Toward statistically valid population decoding models

Péter András[*], Stefano Panzeri, Malcolm P. Young
Neural Systems Group
Department of Psychology
University of Newcastle
NE1 7RU, United Kingdom

SUMMARY

Objective

Although many data-analysis methods have been devised to decode the information from the spike trains of neuronal populations, the problem of building statistically valid stimulus-reconstruction models that provide robust performance has not yet been resolved. In this paper, we present a new method for this problem.

Context

Neurophysiological (e.g., [8,10]) and theoretical (e.g., [3]) investigations indicate that the central nervous system transmits information at the level of neuronal populations. Recent analysis (e.g., [2]) shows that an efficient decoding method for such population codes is the one that uses Bayes' rule. The use of Bayes' rule essentially means that we calculate the posterior probability of the stimuli $P(s \mid \mathbf{r}) = \dfrac{P(\mathbf{r} \mid s) \cdot P(s)}{P(\mathbf{r})}$, where s is the stimulus and **r** is the population response, and we select the most likely stimulus as the decoded or estimated stimulus.

However, Bayesian decoding mechanisms typically have to be applied to relatively small amounts of experimental neurophysiological data (e.g., [9]). In most cases, data are available from single or few unit recordings, and the population response is constructed by methods, such as the pseudo-simultaneous population vector [9], which provide a small sample of the population response space. Further, the performance of the decoding may strongly depend on the validity of the conditional distributions assumed by the

[*] Corresponding author.

Preferred presentation: ORAL

model. Hence the statistical validity of the decoding may be questionable, and it is in general very difficult to validate it statistically.

Results

We propose the formalization of population decoding models as Bayesian networks [4]. Furthermore, we apply Monte Carlo Gibbs sampling [4] of the data space to obtain a large enough sample for statistical validation of the model. Such sampling may be prohibitive if the number of variables is large, as the number of samples necessary increases exponentially with the dimensionality of the data space. To overcome this problem, we propose to use scrambled quasi-Monte Carlo [6] sampling if the number of variables is moderately large. The main advantage of scrambled quasi-Monte Carlo sampling is that it replaces random sampling, which converges to uniform sampling in the long run, by a much smaller randomised deterministic sample with a distribution that is very close to the true uniform distribution.

To exemplify our proposal we applied it to a particular case of stimulus-category coding. We used data acquired from primary visual cortex of cats [7]. The stimuli presented to the animals fell into four categories. Recordings were made simultaneously by two electrodes from two cortical multi-unit sites. We wanted to estimate the category of the presented stimulus on the base of a single observation of the neuronal population responses. To quantify the neural responses, we first used information analysis of the spike trains to find the most informative 100ms period during the presentation of the stimuli [7]. Next, we classified the responses of each recorded multi-units into 4 response classes, giving in total of 16 response classes at the level of the multi-unit pairs.

To study how stimulus reconstruction accuracy scaled with population size, we then built a Bayesian network by constructing the $P(\mathbf{r} \mid s)$ conditional probability table for the neural multi-unit pairs considered (between 5 and 80 in total). To calculate the estimate of the stimulus, we inject the evidence of population response $\mathbf{r}$ into our Bayesian network and we calculate the most likely stimulus. These calculations are formalized as follows:

$$\widetilde{p}(s_i) = p(s_i) \cdot \prod_{k=1}^{n} p(r^k \mid s_i) \tag{1}$$

$$\hat{p}(s_i) = \frac{\widetilde{p}(s_i)}{\sum_{j=1}^{m} \widetilde{p}(s_j)} \tag{2}$$

$$s_{est} = \arg\max_{s_i} \hat{p}(s_i) \tag{3}$$

where $\mathbf{r} = (r^1, \ldots, r^n)$ is the actual population response vector, s_j, $j = 1, \ldots, m$ are the stimuli, and s_{est} is the estimated stimulus provided by the model (the 'arg max' notation means that we take the argumentum, i.e., s_i, of the maximal value, i.e., $\hat{p}(s_i)$). To simplify the computations, in practice we calculate s_{est} as:

$$s_{est} = \arg\max_{s_i} \left(\log p(s_i) + \sum_{k=1}^{n} \log p(r^k \mid s_i) \right) \tag{4}$$

The statistical validation of the model was undertaken by constructing scrambled quasi–Monte Carlo samples of the data space. We used the method proposed by Owen [6] to scramble quasi–random Niederreiter vectors [5] and to obtain the quasi–Monte Carlo samples.

We compared the decoding results obtained with our Bayesian network method to the results obtained both with the method of pseudo-simultaneous vectors [9], and to a model corresponding to randomised data (i.e., we eliminated the stimulus class information by randomly reassigning the stimulus classes to the recorded responses). The Bayesian network model was able to obtain 100% correct stimulus class estimates by considering a random combination of at least 20 multi-unit pairs. In contrast, the model generated by applying the pseudo-simultaneous vector method was unable to obtain 100% correct estimation. We analysed three versions of this latter method: fitting a Gaussian, or Poisson, distribution to the conditional response probability tables, and using the table directly. In all three cases, the model's performance converged to chance performance, i.e., 25% correct, for large numbers of multi-unit pairs. The model generated by using the randomised data was unable to provide a good estimate, its performance converging to the chance performance for large numbers of considered multi-unit pairs. These results suggest that the Bayesian network method is capable of generating a valid model of the data, whilst previously proposed methods are not.

Significance

The main advantage of our method is that it generates a large validation data set, which samples uniformly the whole data space. This allows us to use the available data efficiently to build and validate the model. Another significant advantage is that our method does not impose additional structural constraints on the response probabilities by fitting any analytical distribution to it. This makes the read-out mechanism proposed here particularly efficient when the neuronal responses being analysed do not fit any simple distribution.

[1] deCharms RC, Zador A (2000). *Ann. Rev. Neurosci.*, 23: 613

[2] Deneve S, Latham PE, Pouget A (1999). *Nature Neurosci.*, 2: 740

[3] Lehky SR, Sejnowski TJ (1999). *Neural Computation*, 11: 261

[4] Liu JM, Desmarais MC (1997). *IEEE Transactions on Knowledge and Data Engineering,* 9: 990

[5] Niederreiter H and Shiue PJ-S (Eds.) (1995). Monte Carlo and Quasi-Monte Carlo Methods in Scientific Computing. New York: Springer-Verlag.

[6] Owen AB (1998). Scrambling Sobol' and Niederreiter-Xing points. *J. Complexity*, 4: 466

[7] Panzeri S, Golledge H, Zheng F, Tovee M, Young MP (2001) *Visual Cognition*, in press

[8] Petersen RS, Diamond ME (2000). *J. Neurosci.*, 20: 6135

[9] Robertson RG, Rolls TE, George-Francois P, and Panzeri S (1999) *Hippocampus*, 9: 206

[10] Young MP, Yamane S (1992) *Science* 256: 1227