which you can break down text (words, characters, or n-grams) is called as Token. Tokenization process develop tokens

---

from text. All text-vectorization processes consist of applying some tokenization scheme and then associating numeric vectors with the generated tokens. There are two ways to generate vector for token namely one hot encoding and token embedding. One hot encoding is simplest and easiest way to develop vector for word. But it suffers from different problem like sparseness and high dimensions. Token embedding can be obtained for word using two different method. First is learning word embedding jointly with your main task. In this method, start with random word vectors and then learn word vectors in the same way you learn the weights of a neural network. In keras, it is done with embedding layer. Second method is using pre-trained word embedding. In this method, you have to load your model with word embeddings, which were precomputed using different machine learning tasks.

Sentiment analysis consists of determining whether a comment expresses positive or negative sentiment. IDBM is benchmark database for sentiment analysis. It is database of movie reviews. Sentiment analysis model takes these reviews and classify these in positive or negative class based on sentiment.

**Keyword:**

Sequential Model, Sentiment Analysis

**Procedure:**

1. Load twitter Dataset
2. Preprocess the input data
3. Define your model with embedding layer and simple RNN
4. Train and evaluate your model
5. Plot the result