Multimodal Neurons in LSMs

Kimaya Shikarkhane (19D070053)[1], Tanmay Jain (190020053)[1], Nivedya Nambiar (190070039)[1]
[1]*Indian Institute of Technology Bombay, Mumbai:400076, India*

Multimodal Liquid State Machines

Liquid State Machines are extensive recurrently connected spiking neural networks with the neurons organised in a reservoir. They are a model for real-time computations on continuous streams of data which explain their biological plausibility. In this way, the LSM acts as a filter that acts on a continuous input stream to produce a continuous output stream[3]. The LSMs are in essence a preprocessing layer to the actual learning layer that modifies the interpretation of the input stream by augmenting certain features after correlating them in time, and is hence an unsupervised learning method.

Liquid State Machines are proven to be successful for single modality classification problems like speech classification and image recognition, and these have been explored extensively in literature, with improvements to accuracy being made by introducing various modulating schemes like astrocyte modulation[1]. However the application of liquid state machines for combining multiple modalities has been relatively less explored owing to limited understanding behind their functioning.

Multimodal neurons in LSMs are expected to combine data from different modalities simultaneously to produce a better informed classification output. This is biologically plausible as the human brain indeed uses inputs from audio, visual and tactile sensors to distinguish and classify objects in the real world. Previously, authors in [2] proposed an architecture for multimodal LSMs that includes excitatory connections between neurons in the visual "cortex" and the auditory "cortex" as shown in Fig.1. These cross modal connections were proven to improve the classification accuracy of the LSM as compared to the case without the cross-modal connections by the authors in [2]. In our project we have tried to model our LSM after the same architecture of cross modal connections between auditory and visual parts of the liquid, as will be explained in our Methods section.

Results from Stage 1

In stage 1 of the project, we implemented separate liquid state machines that take audio and visual input to give an output classification for each sample. To train the LSM for visual inputs we used the MNIST 784 dataset while the TI-46 digits dataset was used for the audio inputs. The audio input waveforms were converted to spikes by first using the Lyon ear model followed by BSA encoding using the packages lyon and pyspikes in Python. The Lyon ear model implements filters and automatic gain control to convert the input audio waveform to a cochleogram. The BSA encoding scheme converts the cochleogram, which is a frequency-time representation to a spike pattern for each channel as a function of time. For the images, the pixel values over 28x28=784 dimensions were repeated over the number of iterations of STDP (20).

Each LSM was implemented according to the framework suggested in [1], using the STDP rule as given in the paper. Hence we included astrocyte modulation of the potentiation/depression rates in STDP. The astrocyte modulation aims at equalising the spiking rates of the input neurons and the reservoir neurons. This is expected to bring the LSM to edge-of-chaos operation which maximises accuracy of classification.

In our experiments we obtained the accuracy of the LSM operating on visual inputs to be 96.57% while the accuracy of the LSM operating on audio inputs was obtained as 61.75% for the train set and 52% for the test set. The resulting confusion matrix for the auditory inputs classification is given in Fig. 3 while the confusion matrix for LSM operating on visual inputs is given in Fig. 4. As can be observed in the figure, the test accuracy for the LSM working on audio inputs is very low and there is clear bias in classification, with most misclassifications being done where the input audio is being classified as 9. Especially so for the label 5, as the speech for "five" is quite similar to that of "nine" leading to the misclassification. Similarly for the LSM operating on visual inputs, the digit 4 is misclassified as 9 and vice versa, since the handwritten digits 4 and 9 closely resemble each other.

Project Aim for Stage 2

In this stage of the project, we aim to develop a liquid state machine that can take as input the visual information and the audio information simultaneously and use these to classify the input as belonging to one of 10 classes i.e., the

digits from 0-9. The liquid state machine will employ spike timing dependent plasticity to modify the weights in the network and improve classification accuracy.

Methods

We design a reservoir with $2x(12^3)$ neurons with half the neurons receiving and operating on inputs from the visual side while the other half receive auditory information. The reservoir is preceded by a set of visual input neurons and a set of auditory neurons that spike based on the visual input and auditory input respectively. We have used data from [4] which provides the spike data obtained from processing audio and visual information, hence this data is in effect the spiking pattern of the auditory input neurons and the visual input neurons. The data includes 60000 inputs for training and 10000 for testing with labels as digits ranging from 0-9. The visual input has 784 dimensions (channels) while the auditory input has 507 dimensions. These inputs are simultaneously applied to the reservoir having both the auditory and the visual cortex.

Liquid State Machine - Network connections

The visual input neurons are connected to the first N neurons of the reservoir with the weights being probabilistically initialised according to the scheme mentioned in [1], the same scheme is used for auditory input neurons that connect to the last N neurons of the reservoir. For both the visual and auditory parts of the reservoir, the ratio of excitatory to inhibitory neurons is 80%/20%. Input neurons did not have an excitatory/inhibitory distinction and had random excitatory and inhibitory connections to liquid neurons[1]. The connections between the neurons in the reservoir were made using probabilities based on Euclidean distance between the neurons, following the rule:

$$P(i,j) = C \times exp(-(\frac{D(i,j)}{\lambda})^2)$$

Here, the parameters C and λ in the liquid are as follows:

Type of Connection	C	λ
II (inhibitory to inhibitory)	0.3	9
EI (excitatory to inhibitory)	0.1	
IE (inhibitory to excitatory)	0.05	
EE (excitatory to excitatory)	0.2	

Table 1. Parameter values for different types of connections [1]

The input to liquid connection density is 15%. Based on whether the connection exists between two given neurons, the probability of which is given by P(i,j), the weights are initialised to a value of 3 (for EI, EE connections) and as -3 (II, IE connections). For the input to reservoir connections the weights are randomly initialised as excitatory (+3) or inhibitory (-3).

Liquid State Machine - Spike Generation

We follow the scheme of spike generation and update of membrane voltage as described in [1] for leaky integrate and fire neurons.

$$\frac{dv_i}{dt} = -\frac{1}{\tau_v} v_i(t) + u_i(t) - \theta_i \sigma_i(t)$$

$$u_i(t) = \Sigma_{i \neq j} w_{ij} (\alpha_u * \sigma_j)(t) + b_i$$

$v_i(t)$ is the membrane potential of neuron i in the LSM, while u_i is the synaptic response current. θ is the membrane potential threshold, assumed equal for all neurons in the reservoir, σ is the spike train of the i^{th} neuron. $\alpha_u(t)$ is the synaptic filter - $\alpha_u(t) = \tau_u^{-1} \exp(-t/\tau_u) H(t)$ where H(t) is the heaviside step function. The value of τ_u=1ms and τ_v=64ms. The LIF neurons all have a refractory period of 2ms. The spiking in the liquid neurons are averaged over time (number of iterations) and fed to the output layer which fits this spiking data against labels.

Spike Timing Dependent Plasticity

STDP is derived from the scheme mentioned in [1]. Although the original paper [1] describes an additional modification of the scale of STDP by astrocyte modulation, this has been omitted in this project. The equations for weight update is given by :

$$\frac{dw}{dt} = A_+ T_{pre} \Sigma_o \delta(t - t_{post}^{\ o}) - A_- T_{post} \Sigma_i \delta(t - t_{pre}^{\ i})$$

$$\tau_+^* \frac{dT_{pre}}{dt} = -T_{pre} + a_+ \Sigma_i \delta(t - t_{pre}^{\ i})$$

$$\tau_-^* \frac{dT_{post}}{dt} = -T_{post} + a_+ \Sigma_i \delta(t - t_{post}^{\ i})$$

Here A_+ and A_- are the potentiation and depression learning rates, both equal to 0.15. T_{pre} and T_{post} are the pre/post synaptic trace variables and $a_+=a_-=0.1$, $\tau_+^*=\tau_-^*=10$ms. The weight updates continue till the input is presented, according to the rules mentioned above.

Results

An accuracy of 98.5233% was obtained on the train dataset, and an accuracy of 97.39% was obtained on the test dataset. The confusion matrix displayed in Fig.2 shows the number of inputs correctly classified per class, also showing the predicted label for each true label in the test set.

Future Scope of Multimodal Liquid State Machines

The multimodal LSMs can prove to be useful in various real-life applications like gesture recognition where the video input can be combined with the accompanying speech for gesture classification, emotion recognition where facial features and speech can both be used to classify emotions. From our experiments, the performance of multimodal LSMs is better for classification than unimodal LSMs as is expected from intuition.

However, there is a dearth of exploratory work done in multimodal LSMs hence creating a lacuna that needs to be completed with more experiments to reinforce the utility of multimodal neurons in Liquid State Machines.

Code

Stage 1:

LSM operating on visual input
https://colab.research.google.com/drive/1nIli-o_Ep0Ama7pQzYagO-MA5cyHdGfJ?usp=sharing

LSM operating on auditory input
https://colab.research.google.com/drive/1sQ0u5mYYqIF7gtkSoptwPZr_HDGcCUms?usp=sharing

Stage 2:

Multimodal LSM
https://colab.research.google.com/drive/1Qt-fzyMpbifkymz5KIcUy86mFdAJvsVb?usp=sharing

[1] DOI: 10.48550/ARXIV.2111.01760
[2] DOI: 10.1109/TETCI.2018.2872014
[3] https://igi-web.tugraz.at/PDF/189.pdf
[4] https://zenodo.org/record/3515935#.Y38YC3ZBw2x

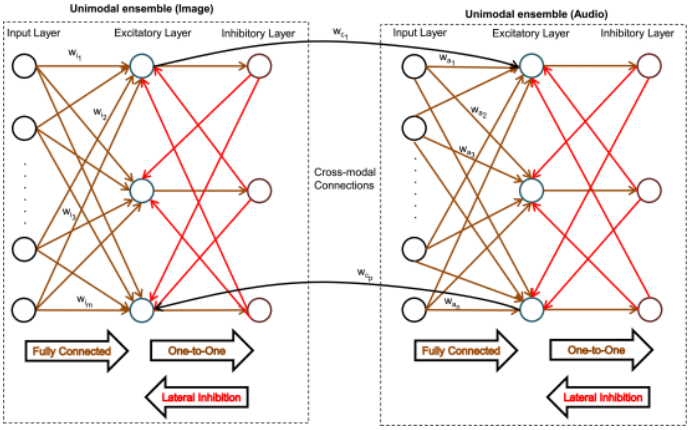

Fig. 1 Schematic of Multimodal LSM proposed in [2]. Image source: [2]

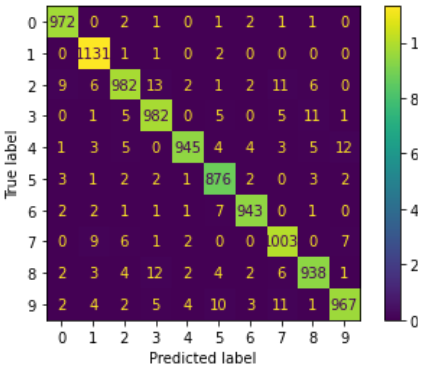

Fig. 2 Confusion matrix showing test performance for multimodal LSMs

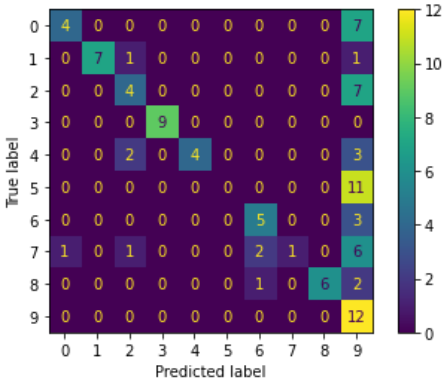

Fig. 3 Confusion matrix for LSM operating on auditory input

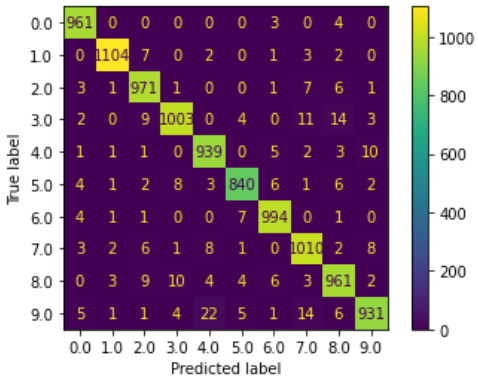

Fig. 4 Confusion matrix for LSM operating on visual input