National Institute of Technology Karnataka, Surathkal
Department of Computer Science and Engineering

Mini project proposal submission (Computer Graphics)
VI Sem. BTech CSE (January - April 2019)

Roll No. in the class	Inst. Reg. No.	Name	Email ID	Phone No.	CGPA and Role
16CO124	166099	Mehnaz Yunus	mehnaz138@gmail.com	9207431140	9.11 Leader
16CO140	166553	Sharanya Kamath	16co140.sharanya@nitk.edu.in	9686191477	8.84 Member

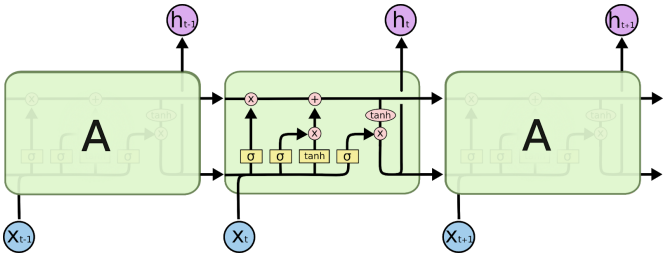
Date: 06.02.19

Title: DNA Sequence Classification using LSTM Networks and GRU

Abstract:

Gene classification is the problem of identifying the functionality of genes using only the sequence information (ATGTGT...) automatically.
Recurrent Neural Networks are generally used for processing sequences of data which evolves along the time axis. But the main challenge behind the problem remains the feature selection process as sequences do not have explicit features. We plan to use Long Short Term Memory networks for classifying dna sequences into various gene sequence function groups.

About LSTM:



Traditional neural networks require starting learning from scratch. RNNs address this issue. They are networks with loops in them, allowing information to persist. LSTMs are a special kind of RNNs explicitly designed to avoid long-term dependency problems. These networks can remember information for long periods of time. Because of this unique property of LSTM networks, we think they would be appropriate for our problem statement.

Methods already used:
Artificial Neural Networks, Deep learning using CNN, RNN, SVM with k-means clustering.

Inputs:
16S dataset downloaded from the RDP Ribosomal Database Project II (RDP-II) by NCBI with a total amount of 3000 sequences.

Outputs:
Accuracy of classification into gene sequence function groups.
Comparison with other classification models.

References:

- Jesse M. Zhang and Govinda M. Kamath Stanford University Learning the Language of the Genome using RNNs
- Daniel Quang and Xiaohui Xie. Danq: a hybrid convolutional and recurrent deep neural network for quantifying the function of dna sequences. bioRxiv, page 032821, 2015.
- Qingda Zhou, Qingshan Jiang, Dan Wei, "A new method for classification in DNA sequence", Computer Science & Education (ICCSE) 2011 6th International Conference on, pp. 218-221, 2011.

Performance measurement criteria (Milestones and dates):

- Dataset understanding and partial Implementation - 3rd week of February 2019.
- Complete implementation - 1st week of March. 2019.
- Paper submission to mentors - 4th week of March 2019.
- Paper publishing or submission - 4th week of April 2019