Data Augmentation of Neural Spike Data for Single Trial Dynamics in M1

Jason Nan
Jacobs School of Engineering
University of California, San Diego
La Jolla, CA, 92023
janan@ucsd.edu

Sidharth Hulyalkar
Jacobs School of Engineering
University of California, San Diego
La Jolla, CA, 92023
shulyalk@ucsd.edu

***Abstract*— Data collection for machine learning is one of the biggest bottlenecks of this field, especially neural data from live subjects. Taking inspiration from object recognition models, data augmentation may be useful for neural data to amplify the number of data points for training your model. In this paper, two methods of data augmentation, shifted binning and noise simulation, are used with neural spike data from the Shenoy Lab [1]. A Kalman filter is used to decode arm movement of a monkey from neural spike data. Performance was compared using the L1 norm to calculate error between predicted and actual arm position. We found that there needed to be a minimum number of raw trials (300) or else the training would fail. After 300, increasing the number of trials did not seem to change the performance by much. However, the shifted datasets showed a 18% increase in performance while the simulated noise decreased performance by about 30%. This reinforces the idea that temporal summation is more weighted in the M1 cortex and data augmentation could be a valid way to extract more information out of a single trial. The same procedure was done after selecting only 20 neurons to use. The same trend is seen, along with increasing performance correlated to increased number of neurons.**

I. INTRODUCTION

One of the biggest challenges in machine learning is gathering data used for training and testing the model. This is especially salient in the field of neural engineering where the data comes from doing experiments from primates. Experiments are not only time consuming and expensive, the necessary data comes from animals which need to have electrodes implanted in their brains to record the neural spikes. After the primate is trained to run the experiment, a single trial could take hours **[2]**. For the neural spike data from the Shenoy Lab, with 728 trials for both testing and training, it could take weeks or months to get enough clean data. Using data augmentation, the required amount of real trials could be lessened by augmenting each data point in a way that creates new data points from the existing set. Intuition for this problem comes from data augmentation done on images for object recognition. The augmentation they do includes, image rotation, random deletion of pixels, color shifting, etc **[3].** This is common practice and shows favorable results. Instead of gathering a large number of different pictures of the same object, only a subset of the image pool is necessary, and the number of data points N can be expanded using data augmentation. For image recognition, it saves both time and increases accuracy of the classifier.

For neural spike data, it would additionally save money, and reduce the strain on the animal subjects used to gather data. Data augmentation of this type can also be used in human BCI applications in which the patient is disabled. Decreasing the amount of data needed would mean less training time for the patient and a more robust BCI device for the public.

In this paper, we will be assessing two different types of data augmentation methods with the data provided by the Shenoy Lab **[1].** The first method is a bin shifting method where the start of the binning will shift by one data point every time for the length of the bin. The second method will be adding gaussian noise to the spike data to simulate measurement noise between trials. The same procedure was repeated on a new training set where 20 neurons were selected at random.

The model is a simple Kalman filter for dynamic time modeling, and the performance is measured using an L1 norm between the predicted position and true position across time.

II. METHODS

A. Dataset

The data we are using is from the Shenoy lab where a monkey is reaching in 8 different directions. There are 91 trials for each of the directions forming a total of 728 trials. The recording is done with a Utah array at 97 neurons, from the start of movement to the end of movement when the target is reached.

The testing set has the same attributes as the training, giving a total of 91x8x2 trials.

B. Data Preprocessing

The raw data for each trial lies in R^D where D is the number of neurons. Each trial ranges in time and is discretized in 1ms increments containing values of either 1 or 0 for if a spike was detected. The data was processed by binning at 20ms increments to get a spike count vector. 15 ms was chosen based on a paper by Shenoy et al **[1]** The same procedure is done to the testing set with no data augmentation.

C. Preseeding

In order to properly configure our dataset we decided to randomly shuffle and reshape the dataset. This ensures that the trials that we are sampling from are completely random and will remove some angle bias. Groups of data were selected from 100 to 700 stepping in 100 to assess the effect of augmentation at different amounts of data. The goal of this is to see better performance in the lower end of trial numbers. At the upper end, the model may have hit a plateau where additional data would only marginally boost performance. This procedure is done 5 times for each group of trials to get rid of any trial biases where some might just fit the model better than others. After this step, there will be 5 repeats for each of the 7 groupings, a total of 35 separate models will be trained for each category (original, shifted, noise).

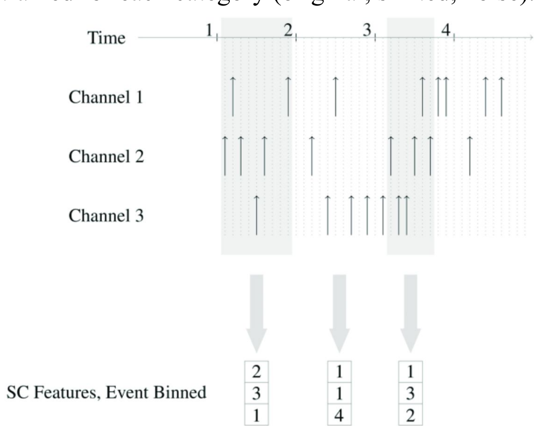

Figure 1: Visualization of the binning style done on the raw data. 20ms of recording data are summed up to determine the total number of spikes in that bin **[4]**

The data points at the end of each trial which could not fill a bin was discarded to make processing easier. Our data then became of the form (NxKxDxT) where:

Description	Variable	Value
Trial #	N	91
Reach angle	K	8
Neuron	D	97
Length of binned trial	T	T

Table 1: Variable assignments where T is left as a variable since the length of each trial is different.

D. Shifting Augmentation

In order to produce multiple data points from a single stream of data we employed a method of shifted sampling. It has been shown that when analyzing neural data from the primary motor cortex(M1) that spatial resolution is consistent over time, while temporal summation is the major factor in encoding motion **[5].** Shifting preserves the temporal summation encoding as spike count, while disregarding the spatial summation portion of the dataset. Another reasoning for this type of shifting is the arbitrary start of this data when recorded. Even the slightest delay in saving the data would have shifted all the data points by a couple samples. The shifting was employed to sequentially resample the same stream at a 1 millisecond shift. Each set of 20ms binned data was then shifted 19 times which resulted in 19 augmented versions of our dataset.

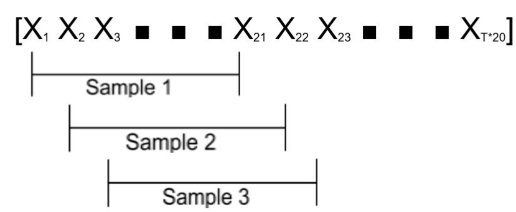

Figure 2: Visualization of the shifting logic where each X is a binned spike count data point.

E. Noise Augmentation

Another approach was to simulate the addition of Gaussian noise to our dataset. The addition of Gaussian noise adds the element of sensor noise, or external measurement noise that in practice could have corrupted the dataset. To simulate the noise a normally distributed random vector was added to the original spike data 19 times to expand the data the same number of times as shifting. The resulting spike count vector could have negative values, as the noise could have removed spikes from already zero spike bins. All negative values were set to zero since negative spike counts do not exist in real life. The result is a set of data expanded by 20x, each with simulated measurement noise.

F. Dimensionality Reduction

The same procedure was done after reducing the number of neurons. In an experiment done by Shenoy, it was shown that there needed to be a minimum of 20 neurons to capture the true nature of the motor cortex **[1].** Trials were done by randomly selecting 20,40,60,80, and 97 (all) neurons to assess if combined with data augmentation, smaller electrode arrays could be used to achieve similar performance as the original set up. The motivation behind this is over time, electrode arrays degrade and could stop recording, and effectively removing the data from a neuron.

G. Training

The Kalman filter is a time-varying dynamic model which was fitted to the data to decode arm state from neural data. It assumes two models, a state and observation model. The state variable Zt represents the 2 dimensional arm state of the monkey over time. The observation variable Xt represents the neural activity of M1 over time. In the assumption of both of these models we select both of them to be modeled as Gaussian distributions.

The arm state is modeled:

$$z = \begin{bmatrix} \textit{horizontal position} \\ \textit{vertical position} \\ \textit{horizontal velocity} \\ \textit{vertical velocity} \end{bmatrix}$$

$$Z_t \sim N(AZ_{t-1}, Q)$$

Z_1 is the initial case, which is modeled as such:

$$Z_1 \sim N(\Pi, V)$$

The Neural Observational is modeled:

$$X_t \sim N(CZ_t, R)$$

The neural observations are modeled using a mean scaled by the transformation matrix C, and covariance matrix R. The variables A,Q,∏,V,C and R are the model parameters that are set in the training phase;

Additionally, in our model we take a Markov assumption that the previous state is fully responsible for the current state. This can be shown through the modeling of the joint state distribution.

$$P(Z_1, ..., Z_T) = P(Z_1)P(Z_2|Z_1)P(Z_3|Z_1, Z_2)P(Z_4|Z_1, Z_2, Z_3) \\ ... P(Z_T|Z_1, ..., Z_{T-1})$$

Through the application of the Markov assumption the probability distribution can be modeled as such:

$$P(Z_1, ..., Z_T) = P(Z_1)P(Z_2|Z_1)P(Z_3|Z_2)\,P(Z_4|Z_3) \\ ... P(Z_T|Z_{T-1})$$

$$P(Z_1, ..., Z_T) = P(Z_1)\prod_{t=2}^{T} P(Z_T|Z_{T-1})$$

To set the parameters to train the model, the joint distribution of the state variables and the observation variables were maximized with respect to the parameters. The likelihood of the joint distribution was modeled as such:

$$L(\theta) = P(\{X\}, \{Z\}|\theta) \\ = P(Z_1)(\prod_{t=2}^{T} P(Z_t|Z_{t-1}))(\prod_{t=1}^{T} P(X_t|Z_t))$$

The log was then taken of this likelihood and from this the respective parameters were calculated by taking the respective partial derivatives. The equations for the parameters are as follow:

$$A = (\sum_{t=2}^{T} Z_t Z_{t-1}^{\top})\,(\sum_{t=2}^{T} Z_{t-1} Z_{t-1}^{\top})^{-1}$$

$$Q = \frac{1}{T-1}\,(\sum_{t=2}^{T} (Z_t - AZ_{t-1})\,(Z_t - AZ_{t-1})^{\top})$$

$$C = (\sum_{t=1}^{T} X_t Z_t^{\top})\,(\sum_{t=1}^{T} Z_t Z_t^{\top})^{-1}$$

$$R = \frac{1}{T}\,(\sum_{t=1}^{T} (X_t - CZ_t)\,(X_t - CZ_t)^{\top})$$

H. Testing

In the testing phase the neural data is used to construct a hypothesis of the location of the arm state. In order to estimate the current arm state based on the previous arm state, at each time step, a one step prediction and measurement update were performed. The one-step prediction models the probability of the arm state given the previous neural state. The measurement update takes the one-step prediction and applies the observational model of the neural state at the current time point. The derivations are shown below:

1) One-Step Prediction:

$$P(Z_t|\{X\}_1^{t-1}) = \int P(Z_t|Z_{t-1})P(Z_{t-1}|\{X\}_1^{t-1})$$

$$\mu_t^{\tau} = E[Z_t|\{X\}_1^{\tau}]$$

$$\Sigma_t^{\tau} = cov(Z_t|\{X\}_1^{\tau})$$

2) Measurement Update:

$$P(Z_t|\{X\}_1^t) = \frac{P(X_t|Z_t)P(Z_t|\{X\}_1^{t-1})}{P(X_t|\{X\}_1^{t-1})}$$

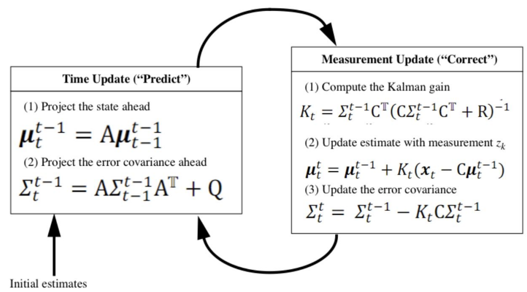

Figure 3: Recursive algorithm for testing

I. *Performance metric*

To quantify the performance of the different augmentations that were done, an L1 norm was used to calculate the error of each model. The single point error increases with time, so to keep things consistent, error calculations were limited to the first 8 bins, which was the minimum amount of bins common to all the testing trials. An average was taken over the first 8 bins and across the 728 trials to get a single error value for each of the 5x7 conditions. The average and variance was calculated across the 5 different seedings and done for each category.

III. RESULTS

Condition	Number of Trials	Augmentation	Total
Original	100	1x	100
	200		200
	300		300
	400		400
	500		500
	600		600
	700		700
Shifted	100	20x	2000
	200		4000
	300		6000
	400		8000
	500		10000
	600		12000
	700		14000
Noise	100	20x	2000
	200		4000
	300		6000
	400		8000
	500		10000
	600		12000
	700		14000

Table 2: Visualization of the data augmentation and number of trials used in all conditions

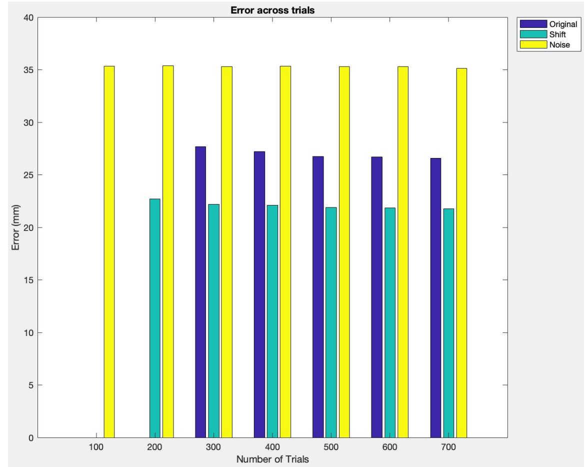

Figure 4: Error across trials, with the different augmentations applied. Missing values at 100 and 200 trials due to training failure, see discussion.

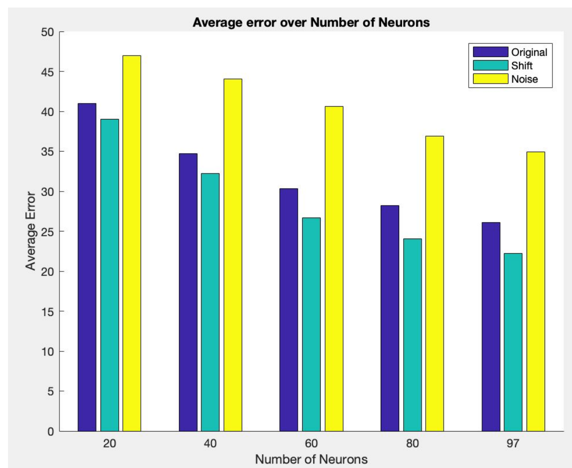

Figure 5: Shows the averaged error across the number of neurons being sampled from

IV. DISCUSSION

For our model, we assumed the system was Gaussian, linear, and Markov for simplicity's sake. This does not accurately model the entire system because of physical limitations of the system. For example, in the state model, the assumption is that the only thing affecting the current arm state is the previous arm state. However this is a flaw due to the edge case where the arm is fully extended in one direction.

The results showed that there is not much variation from 100 trials to 700 trials as seen in figure 4. The missing values for 100 and 200 trials are due to singular matrix errors when calculating the parameters and testing. This could be due to the random seeding choosing trials which had very few spikes, leading to a lot of zeros in the dataset. This would take the matrix close to singular, and provide errors when the inverse is taken. This problem is not seen in the noise augmentation because spikes are being randomly added to the data, resulting in mostly non-zero data points.

In the same figure, it can be seen that the noise is consistently higher than the others, showing that it should not be used. Shifting however, is consistently lower than the original data set, sitting around an error rate of 22 while the original data has an average error of 27 mm. This 5 mm improvement corresponds to 18.5% which is quite substantial in terms of this BCI. The consistent nature of performance across the number of trials is also indicative of the model hitting the plateau and not needing a lot of trials to be able to train the model well.

From figure 5, it can be seen that performance increases with the number of neurons, which is expected. The same trend of performance across data augmentation techniques can be seen as well. Using the shifting data augmentation technique with 60 neurons provides almost identical performance as using all 97 neurons with no data augmentation techniques. Average error of 26.69±0.06 mm compared to 26.07±2.21 mm. This shows that shifted binning is a valid form of data augmentation for neural spike data and can be used to increase performance.

V. Author Contribution

Jason contributed to the code for data augmentation and testing, Sidharth contributed to the code for seeding the trials and training the models. Both worked on writing this paper.

VI. Acknowledgments

We would like to thank Professor Gilja for his instruction in ECE 209 at UCSD and for helping us design this project. We would also like to thank the Shenoy lab at Stanford for providing us with the dataset used in this paper and Professor de Sa from UCSD for her input on dimensionality reduction.

VII. References

[1] Kao, J., Nuyujukian, P., Ryu, S. *et al.* Single-trial dynamics of motor cortex and their applications to brain-machine interfaces. *Nat Commun* 6, 7759 (2015).

[2]: McAndrew, R., & Helms Tillery, S. I. (2016). Laboratory primates: Their lives in and after research. *Temperature (Austin, Tex.)*, *3*(4), 502–508. https://doi.org/10.1080/23328940.2016.1229161

[3]: Shorten, C., Khoshgoftaar, T.M. A survey on Image Data Augmentation for Deep Learning. *J Big Data* 6, 60 (2019). https://doi.org/10.1186/s40537-019-0197-0

[4] Anumula, Jithendar & Neil, Dan & Delbruck, Tobi & Liu, Shih-Chii. (2018). Feature Representations for Neuromorphic Audio Spike Streams. Frontiers in Neuroscience. 12. 10.3389/fnins.2018.00023.

[5]: Taylor, B. A., Fennelly, M. E., Taylor, A., & Farrell, J. (1993). Temporal summation--the key to motor evoked potential spinal cord monitoring in humans. *Journal of neurology, neurosurgery, and psychiatry*, *56*(1), 104–106. https://doi.org/10.1136/jnnp.56.1.104