M

Metric Space Analysis of Neural Information Flow

Conor Houghton
Department of Computer Science, University of Bristol, Bristol, UK

Definition

In the metric-space approach to spike train analysis, spike trains are regarded as points in a metric space, that is, a space with a distance defined between any pair of points.

Detailed Description

Spike Train Metrics

The Victor-Purpura metric (Victor and Purpura 1996) is one common way to define a metric distance between spike trains. The Victor-Purpura metric considers a fictional total cost for transforming one spike train into another using three basic operations: spike insertion, spike deletion, and spike movement. Each basic operation is given an individual cost, one for inserting or deleting a spike and $q|\delta t|$ for moving a spike a temporal distance δt. The cost-per-time, q, is a parameter, with the timescale $2/q$ thought of as corresponding to the temporal precision of spike times in the metric. The distance between the two spike trains is then defined as the cost of the cheapest transformation of one to the other. In other words, the distance between spike trains x and y is given by

$$d(x, \ y) = \min_{\gamma \in \Gamma} c(\gamma) \qquad (1)$$

where γ is a sequence of basic operations transforming x to y, Γ is the set of all such sequences, and $c(\gamma)$ is the cost of γ, that is, the sum of the individual costs of the basic operations that make up γ. This is not the only spike-train metric: there is also the van Rossum metric (van Rossum 2001) which is part of a kernelization approach to spike trains (Paiva et al. 2009), and there are more recent metrics like the synapse metric (Houghton 2009) and the ISI-distance (Kreuz et al. 2007). These are reviewed and compared in, for example, Victor (2005), Houghton and Victor (2012), and Houghton and Kreuz (2013). There are also multineuron metrics (Aronov et al. 2003; Houghton and Sen 2008), which allow metric-based methods to be applied to multiunit data.

In a metric-space approach to analyzing a collection of spike trains, the data is reduced from a set of complicated objects, spike trains, to a much simpler object, the matrix of distances between the trains. When applied to neural information, the information contained in this matrix is used to estimate information theory quantities for the corresponding spike train data; typically the quantity of interest is the mutual information between a sensory stimuli and a spiking response. Of course,

D. Jaeger, R. Jung (eds.), *Encyclopedia of Computational Neuroscience*,
https://doi.org/10.1007/978-1-4614-7320-6_744-2

this reduction of the data to the distance matrix is a simplification and it is likely that the metric will not be sensitive to all the information-carry features of the spike trains. This means that the information contained in the distance matrix will underestimate the true mutual information. However, it is hoped that metric-based methods will be less prone to bias on small data sets than traditional approaches like the direct method described in Strong et al. (1998). There are sophisticated bias-reduction techniques developed for the direct method (Nemenman et al. 2007; Paninski 2003; Panzeri et al. 2007), and it is unclear as yet whether bias is a substantial enough problem to warrant consideration of metric-based methods.

The Transmitted Information

The transmitted information applies to situations when the spiking data represents multi-trial responses to a set of distinct data. In this case, the spike trains can be clustered by stimulus and the transmitted information estimates the mutual information between this "true clustering" and metric-based clustering. The transmitted information is written in terms of a matching or confusion matrix N. If the stimuli are labeled by $s = 1, \ldots, S$ and the responses are grouped into clusters labeled $c = 1, \ldots, C$, then N is an $S \times C$ matrix whose element n_{sc} is the number of responses to stimulus s placed in cluster c. This means that the probability that a response to stimulus s is placed in cluster c is estimated as $p(s, c) \approx n_{sc}/n$, where $n = \sum_{s=1}^{S}\sum_{c=1}^{C} n_{sc}$ is the total number of responses. The transmitted information, h, measured in nats, is derived by substituting this into the usual definition of the mutual information:

$$h = \frac{1}{n}\left(\sum_{s=1}^{S}\sum_{c=1}^{C} n_{sc}\ln n_{rs} - \sum_{s=1}^{S} n_s \ln n_s - \sum_{c=1}^{C} n_c \ln n_c - n \ln n\right) \tag{2}$$

where, in an abuse of notation, $n_c = \sum_{s=1}^{S} n_{sc}$ and $n_s = \sum_{c=1}^{C} n_{cs}$ denote the row and column sums.

Although the transmitted information is straightforward to calculate once the data have been clustered, it does rely on a clustering algorithm as well as a metric, and, in practice, the transmitted information will substantially underestimate the mutual information. In fact, the confusion matrix is often calculated by clustering the data by stimulus, removing individual data points one-by-one and deciding which cluster the test point is closest to. This is not an unsupervised clustering, and this transmitted information does not measure the actual mutual information, it is used as a comparative quantity, for example, it was originally introduced in Victor and Purpura (1996) as a way to determine the optimal value of the parameter q in the Victor-Purpura metric.

A Kozachenko-Leonenko Estimate

In Kraskov et al. (2004) the mutual information is estimated using the Kozachenko-Leonenko k-nearest-neighbors estimator (Kozachenko and Leonenko 1987). This approach is developed in the case where the stimuli are also points in a metric space, rather than points in a discrete case, as was considered for the transmitted information above. The Kozachenko-Leonenko estimator requires the calculation of the rate of change of the probability of finding a data point as a sphere increases; in Kraskov et al. (2004) this rate of change is calculated using the coordinate-based measure and it is assumed the stimulus and response spaces both have a well-defined dimension. These assumptions do not hold in the case of the space of spike trains. However, different ks are used in the stimulus and response spaces. This is done to allow biases in the entropies $H(R)$, $H(S)$, and $H(R, S)$ to cancel, but, fortuitously, this is done in such a way that all the terms that depend on the dimension of the two spaces, or on the volumes of spheres in those spaces, also cancel. This leads to a formula which can be applied to spike trains; in fact, it is possible to derive a Kozachenko-Leonenko estimator for the mutual information directly for metric spaces (Houghton 2015).

Let R be the random variable corresponding to the spiking response and S the random variable corresponding to the stimulus. S is now a continuous random variable and could correspond to a continuous stimulus, such as the position of a foraging rat, or to another spike train. It is assumed that there are metrics on both spaces;

this in turn induces a metric on the space of stimulus–response pairs:

$$d[(s_1,r_1),(s_2,r_2)] = \max\left[d(s_1,s_2),d(r_1,r_2)\right] \quad (3)$$

Now, for a given set of experiments the data consists of a set of stimulus–response pairs: $\{(s_1, r_1), (s_2, r_2), \ldots, (s_n, r_n)\}$. As a kth nearest neighbor approach the estimate relies on a choice of a positive integer k. Let $d(i; k)$ be the distance from (r_i, s_i) to the kth nearest data-point, so $d(i; k)$ is the smallest distance such that

$$\left|\left\{\left(s_j,r_j\right)|d\left[\left(s_i,r_i\right),\left(s_j,r_j\right)\right] \le d(i;k)\right\}\right| = k+1 \quad (4)$$

where the set has size $k+1$ rather than k because it includes the point (s_i, r_i) itself. Now let

$$C_S(i) = \left|\left\{\left(s_j|d\left(s_i,s_j\right) \le d(i;k)\right\}\right| \quad (5)$$

and

$$C_R(i) = \left|\left\{\left(r_j|d\left(r_i,r_j\right) \le d(i;k)\right\}\right| \quad (6)$$

Thus $C_S(i)$ is the number of stimuli with any response within $d(i; k)$ of s_i; this will, of course, include the k data points counted when defining $d(i; k)$, but it may also include points (s_j, r_j) where $d(s_i, s_j) \le d(i; k)$ but $d(r_i, r_j) > d(i; k)$. Hence $C_S(i) \ge k$. The same sort of observation applies to $C_R(i)$. Roughly speaking, a separate k-nearest neighbor estimate is calculated for $H(R)$, $H(S)$ and $H(R, S)$ with $C_R(i) - 1$ and $C_S(i) - 1$ playing the role of k for $H(R)$ and $H(S)$. Now, the estimate of the mutual information is

$$\begin{aligned} I(R;S) = \quad & \psi(k) + \psi(n) - \frac{1}{n}\sum \psi[C_S(i)] \\ & -\frac{1}{n}\sum \psi[C_R(i)] \end{aligned} \quad (7)$$

where $\psi(x)$ is the digamma function of x.

If the stimulus space is discrete, the same approach can still give an estimate of the mutual information (Tobin and Houghton 2013). If there are n_s stimuli and each is presented n_t times, the mutual information is

$$\begin{aligned} I(R;S) = \psi(k) + \psi(n_s n_t) - \psi(n_t) \\ -\frac{1}{n}\sum \psi(C_R(i)) \end{aligned} \quad (8)$$

In Tobin and Houghton (2013) a similar formula is derived using a kernel-density-estimate based approach. An alternative Kozachenko-Leonenko estimate of the mutual information between a discrete stimulus set and spiking responses is given in Victor (2002).

References

Aronov D, Reich DS, Mechler F, Victor JD (2003) Neural coding of spatial phase in v1 of the macaque monkey. J Neurophysiol 89:3304–3327

Houghton C (2009) Studying spike trains using a van Rossum metric with a synapses-like filter. J Comput Neurosci 26(1):149–155

Houghton C (2015) Calculating mutual information for spike trains and other data with distances but no coordinates. R Soc Open Sci 2(5):140391

Houghton C, Kreuz T (2013) Measures of spike train synchrony: from single neurons to populations. In: Schuster H, Pesenson M (eds) Multiscale analysis and nonlinear dynamics: from molecules to the brain. Wiley, Weinheim

Houghton C, Sen K (2008) A new multi-neuron spike-train metric. Neural Comput 20(6):1495–1511

Houghton C, Victor J (2012) Spike rates and spike metrics, Chap. 8. In: Kreiman G, Kriegeskorte N (eds) Visual population codes: toward a common multivariate framework for cell recording and functional imaging. MIT Press, Cambridge, MA

Kozachenko LF, Leonenko NN (1987) Sample estimate of the entropy of a random vector. Probl Peredachi Inf 23(2):9–16

Kraskov A, Stögbauer H, Grassberger P (2004) Estimating mutual information. Phys Rev E 69:066138

Kreuz T, Haas JS, Morelli A, Abarbanel HDI, Politi A (2007) Measuring spike train synchrony. J Neurosci Methods 165(1):151–161

Nemenman G, Lewen W, Bialek W, de Ruyter van Steveninck R (2007) Neural coding of natural stimuli: information at sub-millisecond resolution. BMC Neurosci 8(Suppl 2):S7

Paiva ARC, Park I, Principe JC (2009) A reproducing kernel Hilbert space framework for spike train signal processing. Neural Comput 21(2):424–449

Paninski L (2003) Estimation of entropy and mutual information. Neural Comput 15(6):1191–1253

Panzeri S, Senatore R, Montemurro MA, Petersen RS (2007) Correcting for the sampling bias problem in spike train information measures. J Neurophysiol 98(3):1064–1072

Strong SP, Koberle R, de Ruyter van Steveninck RR, Bialek W (1998) Entropy and information in neural spike trains. Phys Rev Lett 80(1):197–200

Tobin RJ, Houghton CJ (2013) A kernel-based calculation of spike train information. Entropy 15(10):4540–4552. http://www.mdpi.com/1099-4300/15/10/4540

van Rossum M (2001) A novel spike distance. Neural Comput 13:751–763

Victor D (2002) Binless strategies for estimation of information from neural data. Phys Rev E 66(5):051903

Victor D (2005) Spike train metrics. Curr Opin Neurobiol 15(5):585–592

Victor JD, Purpura KP (1996) Nature and precision of temporal coding in visual cortex: a metric-space analysis. J Neurophysiol 76(2):1310–1326