Pitch insensitive LSM - Convolution in f-space

Group-14

Members:
Aman Shaikh - 19D070051
Yogesh Katara- 19D070072
Harsh Choudhary- 200070023

Guide: Vivek Saraswat
Prof. Udayan Ganguly

EE746: Neuromorphic Engineering

Introduction

Speech recognition is a broad study area in computer science that recognizes spoken words and converts them to text

Liquid State Machines (LSMs) are brain-inspired architecture, consisting of a large recurrent network of randomly connected spiking neurons

It has various design parameters giving high flexibility for training

Introduction

Preprocessing: This stage consists of a cascade of a second order filters and follows BSA algorithm to get the input spikes trains. 77 spike trains were generated for a corresponding input speech sample

Liquid Reservoir: Grid of LIF neurons with fraction of excitatory and inhibitory neurons

Linear Classifier: A fully connected layer of spiking readout neurons to recognize the class of the input

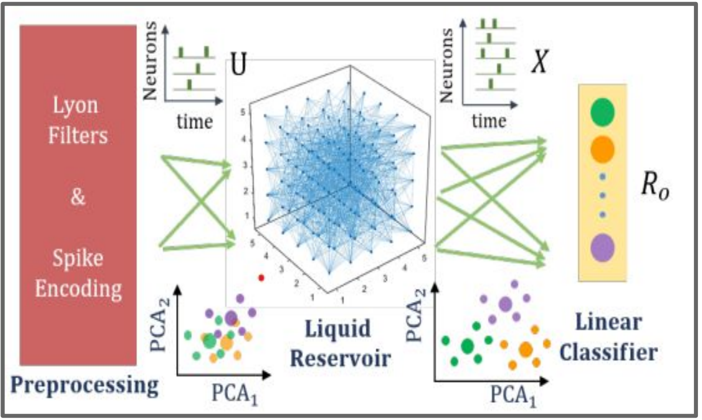

Fig. 1. Implementation of LSM

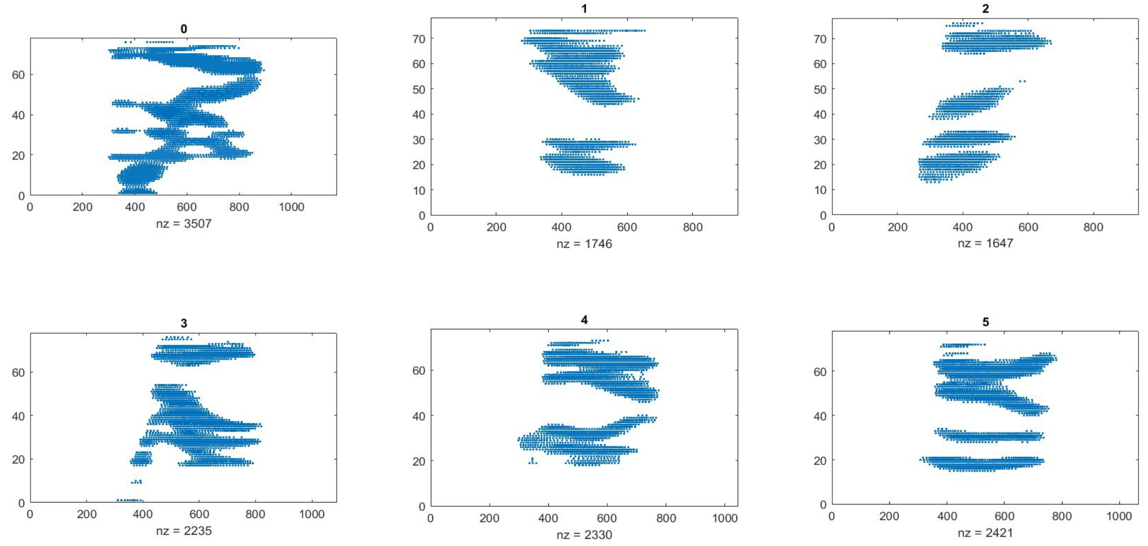

Fig. 2. Raster plots of spoken digits (0-5) for a specific speaker.

Background

- In [1], the authors had perform spoken digit recognition on 500 samples from TI-46 dataset.
- The samples consisted of 5 female speakers, each having 100 samples where each digit 0-9 had 10 samples each.
- Reservoir size used was **5x5x5**.
- 5 fold testing and training was used for 20 epochs.
- The accuracy was obtained to be **99.09%.**

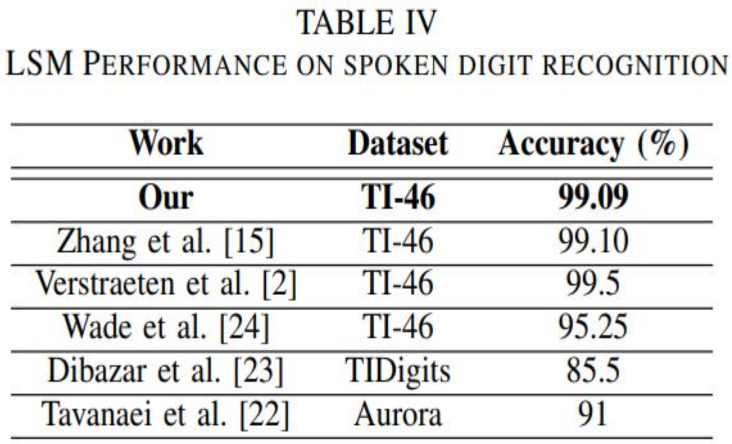

TABLE IV
LSM PERFORMANCE ON SPOKEN DIGIT RECOGNITION

Work	Dataset	Accuracy (%)
Our	**TI-46**	**99.09**
Zhang et al. [15]	TI-46	99.10
Verstraeten et al. [2]	TI-46	99.5
Wade et al. [24]	TI-46	95.25
Dibazar et al. [23]	TIDigits	85.5
Tavanaei et al. [22]	Aurora	91

Table I
Performance comparison [1]

Replication

- The dataset and the codes from [1] obtained were run and the accuracy obtained was **98%**.
- The accuracy is quite high as all female speakers have high pitch.
- Raster plots obtained are quite similar.
- Different digits can be distinguished from their raster plots

Motivation & Methodology

- As a next step towards modelling LSMs as pitch insensitive, the dataset used contains 4 male and 4 female speakers.
- We used 800 samples from TI-46 dataset which consist of 8 speakers, each having 100 samples of digit 0-9 have 10 sample each
- This samples fed to the preprocessing unit to generate the training inputs for LSMs.
- Reservoir size is same as used in paper and 5 fold training and testing is used.
- The accuracy obtained is **95.00%.**

Train on one pitch and Test on another pitch

- Training the LSM on a particular pitch/speaker does not do well when testing on a different speaker.
- Test accuracy is 0% for any random case.
- As the model has never seen another pitch and both the signals are uncorrelated, it will be rarely able to detect it and since the no of data points are not very large , the accuracy comes out to be 0
- The raster plots look different for different pitches.

Train on mixed pitches and Test on another pitch

- After training the model on mixed pitches, the test accuracy on the untrained pitch is 0%.

Effect of Liquid Reservoir size

The accuracy corresponding to the size of the reservoir is given in the table

The accuracy increased with reservoir size.

Computation time increased due to large matrices.

Reservoir Size	Accuracy(%)
5x5x5	95
5x5x8	96.25
8x5x5	96.75
10x5x5	97.5

Table II
Reservoir size vs Accuracy

Effect of filter size used in BSA

- Ben Spiking Algorithm (BSA) is used to encode the input voice signals into spikes.
- Increasing the filter size in BSA, the accuracy increased slightly.
- Increasing filter size, increased spiking points in raster plots
- This provides more data points for learning of LSM.

Filter Size	Accuracy (%)
97	97.5
193	97.625
577	97.875

Table III
Filter size vs Accuracy

Effect of filter size used in BSA

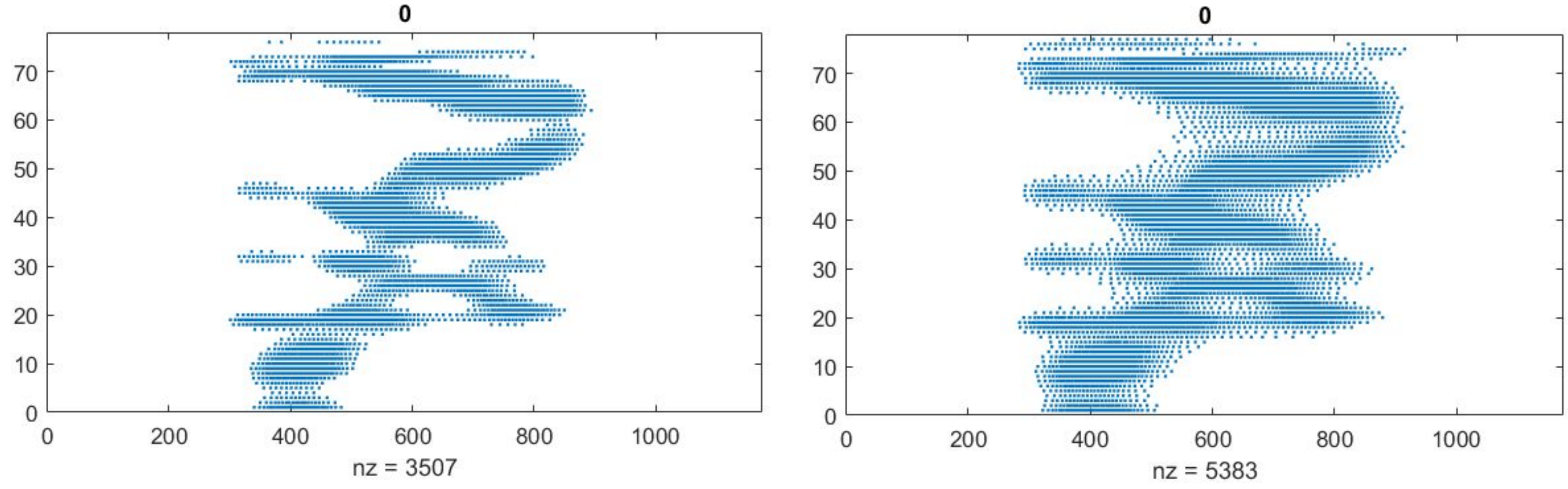

Fig. 3. Output of BSA for same input signal with filter size of 97 and 193 respectively

Results and Conclusions

- LSMs are useful for speech recognition as they are efficient and low power consuming due to their SNN structure.
- The training dataset must have all possible pitches for the model to work.
- Increasing the size of the Liquid Reservoir helps improve the accuracy.
- Increasing Filter size, slightly increasing accuracy but more hardware is needed for the filter.

References

[1] A. Gorad, V. Saraswat and U. Ganguly, "Predicting Performance using Approximate State Space Model for Liquid State Machines," 2019 International Joint Conference on Neural Networks (IJCNN), 2019, pp. 1-8, doi: 10.1109/IJCNN.2019.8852038.