Discovery and Characterization of Neural Circuitry from Behavior, Connectivity Patterns and Activity Patterns

Carey E. Priebe  
Johns Hopkins University  
Department of Applied Mathematics and Statistics  
Baltimore 21218-2682 USA  
ceph@jhu.edu

1. Project Summary

A recent study [1] has generated a neuron-like behavior atlas of the Drosophila larval nervous system. This atlas is a starting point for connectivity- and activity mapping studies to further investigate the mechanisms by which neurons mediate diverse behaviors. Drosophila larvae, with 10,000 neurons, offer an opportunity to determine how an entire nervous system generates behavior. The current atlas is a significant step toward achieving this goal — Timothy O’Leary & Eve Marder [2] write that it will “usher in a new era of integrated methods for deciphering how an entire nervous system generates behavior” and that it “has achieved a technical, multidisciplinary tour de force that will provide a rich source of research questions” — but it needs to be refined to single neuron resolution. Marta Ziatlic and Jim Truman at Janelia Farm have refined the library of neuronal lines used in the current publication, driving experimentation in vivo and in parallel in the larval nervous system. Optogenetically activation of this new collection will allow the application of multiscale unsupervised structure learning to the behavior data and yield a single neuron behavior map covering all neuron types in the larval nervous system. In parallel, Albert Cardona at Janelia Farm is using electron microscopy (EM) to map a wiring diagram of the larval nervous system. Marta Ziatlic is also generating a neuron activity map of the entire Drosophila larval nervous system using calcium imaging, as described for zebra fish in Ahrens et al. [3]. The objective of this project is to develop principled statistical pattern recognition & machine learning methods for clustering neurons based on these three different data sets, each individually and jointly. The extent to which clusters obtained from the three data sets agree, and the manner in which they disagree, will provide a characterization of neural circuitry from behavior, connectivity patterns and activity patterns.

2. Project Description

We propose fusion and inference from multiple disparate data sources for discovery and characterization of neural circuitry from behavior, connectivity patterns and activity patterns. This interdisciplinary project will consist of my team from Johns Hopkins University Department of Applied Mathematics and Statistics working closely with neuroscientists from HiMiris Janelia Farm Research Campus to develop principled statistical pattern recognition and machine learning methods for clustering neurons based on three different data sets extracted from the Drosophila larval nervous system: calcium imaging neural activity data (A for Activity), optogenetic behavior data (B for Behavior), and an electron microscopy (C for Connectome). Drosophila larvae, with n = 10,000 neurons, offer an opportunity to determine how an entire nervous system generates behavior. The clustering results will provide the basis for determining how the structure of neural circuits relates to their function.

Let A be an n × n functional connectivity matrix for the calcium imaging neural activity data; let B be an n × n similarity matrix for the optogenetic behavior data; let C be an n × n connectivity representing the EM wiring diagram, with neurons as vertices [a] = [1, 2, . . ., n]. Given A, B, and C, our task is to develop advanced methods, in conjunction with the Janelia Farm neuroscientists, for clustering neurons based on these three different data sets, both individually and jointly. We will know (almost exactly) which neuron v in the optogenetic behavior data maps to which neuron v in the EM connectome; however, there will be more ambiguity in the mapping of [a] in [v] and [v] in [a] for the associated calcium imaging neural activity data. Methodologically, we will first aspire to determine that these maps are known; then, based on our assessment of the performance of our algorithms for resolving the ambiguity, we will adapt our cross-modality cluster analysis methodology to account for the exact ambiguity.

Step 1: We must participate in the three experimental data collections, with our colleagues at Janelia Farm. First, the data collection itself is proceeding with our input regarding experimental design. Once the data are collected, data processing must ensure that the three data objects — Activity A, Behavior B, and Connectome C — are commensurate (in particular, we require full knowledge of the cross-modality neuron identification maps) and appropriate for subsequent cluster analysis.

Step 2: From each of A, B, and C, we will generate a clustering — a partition of [n], which we will denote P_A, P_B, and P_C, respectively. (There is much hidden in the succinct “we will generate a clustering.” Each of the three objects are fundamentally different, and thus we require three separate clustering methodologies. Furthermore, as we do not know the “correct” number of clusters, this inherently tricky model selection issue must be addressed — thrice. Finally, to facilitate our subsequent analysis, it will be preferable if we have the three clustering and model selection methodologies be as similar to one another as possible, so that when we compare clusterings we are comparing like clusterings. All of these issues will be formally addressed in the context of the Janelia Farm neuroscientists.) A method for comparing partitions, such as the pairwise Adjusted Rand Index — ARI(p,q) for p ≠ q ∈ {A, B, C} — can then be used to assess the similarity of the three clusterings. For example, in the unlikely case that these three ARIs are all equal to one, we conclude that all three data modalities yield identical clusterings. At the other extreme, in the unlikely case that the pairwise ARIs are all equal to zero, we conclude that all three data modalities yield clusterings no more similar to one another than would be expected by chance. The extent to which ARI(p,q) is larger than zero yet smaller than one will be the starting point for our analysis of the similarities and differences of the three neuronal clusterings.

Step 3: Here we will consider a joint analysis of the three data objects. We begin by constructing a 3nn × 3nn omnibus matrix M. The three n × n diagonal blocks of M are given by our three data matrices; that is, M[1, 1 : n] = A, M[1, n + 1 : 2n] = B, and M[2n + 1 : 3n, n + 1 : 2n] = C. The remaining n × n off-diagonal blocks of M are all set equal to the n × n identity matrix I[n], which captures the information that we know the nth neuron in A matches the nth neuron in B matches the nth neuron in C. (These off-diagonal blocks of M will have to be altered to account for ambiguity in our cross-modality neuron identification.) From this omnibus matrix, a SIMACOF algorithm for multidimensional scaling will be used to generate an Euclidean embedding. That is, we will map M[ ] → X, where the first n × n matrix X represents each neuron in each modality as a point in R^n such that the original data similarities and the neuron matching knowledge are simultaneously respected to the extent possible. The choice of the embedding dimension d is another issue of model selection; principled approaches exist.) The significance of the geometry of this embedding is that when all three embeddings are formally constructed for a single animal from each of the calcium imaging neural activity data, the optogenetic behavior data, and the EM wiring diagram — lie close together, then this neurons relationship to the entire collection of neurons is similar across all three modalities. Consider, for example, a single neuron, representative of modality, yielding a partition [P_A] of [n]. Then, using this P_A together with the individual neurons’ three modalities we again generate three partitions [P_B] of [n], denoted P_A, P_B, and P_C. If, for each neuron, the three embedded points lie close together, then these three clusterings will be similar and ARI(p,q) for p ≠ q ∈ {A, B, C} will be close to one. If, on the other hand, the optimization is unable to respect the neuron matchings, because the structure of the three similarity matrices are sufficiently different, then these ARIs values will be closer to zero. Furthermore, the relative geometry of the three modality-specific embeddings will tell us how and why the clusterings differ, providing the basis for a characterization of neural circuitry from behavior, connectivity patterns and activity patterns. For example, we might find that even though a specific collection of neurons cluster together in both P_A and P_B, that is, based on the embeddings of the calcium imaging neural activity data and the optogenetic behavior data — the geometry of the embedding of the EM wiring diagram does not respect this structure and P_C splits the neurons in this collection into multiple clusters. Subsequently, we would be able to investigate, through the geometry of the embedding of the connectome C for this specific collection of neurons, what is in the structure of the EM wiring diagram that precludes clustering consistency across modalities.

References

[1]  
[2]  
[3]  
[4]  
[5]  
[6]  

Figure 2: Artists rendition of the embedding M → X of the 3nn × 3nn omnibus matrix M to X ∈ R^{3nn}. This figure illustrates the example wherein the specific collection of neurons [1, 2, 3] cluster together based on the embeddings of the calcium imaging neural activity data A and the optogenetic behavior data B, but the geometry of the embedding of the connectome C does not respect this structure and splits neuron 1 into a separate cluster. Subsequently, we would investigate, through the omnibus matrix and the geometry of the embedding, what it is in the structure of the connectome that precludes clustering consistency across modalities.