***Text Generation Using LSTM***

*-Soham Banerjee*

**Pre-processing of Data –**

The dataset used was the corpora – Human Action by Ludwig Von Mises which was available in the ‘.epub’ format.

First this was converted to .txt format for ease of use.

The size of the file was quite big to train the LSTM quickly. So, the data was manually reduced to a lesser size (<500 KB)

The reduced file was manually scanned to remove unwanted elements like references, headings, etc. This was done to avoid unnecessary training requirements.

**Model –**

The predictions done are based on characters and not words. The word-based approach can lead to better syntax and correct words but requires longer sequences for training and more data to cover for the increased size of the dictionary of predicted outputs. Breaking of sequences is proper order is required and the text needs to cleaned up much more before training. Hence, the choice.

The model used consisted of 2 LSTM layers with 512 cells in each layer, followed by the fully connected layer for the categories which uses a soft-max activation.

**Training –**

The data was converted in segments of size 100 chars each. Total number of samples obtained were ~180000. The samples of were passed in batches of 128 and trained using an ‘adam’ optimizer. Training was done over 15 epochs and a cross-entropy loss of 1.2006 was obtained after the epochs.

**Observations –**

A specific sequence of chars – “of the cctinns of the nonhistorical sciences and the meaning” tends to follow the occurance of the sequence – “of the”.

The model predicts words which are quite close to the real language. This can further be improved by using much more epochs for training and by using an increased corpora size.

**Predicted Sequences –**

1. Input –

behavior and unconscious but nonetheless serviceable responses of human muscles and nerves. yet, th

Output -

e concepts of the philosophers and the sroduction of the consection of the cctigns of the mone sinles between eonsciousness and the means they applied by the content of human action and anlther abtion as the suoperarian leneers of the cctinns of the

1. Input –

ence to experience does not impair the aprioristic character of praxeology and economics. experience

Output -

concerning the content of human action and anster that the problem of the eouire etent of the cctinns of the nonhistorical sciences and the product of the praxeological mathrial conditions of the

1. Input –

r that to prefer a to b could at the same time be to prefer b to a is simply inconceivable and absur

Output –

d reason and the means they applied by the content of human action and another abtion as the suoperarian leneers of the cctinns of the nonhistorical sciences and the meaning of the cctinns of the nonhistorical sciences and the meaning of the

1. Input –

sition: man's unique aim is to attain happiness, is tautological. it does not imply any statement ab

Output –

out the praxeological categories and therefore prcseiued and the meaning of the cctinns of the nonhistorical sciences and the meaning of the cctinns of the nonhistorical sciences

1. Input –

of inference which has given pragmatic evidence of its expediency, although its satisfactory epistem

Output –

ology can be confined only by the employment of the cctinns of the nonhistorical sciences and the product of the praxeological mathrial conditions of the cctinns of the nonhistorical sciences and the sroduct of the