Functional characterization of neurons in the early visual pathway: An exploration of spike train classification methods

Cássio Mendes Fontes[1], **Ana Luiza Turchetti-Maia** [2], **Jerome Baron** [2], **Antônio Pádua Braga**[1]

[1]Departamento de Eletrônica – Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte – MG – Brazil

[2]Departamento de Fisiologia e Biofísica – Universidade Federal de Minas Gerais (UFMG)
Belo Horizonte – MG – Brazil

cassio.farmaciajr@gmail.com, altmaia@yahoo.com.br

jbaron@icb.ufmg.br,abpbraga@ufmg.br

Abstract. *The understanding of how neurons interact in the visual cortex and what kinds of neurons are responsible for each interaction are still unanswered questions in the field of neurophysiology. To acquire information on such problems, sixteen neurons were mapped with visual stimulus of different sizes, contrasts but with similar patterns on their receptive fields. The data obtained was analyzed with several approaches. PCA was applied as a linear dimensionality reduction technique and the new data was separated with clustering techniques. The acquired data was also represented as the difference of time (moment in time when one spike occured minus the previous one) between spikes. Several attributes were extracted and clustered into different groups and with kernel techniques to visualize the data two large groups of neurons were identified. A support vector machine classifier (SVMC) was implemented in order to verify if an efficient classification could be performed. The main data was also separated as spikes before, during and after direct stimulation using the difference of time between spikes as data. The same attributes were extracted from these matrices and kernel techniques were applied on them to visualize the data. As before, two large groups of data were identified and isolated with clustering techniques which represented two groups of neurons.*

1. Introduction

The receptive field of a visual neuron was defined by [Hartline 1938] as being a region of the retina that when stimulated induced an alteration in the neuron activity. This is known as the *Receptive Field* and is used to characterize neurons in different levels of the visual system.

New evidences have shown that neurons in several visual areas can suffer modulation when stimulated inside and outside of their *Classical Receptive Fields* (CRF). This is known as *center-surround* interaction and was described first by [Kufler 1953]. This means that answers to local attributes are highly influenced by the context in which they are presented.

The usual model to study contextual effects in the visual primary visual cortex is to present a small stimulus centered in the neuron's CRF and increase its size. The usual

cellular response to this stimulation is composed by two parts. The first is that the facilitation of the response increases with the increase of the stimulus size and the second is a suppression after the maximal spacial summation as described by [J.R Cavanaugh 2002].

To further study these neuronal behaviors, specially the contextual effects present in the visual system of birds, owl's wulst visual neurons were mapped and the stimulus was centered in the CRF. Several different conditions of size and contrast were applied. When a spike occurred its moment in time was determined.

With the acquired data for sixteen neurons a linear dimension reduction technique (PCA) was applied followed by a clustering technique which apparently showed two large groups in the original data-set.

Another approach to the problem was the extraction of several features directly from the data. Clustering techniques were applied on them and the results were analyzed with the aid of kernel matrixes. With the isolation of different groups, a *Support Vector Machine* classifier was implemented to verify if a determined neuron could be classified in one of two groups which were identified by the affinity matrix properties.

The acquisition of a new data-set from the original one was also performed. The moment in time which a spike happened minus the moment the previous one did which we will call time interval between spikes, was obtained for every trial condition. From them, several features were extracted and clustering methods were also applied. To visualize the data *Kernel matrices* were used. A classifier was also trained using the data of the clusters to verify if a determined neuron could be classified in one of two groups also seen by the affinity matrixes properties. To further study the neuronal behavior, the data was separated into three different matrices corresponding to spikes before, during and after stimulation. In these new matrices, the time interval analysis was performed and the same features as before were extracted and clustering methods were applied for all three data-sets. *Kernel matrices* were also used to visualize the data.

The result to such analysis was that some neurons presented all experimental conditions only in one of two groups, some others presented most of conditions in one them and others presented part of the conditions in one group and part in a second group. This situation happened for all data-sets analised but the representation of time interval between spikes showed a better class separation specially when analysing the data-sets composed by time intervals before and after stimulation. Which may indicate that the natural spiking behavior of the neurons may be a tool to differentiate them.

Two isolated units were also studied. Clustering and Kernel methods were applied on the data which was represented as time intervalsbetween spikes. A natural separation of high and low contrast occurred with one of the units and when the high and low contrast were isolated and analyzed they also showed a dichotomization depending on the two different conditions. These isolated analysis can indicate that there are two different generator functions for the data which can be isolated unit response and response with contextual interactions with recruiting of neighboring neurons.

2. Materials and Methods

2.1. Data Acquiring Methods and Data Disposition

We used standard extracellular recording techniques in awake burrowing owls [J Baron 2007] to measure the responses of isolated wulst neurons to moving sinusoidal grating of 10 different sizes (0.48, 0.72, 1.08, 2.04, 3, 3.6, 6, 8.99, 11.99 and 17.99 degrees of visual angle) presented at high (97%) and low contrasts (10%). Manual mapping was initially used to determine the location of the receptive field center, and all subsequent measurements were made trough the dominant eye, with gratings optimizaed for direction, spatial and temporal frequencies.

Extracellular potentials were first amplified (x1000), filtered (300 Hz to 7 Khz), digitized (32 kHz), and then stored to disk using custom software in LabVIEW (National Instruments) by Dr. Sergio Neuenschwander. Unit isolation and discrimination was further refined offline with software developed by Dr. Nan-Hui Chen. The animal protocols used in this study were approved by the Ethics Committee for Animal Experimentation (CETEA, license nº 2004/01) of the Federal University of Minas Gerais, and were conducted in conformance with the guidelines established by the National Institutes of Health and the European Communities Council Directive of 24 November 1986 (86/609/EEC).

2.2. Mercer-Kernels

Mercer Kernels allow mapping of input data into feature space. But such mapping is accomplished without truly performing the mapping, only the dot product of input space data is performed [Queiroz 2009]. This is of great use because allows the direct use of Kernel Machines that have a great importance in the supervised learning universe. The Kernel Matrix is considered a Mercer Kernel if and only if it attends the symmetry and semi-definite requirements [Wang 2008]. The Kernel Matrix can be seen as a affinity matrix that captures the similarity between all pairs of points of a data set. There are several ways for constructing the Kernel Matrix. They can be implemented by polynomials, linear, sigmoid and Gaussian functions as can be seen in (1-3), but there are several other ones that can be used to construct the kernel matrix such as laplacian and sigmoid.

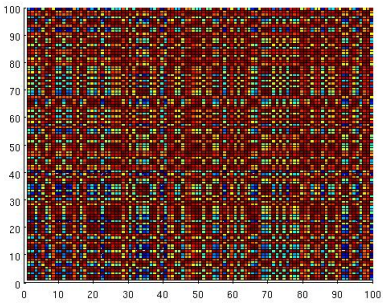
$$K(x_i, x_j) = e^{\frac{(x_i - x_j)^2}{2\sigma}} \tag{1}$$

$$K(x_i, x_j) = x_i x_j^T \tag{2}$$

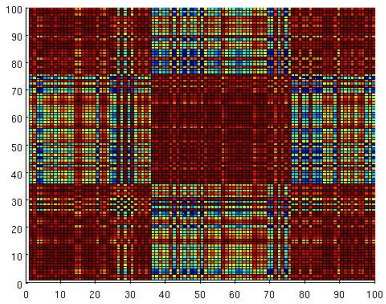
$$K(x_i, x_j) = (x_i x_j^T - 1)^p \tag{3}$$

Kernels contain large quantities of information about the structure of the data. The structural information contained in the Kernel can only be accessed when the data is ordered according to their generator functions. When data is ordered in such a manner, the block-diagonal affinity matrix should appear and reveal the information embedded in the structure of the matrix, as can be seen in Fig. 1(a) and Fig. 1(b)

The present work constructed several Kernel Matrix utilizing linear, quadratic and polynomial functions. These representations of data were used to perform visual analysis



(a) Grafical representation of a kernel without ordination of data according to generator functions



(b) Graphical representation of a kernel with ordination of data according to generator functions

Figure 1. Graphical representation of block-diagonal affinity matrix property

on how the groups were related, specially after the ordination of the data according to the result of the clustering techniques.

2.3. Principal Component Analysis

Principal Component Analysis (PCA) is a way of identifying patterns in data and expressing the data so that the similarities and differences are highlighted. According to [Smith 2002] PCA is a variable reduction procedure and that takes into account that most data acquisition has redundancy.

A principal component can be defined as a linear combination of optimally-weighted observed variables. It is an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data lies on the first coordinate, which is the first principal component.

For a data matrix X^T, with zero mean, where each row represents a different repetition of the experiment and each column gives the results from a particular probe, the PCA transformation is given in (4-5)

$$Y^T = X^T W \tag{4}$$

$$Y^T = V \Sigma W \tag{5}$$

where $V\Sigma$ is the singular value decomposition (svd) od X^T.

Given a set of points in Euclidean space, the first principal component (the eigenvector with the largest eigenvalue) corresponds to a line that passes through the mean and minimizes sum squared error with those points. The second principal component corresponds to the same concept after all correlation with the first principal component has been subtracted out from the points. Each eigenvalue indicates the portion of the variance that is correlated with each eigenvector.

PCA essentially rotates the set of points around their mean in order to align with the first few principal components. This moves as much of the variance as possible (using

a linear transformation) into the first few dimensions. The values in the remaining dimensions, therefore, tend to be highly correlated and may be dropped with minimal loss of information.

2.4. Fuzzy K-Means Clustering

As said in [Peter E. Hart 2007], in the classical K-means procedure, each data point is assumed to be in exactly one cluster as can be seen in (6),

$$P(w_i|X_k,\Theta) = \begin{cases} 1 & \text{if} \quad x \leq 0 \\ 0 & otherwise \end{cases} \tag{6}$$

where the probability $P(w_i|X_k,\Theta)$ is large when the squared Mahalanobis distance $(X_k - \mu_i)^t \sum_{i=1}^{n}(X_k - \mu_i)$ is small. Computing the squared Euclidean distance $||X_k - \mu_i||^2$, find the mean $\mu_m$nearest to X_k and approximate $P(w_i|X_k,\Theta)$. This condition can be relaxed and can be assumed that each sample X_j has some grades membership in a cluster. The fuzzy k-means clustering algorithms seeks a minimum of a heuristic global cost function (7).

$$J_{fuz} = \sum_{i=1}^{c}\sum_{j=1}^{n}[P(w_i|X_k,\Theta)]^b||X_k - \mu_i||^2 \tag{7}$$

Where b is a free parameter chosen to adjust the "blending" of different clusters.

2.5. Support Vector Machines

The *Support Vector Machine* (SVM) is a machine learning model. It is based on the maximization of the separation margin between the groups [Chen and Lin 2003]. Hyperplanes are built in such a manner as to maximize the separation margin and also minimize the training error [Shawe-Taylor 2000] In order to meet such a task, finding the hyperplane, an optimization approach must be applied and the dual problem of the maximization of the separation margin must be solved.

To a binary classification problem, where (x_i, y_i) is the input-output pair and $y_i \in +1, -1$, the dual optimization problem is given by:

Maximize:

$$W(\alpha) = \sum_{i=1}^{N}\alpha_i - \frac{1}{2_i}\sum_{j=1}^{N} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \tag{8}$$

Limited by:

$$\sum_{i=1}^{N} y_i\alpha_i = 0; \forall_{i=1}^{N} : 0 \leq \alpha_i \leq C \tag{9}$$

where $W(\alpha)$ is the weight vector, the C parameter is determined by the user. It sets on the importance of the maximization margin in comparison with a misclassification during the determination of the separation hyperplane. In the present problem for

comparison porpoises the C parameter was determined as one. The α_i is the *La Grange multiplier* and $K(x_i, x_j)$ is the utilized *kernel*.

The optimized weight (w^*) vector is given by (8):

$$W^* = \sum_{i=1}^{N} y_i \alpha_i^* x_i \tag{10}$$

The parameter b* is given by (9):

$$\begin{aligned} b^* \quad = \quad & -\frac{1}{2}\left[\underbrace{max}_{i|y_i=-1}\left(\sum_{i=1}^{N_{SV}} y_i \alpha_j K(x_i, y_j\right)\right] \\ & -\frac{1}{2}\left[\underbrace{min}_{i|y_i=+1}\left(\sum_{i=1}^{N_{SV}} y_i \alpha_j K(x_i, y_j\right)\right] \end{aligned} \tag{11}$$

The decision function is given by (10):

$$f(x) = sgn\left(\sum_{i=1}^{N_{SV}} y_i \alpha_i^* K(x_i, x) + b^*\right) \tag{12}$$

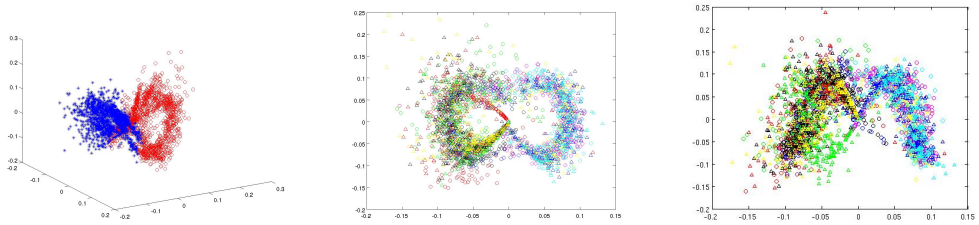
This means that if $f(x) \geq 0$, x belongs to the $+1$ class otherwise x belongs to the -1 class. N_{SV} is the number of support vectors. The SVM applies directly to linearly separable problems. To separate problems that are not linear, the entry data has to be nonlinearly mapped to a feature space where the data is linearly separable.

3. Results and Discussion

Several approaches were tested to explore the data at hand. The first was the use of PCA on the raw data even before representation of time intervals between spikes as can be seen in Fig. 3. The first three principal components presented eigenvalues much higher than all the other ones, so they were selected and plotted threedimensionaly as can be seen in Fig. 2(a)

Apparently two different groups can be identified in the data. This becomes clearer when the data is plotted bidimensionaly as can be seen in Fig. 2(b) which are the first and second principal components and also in Fig. 2(c) which are the first and third principal components. In both figures, each color and symbol represents the data of a different neuron.

Using a clustering algorithm (FCM) the groups are separated according to Fig. 2(a). The blue points belong to one group and the red points belong to another. Although was expected that the separation occurred in such a manner that one group would have one set of neurons and the other group would have another set, the result of the FCM pointed that some conditions of one neuron are in one group and other conditions are in one other group.



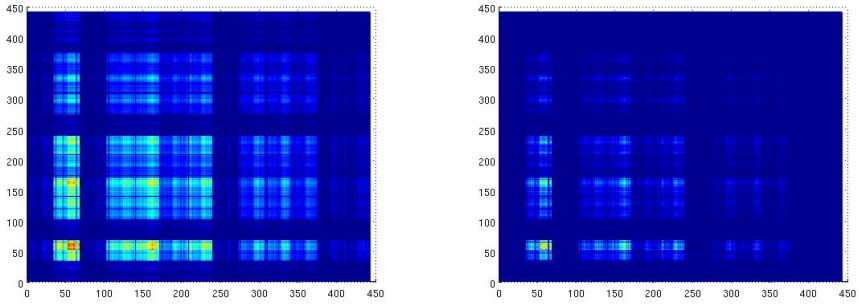
(a) Graphical representation of the three most relevant principal components

(b) Graphical representation of the first and second principal components

(c) Graphical representation of the first and third most relevant principal components

Figure 2. Graphical representation of the most relevant principal components

With the need of further data exploration to better understand the information embedded in the data, after the values inside each cell were transformed to seconds, several attributes were extracted directly from the data matrix. The number of spikes per condition, mean, stat-deviation, variance, maximum, minimum, skewness and kurtosis of times per condition.



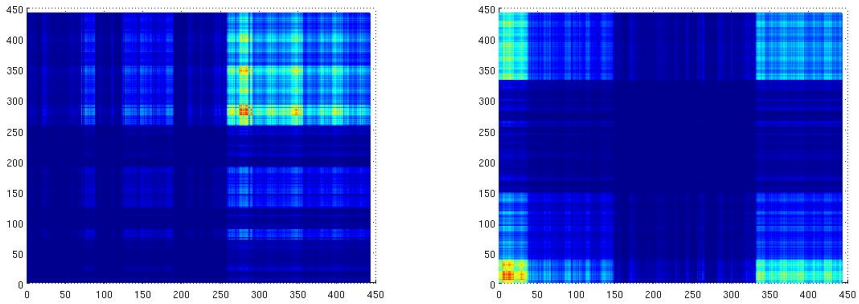
(a) Linear kernel matrix representation of extracted attributes

(b) Quadratic kernel matrix representation of extracted attributes

Figure 3. Kernel matrix representation of extracted attributes

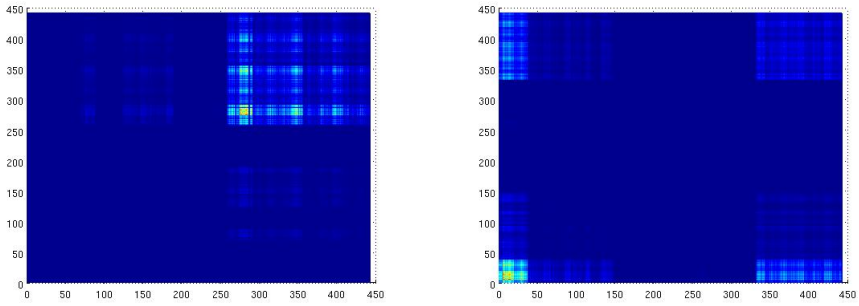
The new data representation was then used to identify possible pattern distributions. A linear and a polynomial kernel matrices were built with these new representations, as can be seen in Fig 3(a) and Fig 3(b). Cluster techniques (FCM) were applied on the data for two and four groups. The data was ordered according to the new groups. It is important to highlight that the data was ordered according to experimental conditions since the data for all neurons was combined before the clustering algorithm was applied on the data-set. For two clusters ordination the groups showed a small relation with each other for a linear mapping as can be seen in Fig. 4(a) and for the four groups ordination, the three main groups were identified as can be seen in Fig. 4(b). The polynomial kernels for two and four groups showed a better separation for the different clusters as can be seen in 5(a) and 5(b).

To continue the data exploratory analysis, the original data matrix was manipulated, and a new matrix with the time interval between spikes was created. With this new representation, several attributes such as mean, stat-deviation, skewness, kurtosis, maximum and minimum of the time intervals and the number of spikes were extracted



(a) Linear kernel matrix representation of 2 groups ordination (b) Linear kernel matrix representation of 4 groups ordination

Figure 4. Linear kernel matrix representation of different groups ordinations



(a) Polynomial kernel matrix representation of 2 groups ordination (b) Polynomial kernel matrix representation of 4 groups ordination

Figure 5. Polynomial kernel matrix representation of different groups ordinations

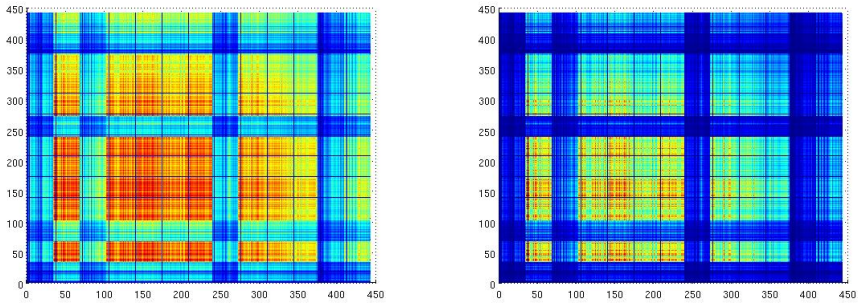
[Guyon and Elisseeff 2003]. A linear and a polynomial kernel matrices were built as can be seen in Fig. 6(a) and in Fig. 6(b).

A clustering method (FCM) was applied on the data-set, for two and four groups. The results of the polynomial mapping can be seen in Fig. 7(a) and in Fig. 7(b). This result shows that there maybe two great groups in the data. Again, the data was not separated by neuron and a large matrix of experimental conditions was built to apply the clustering method.

This result can be taken as a good evidence that there are two large groups in the data set. To verify which elements composes the two groups, the data was separated and ordered. This showed an interesting result since the separation occurred naturally by neuron. The neurons 01,03,08,12,13 and 15 belong to the first group and the neurons 02,04,05,07 and 16 belong to the second group . The neurons 6,9,10,11 and 14, had part of the data grouped in the first group and the other half in the second one (according to the results of the FCM).

These results come to show that there maybe be two different sets of neurons which present different behaviors in such a manner that they can be grouped in separate classes.

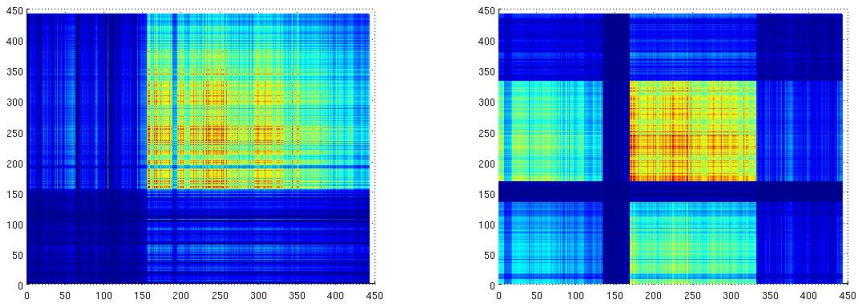
With this information, a linear and a polynomial SVM classifier were imple-



(a) Linear kernel representation of time interval between spikes

(b) Polynomial kernel representation of time interval between spikes

Figure 6. Kernel representations of time interval between spikes



(a) Polynomial kernel representation of two groups separation for the time interval representation

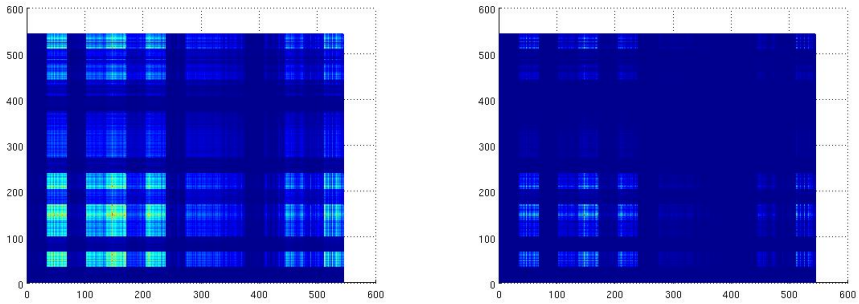
(b) Polynomial kernel representation of four groups separation for the time interval representation

Figure 7. Polynomial kernel representation different groups separation for the time interval representation

mented. Random data was taken from the data-set. For training 70% of the data was used and 30% was left for testing. This was performed ten times and the accuracy and sensibility for the classifications were averaged. For training and testing the linear classifier presented good performances with accuracy around 93% and sensibility around 90%. The polynomial classifier presented an average 98% accuracy and an average 96% sensibility.

To further explore the problem, the spikes referring to the period before stimulation, during stimulation and after stimulation were separated into three different data-sets. The time interval between spikes was obtained and on this new representation the analysis were performed. The same features as before were extracted from each of the data sets. Linear and polynomial Kernel matrices was built for each representation as can be seen on Fig. 8(a) and Fig. 8(b) which are representations for the post stimulation data matrix. As before FCM was applied to the data matrix to cluster the data into two and four groups. After ordering the groups linear and polynomial kernels were also constructed as can be seen in Fig. 9(a), Fig. 9(b) and Fig 10 which are the respective kernels

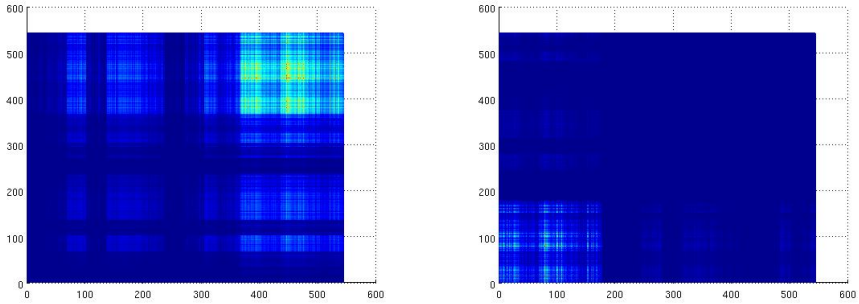
Comparing the kernel representation of two and four clusters, can be seen that there are in fact two large groups of data, which point to two generative functions. With



(a) Linear kernel representation of post stimulation attribute data matrix

(b) Polynomial kernel representation of post stimulation attribute data matrix

Figure 8. Kernel representations of post stimulation attribute data matrix



(a) Linear kernel representation of post stimulation attribute data matrix ordered according to two cluster separation

(b) Polynomial kernel representation of post stimulation attribute data matrix ordered according to two cluster separation

Figure 9. Kernel representations of post stimulation attribute data matrix ordered according to two cluster separation

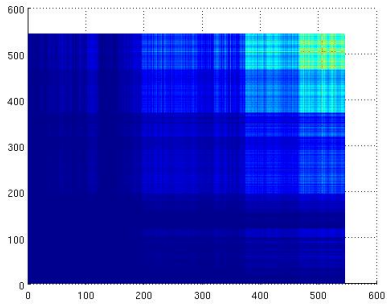


Figure 10. Linear kernel representation of post stimulation attribute data matrix ordered according to four cluster separation

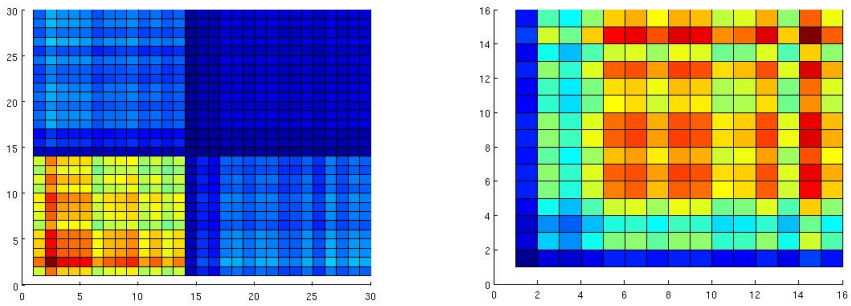
this result, analyzing the data in the two clusters for the three data sets interesting results were seen. The pre and post stimulation data-sets when separated by conditions, showed a natural separation of neurons which some were in the first and others in the second group, but this separation was better than the ones seen before. The neurons in group 1 were 01,03,06,07,08,09,10,11,12,13 and 15 and the ones in group two were 02,04,05 and 07.

Neurons 14 and 16 showed conditions in bought groups. The stimulation data-set when grouped did not separate as well as the pre and post stimulation. The neurons in the first group were 01,03,08,12,13,14,15 and all other neurons had conditions in group one and group two. These results show that two groups of neurons can be identified and isolated using the natural spiking behavior of the neuron.

Analysis on two isolated neurons were also performed. These neurons presented thirty stimulation conditions, 15 of low contrast and 15 of high contrast. The data of low and high contrast for each experiment condition was analyzed together and separated. The features extracted before were also extracted for these two neurons and FCM for two and four clusters was applied on the attribute matrices. One unit when grouped into two clusters with all thirty conditions showed a natural separation of high and low contrast as can be seen in Fig. 11(a).

When the high and low contrast conditions are separated, following the same methodology as before, two groups can be identified in bought high and low contrasts groups as can be seen in Fig. 11(b)

These results show that different neurons may have different responses to high and low contrast levels in such a manner that in some neurons contrast may not cause a significant change in response. Can also be seen that there may be two different groups of responses to stimulation which can be those by a single unit and with the influence of other units in the sense of contextual interaction.



(a) Kernel representation of isolated neuron which presented two groups high and low contrast

(b) Kernel representation of isolated neuron with low contrast condition after ordination according to FCM result

Figure 11. Kernel representation of isolated neurons

4. Conclusion

The current work mainly focused in data exploration. Several approaches were attempted. Dimension reduction techniques showed a path that can be further explored specially if the data is labeled. This technique can be used to map the entry data and train a neural network. This can be of great use, if a neuron classification task is necessary.

Other point which was found in the present work was that depending on how the data was represented and on which features were extracted, if a new data-set is composed with all neurons and two groups are separated according to the responses to a determined stimulus, when the data is ordered, one set of neurons naturally forms one

group of neurons and another set of neurons forms another group, specially if only the information before and after stimulation is taken into account. With this information the data may be labeled according to a classifier that was obtained with the current data sets [Haykin 1994].

Another point which is of high relevance is that some neurons may present responses that are influenced by contrast and other may not present such characteristic. Besides that, was also seen that there may be two sets of responses,to the stimulus when only high or low contrast responses are analyzed. These two sets of responses may be responses of one single cell and responses of contextual interaction, when other cells of the neuron neighborhood are recruited to form the final response of the neuron.

A future work with this data set may oversee the use of two regression models to approximate the functions that correlate the number of spikes in a determined time interval with the stimulus size and contrast. This can be inferred, since there are two families of neurons with different behaviors, perhaps two regression models can be more adequate. Also a deeper analysis of all units isolated to better understand if some neurons really respond better to change of contrast or not and if there are really two types of responses the one of the single cell and the one with recruited neurons along with the original cell.

5. Acknowledgments

We would like to thank Sergio Nerenschwander for providing the SPASS data acquisition software and Nan-hui Chen for allowing us to use his spike-sorting software. The experimental part of this work was supported by the Research Support Foundation of the State of Minas Gerais (FAPEMIG) and the FINEP research grant ”Rede Instituto Brasileiro de Neurociência” (IBN-Net) # 01.06.0842-00; A.L.T.M received a graduate scholarship from the Brazilian National Council for Scientific and Technological Development (CNPq).

References

Chen, Y. and Lin, C. (2003). Combining svms with various feature selection strategies. Technical report.

Guyon, I. and Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182.

Hartline, H. (1938). The response of single optic nerve fibers of the vertebrate eye to illumination of the retina. *American Journal of Physiology*.

Haykin, S. (1994). *Neural Networks: A Comprehensive Foundation*. Macmillan, New York.

J Baron, L. P. M. O. D. B. L. S. . N. (2007). Directional responses of visual wulst neurons to grating and plaid patterns in the awake owl. *European Journal of Neuroscience*.

J.R Cavanaugh, W. B. J. M. (2002). Nature and interaction of signals from the receptive field center and surround in macaque v1 neurons. *Journal of Neurophysiology*.

Kufler, S. (1953). Discharge patterns and functional organization of mammalian. *Journal of Neurophysiology*.

Peter E. Hart, R. O. D. D. G. S. (2007). *Pattern Classification*, chapter 9, pages 528–530. WILEY, second edition.

Queiroz, F. . B. A. P. . P. W. (2009). Sorted kernel matrices as cluster validity indexes. *2009 IFSA World Congress*.

Shawe-Taylor, N. C. J. (2000). *An Introduction to Support Vector Machines and other kernel-based learning methods*, chapter 6, pages 93–122. Cambridge, first edition.

Smith, L. (2002). A tutorial on principal components analysis.

Wang, W. Z. D. Z. K. (2008). On kernel difference-weighted k-nearest neighbor classification. *Pattern Analysis & Applications*.