Nick Riina

5/4/2021

Q355 Final Project Paper

Exploring the A-MLP

Introduction

The function of glial cells has gone through several stages of scientific understanding. Denoted by their name 'glial' meaning glue, these cells were originally thought to be packing material in the brain. This conception improved as scientific investigation continued, identifying major categories of these cells and their respective functions. Astrocytes, the most abundant of these cells (Maurizio et al 2019), have been conceived of as supportive cells for neurons: delivering nutrients, directing blood supply, and assisting in growth. In 1990, Anne Cornell-Bell and colleagues determined that glutamate evoked a calcium concentration increase in astrocytes. (Cornell-Bell et al., 1990) This internal calcium concentration was soon-after demonstrated to propagate to nearby astrocytes, hypothesized by many to be an astrocytic communication network. (Maurizio et al 2019) These discoveries prompted investigation into the roles astrocytes may have in neural information processing and in cognition in general. As a part of these investigations, computational approaches to simulating the behavior of astrocytes have been proposed. Several studies have shown evidence that their astrocytes are improving the learning-time needed and the accuracy reached with their networks, however these models have significant differences (Gonzalez et al. 222), (Wade et al. 2011), (Sajedinia 2015), (Ikuta 2011). In this project I reproduced one of these astrocyte models that run alongside a multi-layer perceptron (MLP).

Proposed Astrocyte models

The computational approaches to illuminating the functional role of astrocytes in neural learning has a diffuse literature. Studies focused on slightly different conceptions of astrocytes generate widely different models that likely have spurious roles. Despite the differences between these models, many have a few items in common:

1. Astrocytes are thresholded, needing sufficiently large neural activity to initiate calcium release into the cytoplasm.
2. Astrocyte calcium concentration decays over time.
3. Internal calcium concentration results in astrocytic input passing back into the neural network.
4. One astrocyte unit per neural unit, following the cortex distribution of astrocytes of neurons (Gergel 2018).

Some other features have been incorporated by some models but not all:

1. Astrocyte activity moves at a slower time scale than neurons, due to the properties of chemical signaling in contrast to the electrical signaling of neurons
2. Astrocytes as able send excitatory input (glutamate) and inhibitory input (ATP) into synapses
3. Astrocyte waves can propagate to nearby astrocytes
4. Single Astrocytes can monitor and affect several neurons or synapses simultaneously
5. Astrocytes may receive input from the presynaptic neuron, the postsynaptic neuron, and the synaptic concentration itself.

Below are some images to demonstrate the plurality of astrocytic models:

Figure from Wade et al. 2011.

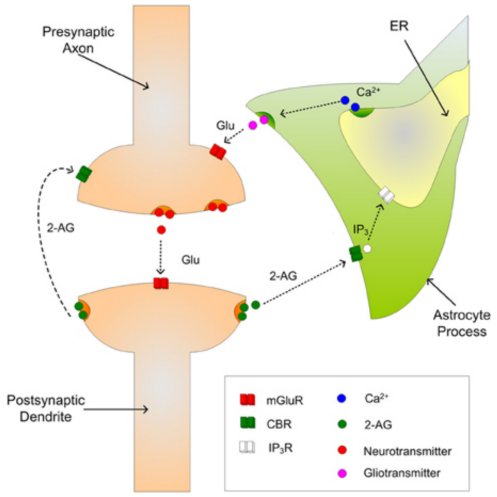


Figure from Sajedinia 2015.

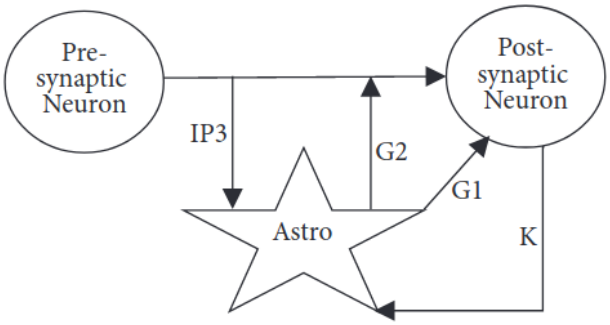


Figure representing Gergel 2018.

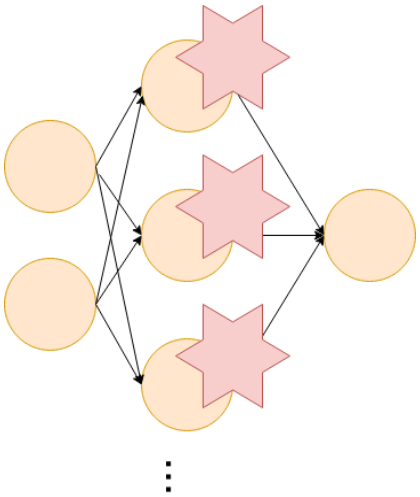
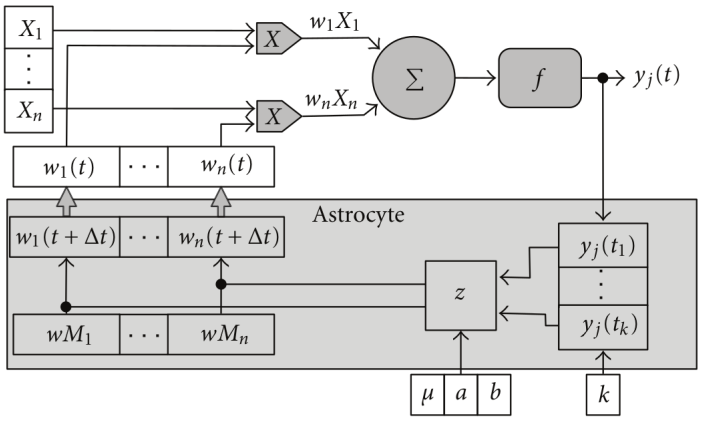


Figure from Sajedinia 2018.



In the top two models, input to the astrocytes follows a more biologically realistic interpretation, with input coming from both pre- and post-synaptic neurons. In the lower two models, astrocytic activity is calculated only with the corresponding neuron's activity.

For this project I chose one of the simpler models which computed astrocytic behavior based on neural activity instead of any biological-corresponding signaling pathway. These models were also convenient because their initial experiments both utilized MLPs, a relatively common artificial neural network. Out of the models of this sort there was still a choice between two classes of behaviors for the astrocytes:

1. Glial learning
2. Tripartite synapse

Glial learning describes when an astrocyte influences the network by directly changing the weights of the corresponding neuron. Tripartite synapse algorithms describe when the activity of an astrocyte is a number, added into the activation function of its corresponding neuron. Both of these astrocyte models apply changes to the networks function based on neural activity, however in different ways. Tripartite models contribute similarly to how other neurons do, while the other algorithm (glial learning) acts as a pseudo learning rule, adjusting the weights of the MLP over the duration of use (Sajedinia 2018).

This experiment utilized a tripartite model, referred to in this paper as the Gergel AMLP (in reference to Gergel et al. 2018). This model was chosen for its simplicity, having less parameters than the glial learning model. Each Gergel astrocyte may be completely described with three parameters: a neural-firing threshold, a decay rate, and weight. The glial learning models, on top of these three, also contained a weight -increase percentage, weight-decrease percentage, and running activity parameters (Sajedinia 2018). The tripartite model was also chosen for biological plausibility of the astrocytic activity itself. When the neural firing threshold is reached, the astrocyte activity initializes at 1 and decays based on its decay rate, which corresponds to the release and decay of a cytoplasmic calcium in biological astrocytes. The glial learning model applied changes to the weights based on neural activity directly, with no

modeling of internal activity. A third reason the tripartite model was chosen was for interpretability of results. Each astrocyte's activity is a number, much like those of neurons. This type of behavior was hypothesized to be more interpretable akin to its similarity to how artificial neurons are modeled.

In this report, I focused on recreating the Gergel AMLP in order to illuminate the behavior of the astrocytic units. Future studies will be needed to compare these models against biological data.

Methodology

The astrocyte model and parameter-update rules were adopted from a paper by Peter Gergel and Igor Farkas (2018). The paper provided only equations, so all astrocyte and MLP code was created by hand in python 3. Numpy and Matplotlib libraries were both used.

MLP Parameters:

Equation 1 - Hidden unit activation rule for MLP

$$h_i(t+1) = f(\sum_{j=0}^{M} w_{ij} x_j(t) + \alpha \psi_i(t))$$

The activation rule for all neural units was a sigmoid activation function. The neurons did not contain biases (not explicitly mentioned by Gergel and Farkas) and all networks were trained with backpropagation. For both experiments, the MLP had one input layer, one hidden layer, and one output layer. The amount of input and hidden units varied with the experiment, the output layer was constant at 1 neuron. The performance of the MLP was tracked with the Mean of Squared Errors following Equation 2.

Equation 2 - Mean of Squared Error

$$\frac{\sum(actual - prediction)^2}{number\ of\ output\ units}$$

Table 1 - Astrocyte Parameters:

Parameter Symbol	Biological Correlate
θ	Threshold for initiating astrocyte activity in response to neural activity
γ	Decay rate for astrocytic activity
α	Astrocytic weight for sending input back to neuron

Equation 3 - Astrocyte Activity Update rule

$$\psi_i(t) = \begin{cases} 1, & \text{if } \theta < h_i(t-1) \\ \gamma\psi_i(t-1), & \text{otherwise} \end{cases}$$

Equation 4 - Decay Rate update rule(s)

$$\gamma_i(t+1) = 1 - \langle\psi_i(t)\rangle_t$$

$$\gamma_i(t+1) = \gamma_i(t) + \eta_\gamma(1 - \langle\psi_i(t)\rangle_t - \gamma_i(t))$$

Equation 5 - Threshold update rule(s)

$$\theta_i(t+1) = \langle\psi_i(t)\rangle_t$$

$$\theta_i(t+1) = \theta_i(t) + \eta_\theta(\langle\psi_i(t)\rangle_t - \theta_i(t))$$

Gergel and Farkas offered two potential update rules for the θ and γ values. The first rule (rule 1) updated the threshold by changing it to the average astrocyte activity of that astrocyte (each astrocyte had its own θ and γ). Rule 1 also set the decay rate to the inverse of average activity.

Rule 2 (underneath rule 1) incrementally adjusted these values toward average activity (or inverse for the decay rate). The learn rate and rule for determining average astrocyte activity were not discussed in the paper. The learning rule chosen for this experiment was Rule 1. Astrocytic weights were updated with a small variation of backpropagation discussed by Gergel and Farkas. This backpropagation followed the same formula as backpropagation for the hidden weights, scaled by the current astrocyte activity rather than the activity of the hidden unit.

Several variations of the Gergel AMLP were used (See Table 2). Within the parenthesis of these different models, combinations of the three parameters appear. For each of these models, the parameters within the parenthesis are dynamic, those not in the parenthesis are stable. The stable values were found in an iterative process referred to as Grid search explained below.

Table 2 Model Variations:

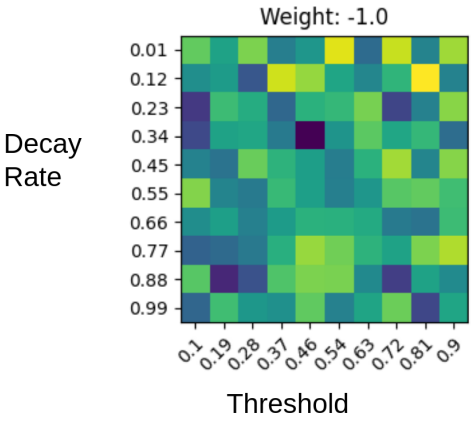
Model Variation	
AMLP	Parameters are stable over entire network and duration - values determined through gridsearch
AMLP(α)	Astrocytic weights (α) trained with backpropagation
AMLP(θ)	Neural firing threshold (θ) adjusted individually for each astrocyte
AMLP(γ)	Astrocytic Decay Rate (γ) adjusted individually for each astrocyte
AMLP(θ,γ)	θ and γ both adjusted
AMLP(α,θ,γ)	θ, γ, adn α adjusted

Table 3 Grid Search Iterations

Parameter	Values tried in Gridsearch
α	-1.00, -0.78, -0.56,-0.33, -0.11, 0.11, 0.33, 0.56, 0.78, 1.00
γ	0.01, 0.12, 0.23, 0.34, 0.45, 0.55, 0.66, 0.77, 0.88, 0.99
θ	0.10, 0.19, 0.28, 0.37, 0.46, 0.54, 0.63, 0.72, 0.81, 0.90

Gridsearch consisted of iterating through a set of values (Table 3) in order to find the best combination. Each trio was trained for 5 iterations, 1500 epochs of training each, with the average mean of squared errors (MSE) recorded (a single box in Figure 1). The trio with best (lowest) SSE was taken for the stable values (values not in parenthesis in models). A gridsearch was run for each experiment to find ideal values. Results were graphed as a heatmap with lighter values indicating better performance.

Figure 1 - Gridsearch Example



Gridsearch graphing code adopted from:

https://matplotlib.org/stable/gallery/images_contours_and_fields/image_annotated_heatmap.html

Experiments: N-Parity

The N-parity dataset was composed of a set of inputs and outputs. The inputs are lists of length *N*, (N as in N-Parity) containing only 1s and 0s. The job of the network is to determine whether there are an even or odd number of 1's in the input. The output was a series of 1's and 0's, with 1 denoting even and 0 denoting odd.

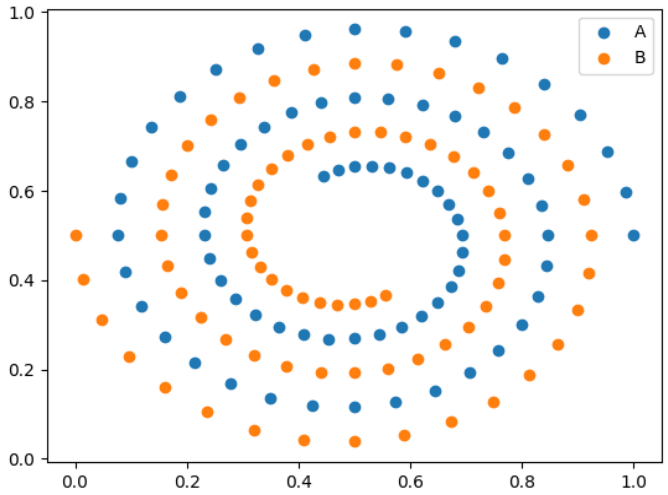
The experiment was run from 2-parity to 9-parity. The number of hidden units in the MLP was also N as specified by Gergel and Farkas.

The dataset was randomly generated with 2^N data points. For this experiment there was no validation as specified by Gergel and Farkas, so the entire dataset was used for training. There was no public dataset specified in the paper so this was coded from scratch using Python 3, Numpy and Matplotlib libraries.

Experiments: Two-Spirals

The two spirals dataset consisted of a set of (x,y) coordinates belonging to two interwoven spirals.

Figure 2 - Two Spirals Example

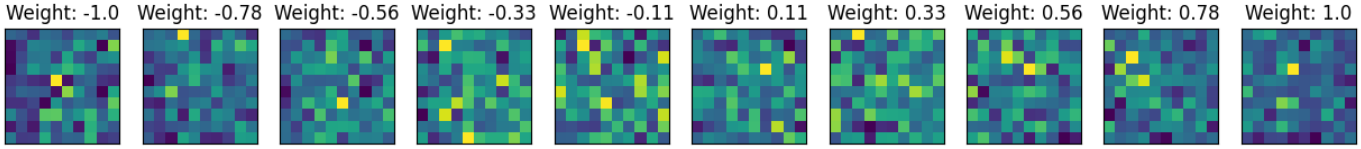


The dataset was divided into two sets: a train and validation set, with 80% of the data points in the train set. The size of each spiral (how many points in each spiral) were not identified in the paper, so in my experiments I used 30 points per spiral (see Figure 1). 30 points was empirically determined to be a good balance between computation time and achieved accuracy. The tightness (how close each spiral is to the other) was not specified in the paper and was arbitrarily chosen by the two-spirals generation code found online. The online code created a list of (x,y) coordinates dividing into two interleaved sets of spirals. Spiral code found at: https://conx.readthedocs.io/en/latest/Two-Spirals.html

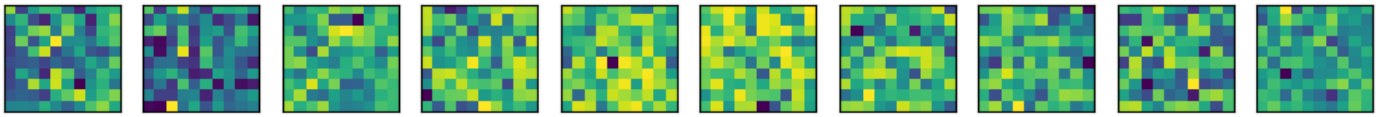
Results: N-Parity

Figure 3 - Grid Search Results

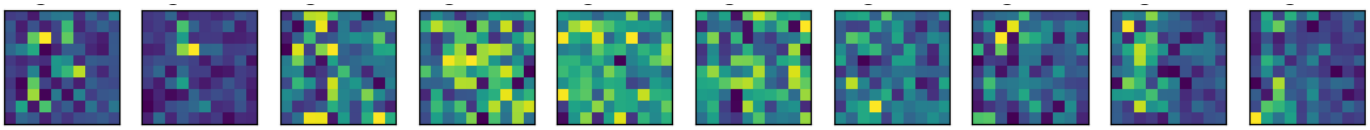
5-Parity first run (5 iterations for 1500 epoch)



5-Parity Second run (5 iterations for 3000 Epoch)



5-Parity Run, Learn Rule 2 (5 iterations for 3000 Epoch)



Each box corresponds to an average MSE between 5 instances, with lighter colors denoting better performance

Table 3 GridSearch Iterations

Parameter	Trio Used	Other GridSearch results	
α	1.0	0.11	- 0.78
Θ	0.46	0.1	0.28
γ	0.34	0.01	0.99

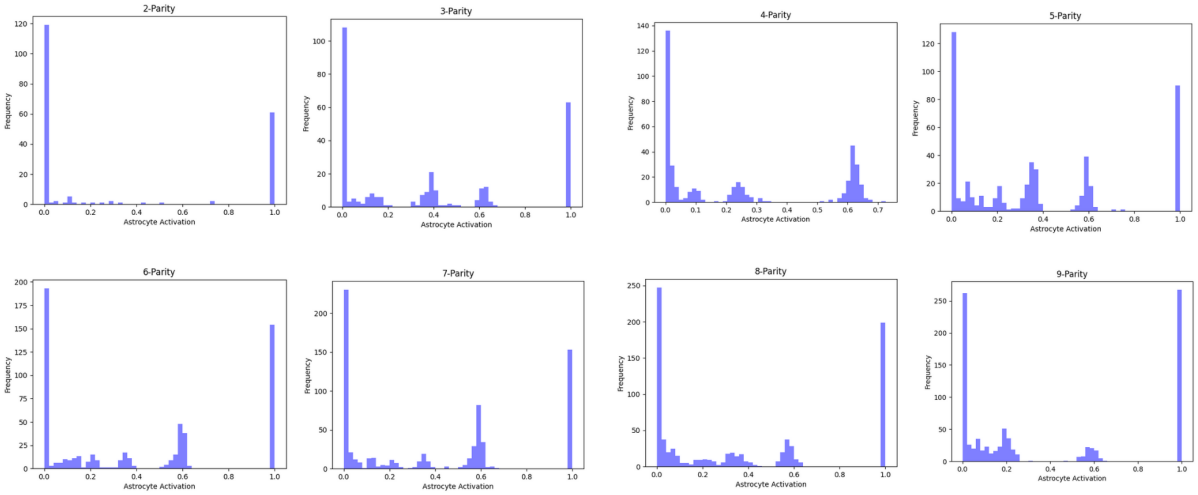
Gergel and Farkas indicated that they used the first learn rule and ran each combination 5 times, however the exact way of choosing the best 'trio' was not specified. In this paper, the first achieved trio of values was used for experiments due to limited time. When subsequent grid searches were run with the same initial starting conditions (5 iterations, 1500 training epoch's), the results were inconsistent (see Table 3). After most of the experiments were conducted, other grid searches were run with twice as many training Epochs (Figure 3). The results seem to indicate that longer grid searches yield a clearer gradient between better and worse combinations of values.

Table 4 N-Parity Mean of Squared Error Results

Model	4-Parity	6-Parity
MLP	0.0702 +/- 0.019	0.1185 +/- 0.0035
AMLP	0.1014 +/- 0.006	0.1227 +/- 0.0021
AMLP(α)	0.1028 +/- 0.023	0.1121 +/- 0.0091
AMLP(Θ)	0.0594 +/- 0.028	0.1217 +/- 0.0031
AMLP(γ)	0.1023 +/- 0.014	0.1231 +/- 0.0024
AMLP(Θ, γ)	0.1031 +/- 0.015	0.1179 +/- 0.0048
AMLP(α, Θ, γ)	0.0855 +/- 0.024	0.098 +/- 0.0098

Values within parentheses were updated with their respective learning rules, highlighted values were the experimental best values. Results for N-Parity were not significant ($p < 0.5$). Each datapoint was the result of 100 instances, each trained for 10,000 Epochs

Figure 4 - Final Astrocyte Activity for N-Parity



The activity of the astrocytic units of 100 instances of the most successful model (AMLP(α, Θ, γ)) were gathered at the end of training 1500 epochs, for 100 instances each (Figure 4). The majority of Astrocytes seem to form two regimes: fully activated, and not active at all.

Results: Two-Spirals

Table 5 Two Spirals Gridsearch Results

Parameter	Trio Used
α	1.0
Θ	0.28
γ	0.66

Grid search was only run one time for the two-spirals dataset due to time constraints.

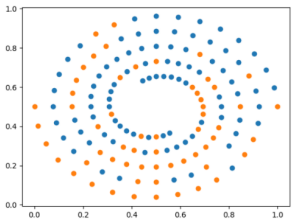
Table 6 Two-Spirals Mean of Squared Error

Model	Train	Test
MLP	0.1258 +/- 0.0002	0.1272 +/- 0.0003
AMLP	0.1093 +/- 0.0064	0.1613 +/- 0.0199
AMLP(α)	0.1025 +/- 0.0086	0.1393 +/- 0.0183
AMLP(Θ)	0.1085 +/- 0.0070	0.1776 +/- 0.0219
AMLP(γ)	0.1078 +/- 0.1113	0.1508 +/- 0.0192
AMLP(Θ, γ)	0.1008 +/- 0.0111	0.1709 +/- 0.0.0371
AMLP(α, Θ, γ)	0.09484 +/- 0.01128	0.1614 +/- 0.0181

Results for Two-Spirals were statistically significant; the basic MLP without any astrocyte units had the lowest MSE with the validation dataset. Each datapoint was the result of 100 instances, each trained for 5,000 Epochs.

Figure 5 - Two Spirals Results Visualization

AMLP(α,Θ, γ) Trial
MSE = 0.095



MLP Trial
MSE = 0.125

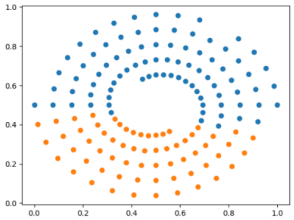
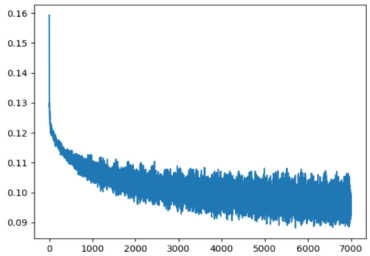


Figure 6 - Two Spirals Results Learning Visualization

AMLP(α,Θ, γ) Trial
MSE = 0.095

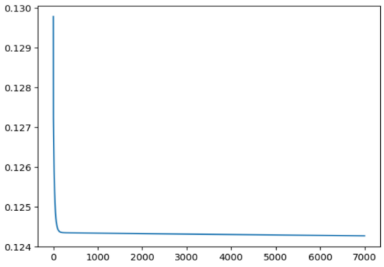
MSE



Epoch

MLP Trial
MSE = 0.125

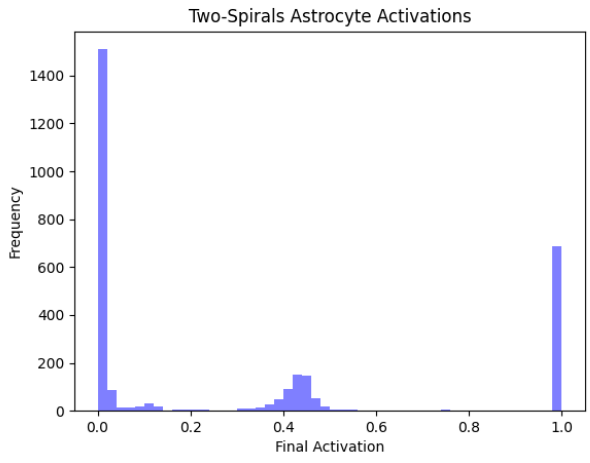
MSE



Epoch

Figures 5 and 6 show a single instance of the AMLP(α, Θ, γ), which achieved the best results for the Two-Spirals Experiment. The learning plots looked similar between the different models, however direct comparisons between the learning trajectories still need to be performed. What is apparent from Figure 6 is that the astrocytes (in the most dynamic model) seemingly help the network avoid the same local minima asymptote as seen in the MLP trial (Standard Deviation = 0.0002)

Figure 7 - Final Astrocyte Activity for Two-Spirals



Discussion

The results of this experiment were compared to the results of Gergel and Farkas' original paper (Gergel 2018). The experiments and analyses were performed as close to the methods of the original paper as possible. Both experiments found the Gergel AMLP to lead to overall lower-MSE scores achieved in networks, with their influence seemingly to improve as the network size increased. The method this improvement came from is relatively unclear. Figure 6 shows how the astrocyte units seemed to help the network avoid getting stagnant within a single local minima, continuing to decay slowly instead of reaching a single minima like that observed in the MLP learning trials. However, the different models had different dynamics, and likely affected the learning of the networks in slightly different ways.

In both experiments (Gergel and Farkas as well as this one), models with dynamic thresholds were typically the best performing (in all trials except Gergel 2018 6-Parity). When one references the threshold learning rule (Equation 5), one can see that astrocytes with no activity have thresholds that decrease. Eventually the threshold will reach a value of 0 (or close to 0) enough to interject some noise into the neuron. This activity would cause neurons that 'die' or go long periods without firing to gain sudden life and activity. This leads to the hypothesis that these astrocytic units (with dynamic thresholds) are working as a sort-of dynamic noise value, interjecting more noise into neurons that fire very little than neurons that fire very often. When a neuron fires often, the corresponding astrocyte also receives lots of input, driving the threshold to higher and higher values, making it harder for the astrocyte to fire.

Table 7: Gergel 2018 Results

Task	Best Model	Stable Parameters
4-Parity	AMLP(Θ, γ)	a = -0.5 , Θ = 0.5 , γ = 0.5
6- Parity	AMLP	
Two- Spirals	AMLP(Θ, γ)	a = -0.1 , Θ = 0.1 , γ = 0.5

Table 8: Summarized Results

Task	Best Model	Stable Parameters
4-Parity	AMLP(Θ, γ)	a = 1.0 , Θ = 0.46 , γ = 0.34
6- Parity	AMLP(α, Θ, γ)	
Two- Spirals	AMLP(α, Θ, γ)	a = 1.0 , Θ = 0.28 , γ = 0.66

A full interpretation into the ways these astrocyte units are helping the MLP achieve lower MSE scores is flawed due to the inconsistency of the stable results. Further, disparities between Gergel and Farkas' findings and the findings of this experiment also cast doubt into the significance of these conclusions. Gergel and Farkas achieved a lower MSE for most of their models, this could have been a result of the construction of the dataset, the existence of a bias term that is just omitted from the published paper, or a normalization process not explicitly stated. Gergel and Farkas' also found entirely different ideal stable astrocytic parameters as well as different best performing models.

Figure 8: N-Parity MSE comparison to (Gergel 2018)

Riina MSE

Model	4-Parity	6-Parity
MLP	0.0702 +/- 0.019	0.1185 +/- 0.0035
AMLP	0.1014 +/- 0.006	0.1227 +/- 0.0021
AMLP(α)	0.1028 +/- 0.023	0.1121 +/- 0.0091
AMLP(Θ)	0.0594 +/- 0.028	0.1217 +/- 0.0031
AMLP(γ)	0.1023 +/- 0.014	0.1231 +/- 0.0024
AMLP(Θ, γ)	0.1031 +/- 0.015	0.1179 +/- 0.0048
AMLP(α, Θ, γ)	0.0855 +/- 0.024	0.098 +/- 0.0098

Gergel (2018) MSE

Model	4-parity	6-parity
MLP	0.081 ± 0.060	0.065 ± 0.035
A-MLP	0.083 ± 0.086	$0.059 \pm 0.034*$
A-MLP(α)	0.080 ± 0.065	0.072 ± 0.054
A-MLP(θ)	0.083 ± 0.075	0.065 ± 0.036
A-MLP(γ)	0.087 ± 0.065	0.062 ± 0.034
A-MLP(γ, θ)	$0.074 \pm 0.051*$	0.063 ± 0.055
A-MLP(α, γ, θ)	0.092 ± 0.072	0.078 ± 0.056

Figure 9: Two-Spirals MSE comparison to (Gergel 2018)

Riina MSE

Model	Train	Test
MLP	0.1258 +/- 0.0002	0.1272 +/- 0.0003
AMLP	0.1093 +/- 0.0064	0.1613 +/- 0.0199
AMLP(α)	0.1025 +/- 0.0086	0.1393 +/- 0.0183
AMLP(Θ)	0.1085 +/- 0.0070	0.1776 +/- 0.0219
AMLP(γ)	0.1078 +/- 0.1113	0.1508 +/- 0.0192
AMLP(Θ, γ)	0.1008 +/- 0.0111	0.1709 +/- 0.0.0371
AMLP(α, Θ, γ)	0.09484 +/- 0.01128	0.1614 +/- 0.0181

Gergel (2018) MSE

Model	Train	Test
MLP	0.075 ± 0.067	0.094 ± 0.066
A-MLP	0.073 ± 0.067	0.088 ± 0.068
A-MLP(α)	0.050 ± 0.049	0.078 ± 0.050
A-MLP(θ)	0.034 ± 0.045	0.049 ± 0.046
A-MLP(γ)	0.068 ± 0.065	0.085 ± 0.063
A-MLP(γ, θ)	$0.030 \pm 0.035*$	$0.051 \pm 0.041*$
A-MLP(α, γ, θ)	0.060 ± 0.051	0.095 ± 0.051

The most alarming difference came from the validation error in the two-spirals experiment (Figure 9). This experiment achieved the lowest validation error for the MLP, while the best performing AMLP model achieved the lowest validation error for Gergel and Farkas. This disparity could also be caused by several factors. One causal reason could refer to the creation of the datasets and the treatment of simulated time. A standard MLP computes an answer per input, with subsequent inputs having no direct effect on each other, besides their error. For the AMLP models, previous activity may affect current activity, either by influencing the current threshold or decay rate, or by means of previous activation causing the astrocytic unit to add input to a neuron that did not pass the firing threshold for that datapoint. As a result of the novel affects subsequent training data points may have on each other, its realistic that different configurations of astrocytes are advantageous for particular orderings of the data points. This would sensibly lead to a poor validation error, as the testing dataset could not possibly follow the same progression without being the same dataset in the same order. The lack of this observation in Gergel and Farkas' results casts doubt on whether or not this process occured in their experiments. These interactions may also change how gradient descent based learning may ideally work. This experiment utilized basic back propagation as typically used with an MLP, however the astrocyte activity is influenced by subsequent trials, which may be more appropriate for a back-propogation through time, or genetic algorithm that places less weight on

the current state of the astrocyte for each training datapoint. The effect subsequent trials have on each other may also be mitigated by having a time parameter that runs orthogonal to the simulated time of the experiment, as in Sajedinia 2015 and 2018. In these experiments, a single input is presented to the MLP. Then astrocyte activity is calculated over a time window, all for this same input. Once this window was finished, the overall effect of the astrocytes was sent through the network. This particular paper was a glial learning algorithm, so this influence was a change in synaptic weights which pairs with this orthogonal treatment of time better than the Gergel AMLP, which treats the neural-astryctic interaction as ongoing.

Another source of disparity is in the final activations achieved.

Figure 9 - Gergel 2018 N-Parity Astrocyte activity results

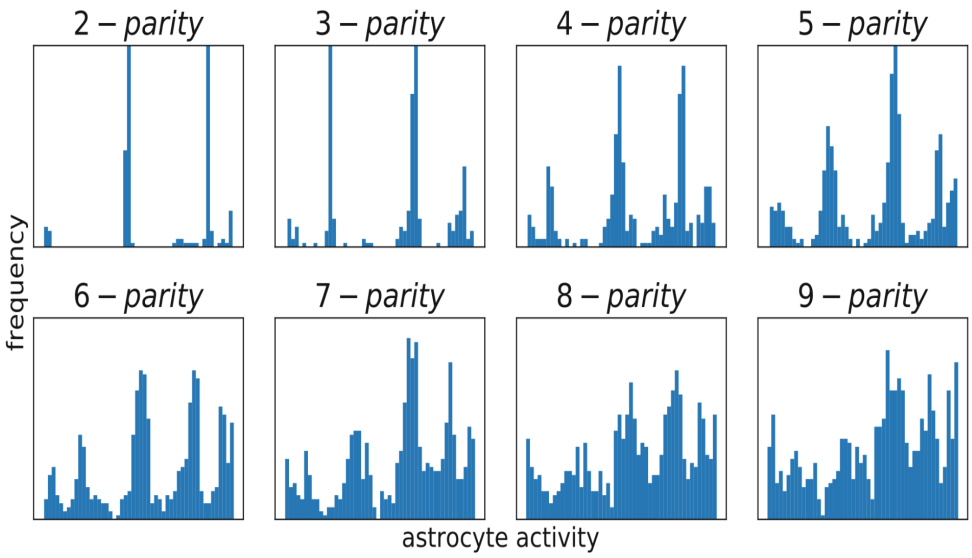
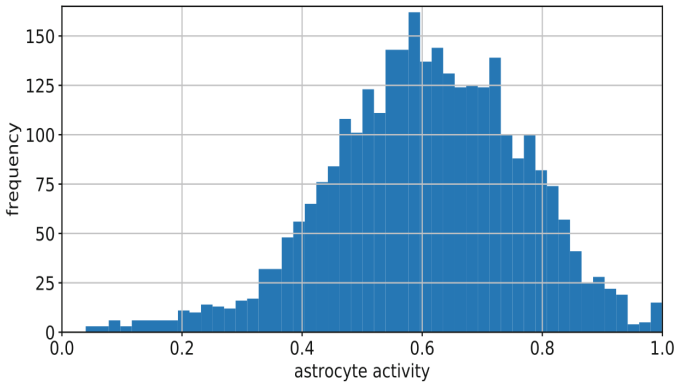


Figure 10 - Gergel 2018 Two-Spirals Astrocyte activity results



When comparing Figure 9 to Figure 4, and Figure 10 to Figure 7, it is obvious that the Gergel Astrocytic units reached a greater diversity of activations at the end of training trials. This could have been caused by the Gergel paper actually using Decay Rate and Threshold update Rule 2 instead of Rule 1 (See Equation 2 and 3), with some learning rate that was not specified. The sentence in the paper was ambiguous, and the omission of the relevant parameters led researchers to use Rule 1. These results also could come from different models being visualized. Gergel and Farkas found a different 'best' combination of parameters for each experiment, which would cause their astrocytes to behave differently than mine for stable models. Gergel and Farkas did not specify which model was used to create these images. Figure 4 and 7 were made using the best performing model for each experiment, however having different 'best' models with different 'best' parameters sensibly leads to different results.

If this experiment was repeated, larger grid searches would be performed, and different parameter selection methods would be tried to find a way that produced consistent results. The grid searches performed after the project presentation (Figure 3) showed potential for stable results. These comparisons would also be improved with communication between the original authors to find out for certain the update rules, creation of datasets, and methods of finding 'best' values.

Further, if this experiment was repeated with more time, more comparisons would be made between the model variations. The specific learning trajectories, the final state of the astrocytic parameters, and progression of these values over training would be formally compared. A

This and further experiments are needed to illuminate the neurobiological role of astrocytes, and how they affect intelligence and cognitive functions in humans. Studies that incorporate the slower activation of astrocytes compared to neurons and the tendency for astrocytic activity to

propagate to nearby astrocytes are more likely to uncover the ways biological astrocytes interact with neurons. Further, most astrocytes cover several synapses between different neurons, so the 1:1 neuron:astrocyte models is also a simplification that leads to disparities between biological and artificial astrocytes. Studies of this nature also have the potential to illuminate concepts in neural learning that can help a network avoid local minima.

There have been more biologically realistic, as mentioned in the introduction (Gonzalez et al. 222), (Wade et al. 2011), (Sajedinia 2015), (Ikuta 2011). Most of these studies have demonstrated similar results to Gergel 2018, and this experiment: that Astrocytic units can improve the overall learning capabilities of neural networks of architectures ranging from basic MLP's (Gergel 2018, Ikuta 2011), to recurrent networks (Sajedinia 2015), to Echo-state Networks and reservoirs (Sajedinia 2015). However few of these studies have explored the exact activity of the astrocytes and *how* they assist the learning of these networks. Most of the hypotheses place the burden of explanation as a type of tunable, dynamic noise. However models with connectivity between the astrocytes also include hypotheses that treat artificial neural-astrocytic networks as complementary learning systems, whereas the differences between neurons and astrocytes are exploited for learning (Sardinia 2015).
The next step for this experimentation team is to add more biological aspects to this model. Connections between astrocytes and the slow-chemical activation of astrocytes in comparison to the electrical signaling of neurons seem to this research team to be the most relevant while keeping the astrocytic model simple for efficient computation. Ideally research in this direction would uncover the principles by which an astrocytic network may assist different architectures of neural networks in learning, as well as identifying a simple astrocytic network that may increase the capabilities of neural networks while remaining computationally efficient.

Citations:

Alvarellos-González A, Pazos A, Porto-Pazos AB. (2012) Computational models of neuron-astrocyte interactions lead to improved efficacy in the performance of neural networks. Comput Math Methods Med. 2012;2012:476324. doi: 10.1155/2012/476324. Epub 2012 May 9. PMID: 22649480; PMCID: PMC3357509.

C. Ikuta, Y. Uwate and Y. Nishio, "Performance and features of Multi-Layer Perceptron with impulse glial network," *The 2011 International Joint Conference on Neural Networks*, 2011, pp. 2536-2541, doi: 10.1109/IJCNN.2011.6033549.

Cornell-Bell, A. H., Finkbeiner, S. M., Cooper, M. S., and Smith, S. J. (1990). Glutamate induces calcium waves in cultured astrocytes: long-range glial signaling. Science, 247(4941):470–473.

Gergel, Peter & Farkas, Igor. (2018). Investigating the Role of Astrocyte Units in a Feedforward Neural Network. Faculty of Mathematics, Physics and Informatics, Comenius University in Bratislava. Mlynská dolina, 84248 Bratislava, Slovak Republic

Maurizio de Pittà, Hugues Berry. A Neuron-Glial Perspective for Computational Neuroscience. Computational Glioscience, Spinger, pp.3-35, 2019, Springer Series in Computational Neuroscience, 978-3-030-00817-8. ff10.1007/978-3-030-00817-8_1ff. ffhal-01995849f

Sajedinia Zahra. Glia-Augmented Artificial Neural Networks: Foundations and Applications. A thesis submitted to the School of Graduate Studiesin partial fulfillment of the requirements for the degree ofMaster ofComputer ScienceDepartment of Computer Science .Memorial University of Newfoundland. May 2015

Sajedinia Zahra, Hélie Sébastien (2018). A New Computational Model for Astrocytes and Their Role in Biologically Realistic Neural Networks. Computational Intelligence and Neuroscience

Wade JJ, McDaid LJ, Harkin J, Crunelli V, Kelso JAS (2011) Bidirectional Coupling between Astrocytes and Neurons Mediates Learning and Dynamic Coordination in the Brain: A Multiple Modeling Approach. PLOS ONE 6(12): e29445. https://doi.org/10.1371/journal.pone.0029445

Relevant URL's

Two-Spiral generation: https://conx.readthedocs.io/en/latest/Two-Spirals.html

Heatmap plotting for Grid Search: https://matplotlib.org/stable/gallery/images_contours_and_fields/image_annotated_heatmap.html

Github with project code including datasets: https://github.com/nriina/AMLP_Project