Merging race networks and Kohonen Self Organizing Map

Denis Cousineau
denis.cousineau@umontreal.ca

References

(a) Cousineau, Lacroix, Hélie (in press). Redefining the rules. Connection Science.

(b) Cousineau, D. (submitted). Merging race models and adaptive networks. Psychonomic Bulletin & Review.

(c) Kohonen, T. (1984). Self-organization and associative memory. Berlin: Springer-Verlag.

(d) Huber & Cousineau (in press). A race model of forced-choice RT. Cognitive Science.

Objective

Sampling models, such as random walk and race models, are very powerful and plausible psychological models. However, there is no learning rule for such models. This reduces their range of applicability in psychology.

In ref. (a) and (b), we showed how to transform a race model into a *race network* using a feed-forward architecture and a modified Δ rule.

Along the way, we *generalized* the matrix operations, which turned out to be useful to summarize many different types of networks.

The objective of this poster is to describe *an unsupervised* race network. It is built around the notion of *Self-Organizing Map (SOM)* developed by Kohonen in (c).

General approach

a) Turning a supervised network (such as the Race Network) into a network with a 2-dimension surface of output units is not a problem. Further, both standard and redefined matrix operations are readily generalized using tensor operations.

b) Turning the Δ rule into an unsupervised learning rule:

In PRN, zero is the wanted signal;

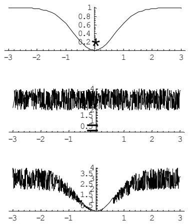
By convolving:

(1 – *Hat*) function centered on the winner with

the observed responses, we get

the expected responses **E**.

Thus: $(\mathbf{E} - \mathbf{O}) = (1 - Hat)\mathbf{O} - \mathbf{O} = -Hat\,\mathbf{O}$.



This error correction signal is inserted in the Δ rule:

$$\Delta\mathbf{W} = \alpha\mathbf{I}_{)+(}(\mathbf{E}-\mathbf{O})$$
$$= \alpha\mathbf{I}_{)+(}(-Hat\,\mathbf{O})$$

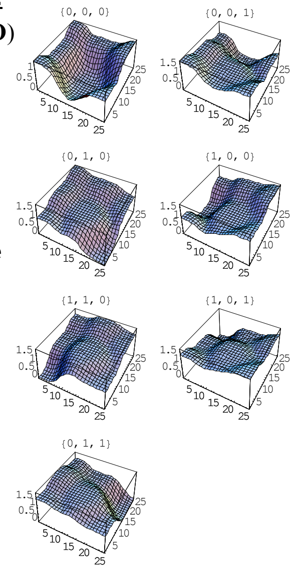
Results

Repres.	Training	Learning
$\{0, \infty\}$	$\mathbf{O} = \mathbf{I}_{\binom{+}{\vee}}\mathbf{W}$	$\Delta\mathbf{W} = \alpha(t)\mathbf{I}_{)+(}(-Hat^{1/\sigma^2(t)}\mathbf{O})$

α(t) and σ(t) are still time-dependant parameters…

However, assuming that the *Hat* is not perfect (height fixed at α, say 0.80), we obtain that $\alpha(t) = 1-(1-\alpha)^t$, an exponential function.

As of σ(t), it seems to depend on how rapid the response *Min*(O) is (a rich-gets-richer effect, found empirically in ref. d)



In conclusion

- The parameters α and σ may have a simple physical explanation in the context of a race model;
- The *Hat* function suggests that lateral activation is a signal that requires time to travel, again a natural assumption in the context of a race model;
- The rich-gets-richer effect might be related to a massive amount of redundancy at a micro level.

Appendices

A New Notation

Most neural networks are built using matrix notations.

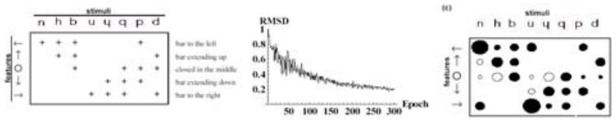
An inner product joins pairs of value using × and aggregates columns using Σ. This can be made explicit with: $\mathbf{I}\,.\,\mathbf{W} = \mathbf{I}_{\binom{\times}{\Sigma}}\mathbf{W}$

A race model needs to find the smallest of additive delays. This can be turned into vector-and-matrix notation with: $\mathbf{I}_{\binom{+}{\vee}}\mathbf{W}$

Likewise for outer product: $\mathbf{I}_{)\times(}\mathbf{W}$ or $\mathbf{I}_{)+(}\mathbf{W}$

Overview of the supervised networks

They are supervised because a teacher provides **E**, the expected response. In the Δ rule, changes are proportional to the size of the input and the amplitude of the error.

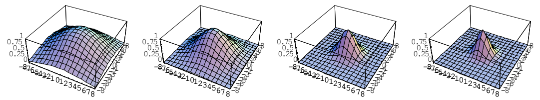


	Representation	Transmission	Learning
Perceptrons: (strength-based)	$\{1, 0\}$	$\mathbf{O} = \mathbf{I}_{\binom{\times}{\Sigma}}\mathbf{W}$	$\Delta\mathbf{W} = \alpha\mathbf{I}_{)\times(}(\mathbf{E}-\mathbf{O})$
Race networks: (time-based)	$\{0, \infty\}$	$\mathbf{O} = \mathbf{I}_{\binom{+}{\vee}}\mathbf{W}$	$\Delta\mathbf{W} = \alpha\mathbf{I}_{)+(}(\mathbf{E}-\mathbf{O})$

A limitation of these networks: they are supervised…

Overview of the Kohonen networks

Kohonen networks learn to recognize patterns, but more importantly, they organize the knowledge using proximity relations. It uses the notion of neighborhood during learning (*Hat*) whose width reduces with learning:



Representation	Transmission	Learning
none	$\mathbf{O} = \mathbf{I}_{\binom{-}{\|\,\|}}\mathbf{W}$	$\Delta\mathbf{W} = \alpha(t)\times Hat^{1/\sigma^2(t)}\times\mathbf{I}_{\binom{-}{List}}\mathbf{W}$

where the *Hat* is always centered on the winner *Min*(O).

Problems with this network:

i) The height α(t) and width σ(t) of the neighborhood are time-dependant parameters reducing with practice. They are unprincipled.

ii) There is no privileged representation. Hence, the Δ rule is inapplicable.