Hierarchical Odor or Speech Recognition — Bayesian Inference embodied in Neuronal Spikes

David J.C. MacKay
University of Cambridge

March 20th, 1996

Abstract

A note on recognition of decomposable objects with uncertainty about intensity.

I apologise this note is rather rough! I trust it will convey the basic idea.

This unpublished work was motivated by the original work of John Hopfield.

1 Concepts

JJH's odoriforous paper emphasised these concepts: intensity of components could be represented by a time advance; and 'smell phonemes' could be recognized by neurons receiving appropriate conjunctions of inputs with appropriate time delays. This has the consequence that the phoneme detectors will give a response whose timing retains intensity information. And the number of spikes will represent in some rough sense the probability that the phoneme is actually there. Thus the scheme can be continued hierarchically with a 'smell word' recognizer being composed by an appropriate conjoining of outputs of phoneme recognizers.

This note explores the probabilistic issue in a bit more depth. Let us denote an nth level phoneme hypothesis by $\mathcal{H}_p^{(n)}$; this hypothesis states 'smell phoneme (n, p) is present' but does not state the intensity of the phoneme, which we could denote $\alpha_p^{(n)}$. (n is a label for the level of the hierarchy, with $n = 1$ being raw sensor components, $n = 2$ being simple phonemes, $n = 3$ being compound phonemes, etc.; p is the ID of the particular phoneme in question.)

Let us retain the idea that timings represent intensity on a log scale, and aim for a model in which the number of action potentials has a precise probabilistic meaning, and the distribution of those action potentials in time represents the uncertainty about the intensity. Uncertainty about intensity will be represented by a variety of timings, and poor fit to data by the model in question is represented by a small number of spikes.

Let us proceed by induction, looking at a two level system.

1.1 The ideal computations

In an ideal Bayesian approach we can describe the probability of the data with a latent variable model. Let's look at a two level system, and assume that the αs are defined such that if one phoneme alone is present at intensity $\alpha_p^{(2)}$ then its children $\{q\}$ must all be present with intensities $\alpha_q^{(1)} = \alpha_p^{(2)} + \tau_{pq}$. Note that there may be other hypotheses p' which have the same children $\{q\}$ and which would predict other intensities $\alpha_q^{(1)} = \alpha_{p'}^{(2)} + \tau_{p'q}$. The matrix $\tau_{p'q}$ stores the fingerprints of all these phonemes.

1.1.1 What we want to know

In order to compare competing hypotheses p and p' in the light of data D, we want to know the relative values of $P(D|\mathcal{H}_p)$; and when communicating on to the next level the intensity that we think we have observed, we want to compute $P(\alpha_p|D, \mathcal{H}_p)$. These are the 'two levels of inference' and Bayes theorem says:

$$P(\alpha_p|D, \mathcal{H}_p) = \frac{P(D|\alpha_p, \mathcal{H}_p)P(\alpha_p|\mathcal{H}_p)}{P(D|\mathcal{H}_p)} \tag{1}$$

where the normalizing constant is:

$$P(D|\mathcal{H}_p) = \int d\alpha_p P(D|\alpha_p, \mathcal{H}_p)P(\alpha_p|\mathcal{H}_p). \tag{2}$$

Ideally, we want to communicate to our superiors these two objects: the probability density $P(\alpha_p|D, \mathcal{H}_p)$ and the 'evidence' $P(D|\mathcal{H}_p)$. In fact, on second thoughts — maybe our superiors don't want us to go putting in any prior over α_p. What they would like to hear from us is simply the likelihood function $P(D|\alpha_p, \mathcal{H}_p)$. How does this likelihood function relate to the quantities inferred at the previous level of our hierarchy? OK, let's assume that our detectors at level 1 are diligent sub-Bayesians who know that their duty is to compute $P(d_q|\alpha_q^{(1)})$. Then, thinking of the data D as the set of $\{d_q\}$ for all relevant children,

$$P(D|\alpha_p^{(2)}, \mathcal{H}_p^{(2)}) = \prod_q P(d_q|\alpha_q^{(1)} = \alpha_p^{(2)} + \tau_{pq}). \tag{3}$$

2 Monte Carlo computation

Let us proceed by induction, as I said before. Assume that Mr. q produces spikes as a Poisson process over the time variable $\alpha_q^{(1)}$) with density proportional to the likelihood function $P(d_q|\alpha_q^{(1)})$. Assume that Mr. p has spines which compute the conjunction (to some precision in time) of spikes with appropriate time delays, so that Mr. p only fires if he receives simultaneous spikes from Messrs. $\{q\}$ down delay lines with delays $\{\tau_{pq}\}$. Then what is the density of firing of Mr. p? Clearly, the probability of the conjunction occurring will be proportional to the product of the densities, so the spikes produced by Mr. p will have a density proportional to the likelihood $P(D|\alpha_p^{(2)}, \mathcal{H}_p^{(2)})$. The constant of proportionality may be different from the one assumed at the preceding level, but within any given level, two models who share the same children will be able to compare their evidences, as defined in equation (2) simply by counting up the number of spikes! Whether the constant of proportionality gets bigger or smaller will depend on how many events are conjoined at one spine and what the time resolution of the conjunction detector is.

In conclusion, the decomposed representation computes precisely the desired quantities and represents them in a Monte Carlo fashion.

2.1 Possible refinement

A hypothesis doesn't really just predict the occurrence of events with appropriate intensities. It also predicts the non-occurrence of other events. These relationships could be implemented by adding to the list of conjoined events at spines a list of other non-events which must also match. These would be inhibitory connections, in hardware terms.

I haven't figured out if the Monte Carlo implementation could pull off this aspect of the computation of $P(D|\mathcal{H}_p, \alpha_p)$ faithfully, however.