Title
Neural Voting Machines

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 25(25)

ISSN
1069-7977

Authors
Richards, Whitman
Seung, Sebastian

Publication Date
2003

Peer reviewed

Neural Voting Machines

Whitman Richards & Sebastian Seung
Artificial Intelligence Lab
MIT NE43-767, Cambridge, MA 02139
{wrichards,seung}@mit.edu

Abstract

In theories of cognition that view the mind as a system of interacting agents, there must be mechanisms for aggregate decision-making, such as voting. Here we show that certain voting procedures studied by social scientists can be implemented as recurrent neural networks. For example, a standard "winner-take-all" network can determine which of a number of competing alternatives garners a plurality of votes. Similarly, in the special case where voters share a model governing the different rankings of alternatives, the Borda procedure can easily be computed. In the face of voter un-certainties, this Borda network returns the maximum likelihood choice.

1.0 Introduction

Information aggregation in neural networks is a form of collective decision-making. The winner-take-all procedure is probably the most favored method of picking one of many choices among a landscape of alternatives (Hopfield & Tank, 1986; Maas,2000.) In the social sciences, this is equivalent to choosing the plurality winner, which is but one of a host of procedures that could be used to choose winners from a set of alternatives. More importantly, in the presence of uncertainty about choices, the plurality winner is not the maximum likelihood choice (Young, 1995.) To obtain a glimpse into some of the problems associated with winner-take-all outcomes, consider the analogy where the input landscape is a population of voters. Let the number of voters sharing the same opinions correspond to the input weights in a neural network. Then the plurality winner - that outcome shared by most of the voters -- only needs to receive more votes than any other alternative in the choice set. Hence it is possible for the winner to garner only a very small percentage of the total votes cast. In this case, uncertainty and errors in opinions can have a significant impact on outcomes, such as when only a few “on-the-fence” voters switch choices. We sketch two other procedures that yield more reliable and robust winners. These procedures utilize information about relationships among alternatives.

2.0 Plurality Voting

To provide background, the winner-take-all procedure is recast as a simple voting machine. Let the number of voters v_i sharing the same preference for a winner be inputs to the nodes in the network. Then the outcome will be

$$\text{plurality\_winner} = \text{argMax}(i)\{v_i\} \quad [1]$$

which can be found using a recurrent network whose dynamics is described elsewhere (Xie, Hahnloser & Seung, 2001 .)

3.0 Borda Method

To improve the robustness of outcomes, we now follow recommendations in Social Decision-Making, and relax the constraint that *only* first choices will be considered in the voting process (Runkel, 1956; Saari & Haunsberger, 1991; Saari, 1998) Specifically, we include second and third-rank opinions, weighting these inversely to their rank when the tally is taken (Borda, 1784.) To further simplify the computation and network design, we assume that the alternative choices are related by a model **M**n that is held in common by all voters. This model relates the n alternatives under consideration by their similarity to one another.

The shared model **M**n can be represented either as a graph, or as a matrix **M**ij. If **M**n is represented as a graph, the vertices would correspond to the alternatives, and the edges ij join nodes that share a common property. (See Fig. 1 for an example.) If ak is a voter's first choice, then the second choices will be those alternatives one edge-step from ak in **M**n. The result is that the total of m voters can now be divided into n different types, identified by their first choice selection.

If the shared model **M**n is represented as a matrix **M**ij, the entry "1" indicates the presence of the edge ij, and 0 otherwise (Harary,1969). For the graphical model of Fig 1, we would have:

$$\mathbf{M}ij \quad = \quad \begin{matrix} 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{matrix} \qquad [2]$$

For simplicity, we assume that the edges of Mn are undirected, meaning that if alternative a1 is similar to alternative a2, then a2 is equally similar to a1. However, directed edges require only a trivial modification to our scheme. Note that if all voters respect Mn in their ranking of choices, as we specify here, then the effective role of Mn is to place conditional priors on the choice domain. Each voter's ranking of alternatives is now not arbitrary, but is also reflecting information about choice relationships (Richards et al, 1998.)

With **M**n expressed as the matrix **M**ij we can include second choice opinions in a tally by defining a new voting weight v*i as

$$v^*i \; = \; \{\; 2\, vi \; + \; \Sigma \;\; \mathbf{M}ij \; \mathbf{v}j \;\; \} \quad [3]$$

where now first-choice preferences are given twice the weight as second-ranked choices, and third or higher ranked options have zero weight. The outcome is then

$$\text{winner\_Borda} \; = \; \text{argMax(i)} \; \{v^*i\} \; [4]$$

The neural network required to execute this tally is shown in Fig 1. It is a simple modification of the standard winner-take-all network, with a doubling of the input weights from each excitatory node to its recurrent partner (double arrows), and with single excitations to non-partner nodes that are adjacent in the model **M**n. (The recurrent layer does not show all the recurrent connections.) For the inputs vi given in the model **M**n, the Borda winner is node 3. Note that the more common winner-take-all plurality procedure would pick node 1.

4.0 Robustness

Figure 2 shows the benefit of the Borda procedure over classical winner-take-all plurality methods, when some information about alternatives in the domain is known. The models **M**n used were connected random graphs with edge probability 1/2. (See Richards et al, 2002 for more details.) A set of weights on the nodes was chosen from a uniform distribution. Winners were calculated using both the Plurality and Borda procedures for the same set of weights. Then each of these weights were diddled by picking the second weight from the interval 0.5 to 1.5 of the first. The graph shows the percent of time the first and second winners were the same vs the number of alternatives in **M**n. (There are over 100 trials per data point.) For the Borda (B) procedure, even for n = 48, the changes of weights (or voting strengths) only affected 20% of the outcomes, whereas for the maximum weight, Plurality procedure (M), over 70% of the outcomes differ. Not surprisingly, the Borda and Plurality winners are increasingly different as n increases, with only 2% agreement for n=48 (solid circles.)

5.0 Other Voting Procedures

Our Borda Count used only first and second choice preferences in the tally, with respective weightings of 2 and 1 times the voter type's own weight. Let this bias be recast as a vector {1, 1/2, 0 }, where the 0 is the weight applied to all preferences ranked after second choices. Then it is clear that the bias for the Plurality method is {1,0,0}. Yet another procedure would be to vote for the "top two" choices, using the bias vector {1,1,0}. More generally, the Borda bias vector can be seen as {1, b, c} with 0 < b < 1 and c = 0 for our simplified preference rankings. Hence the Top-Two and Plurality procedures are extremes of a generalized Borda count.

Another obvious manipulation is to increase the depth of the preference ranking, thereby incorporating more information about the relationships among alternatives. As mentioned, for the standard Borda method, the elements of the bias would then be integers inverse to the depth of the rankings. A still different procedure that also incorporates more infor-mation than the generalized Borda method is to conduct a tournament, where alternatives are compared pairwise. The winner is then that alternative that beats all others. Note now there is no need to decide values for "b" in the Borda bias vector. This is the Condorcet Method (Condorcet, 1785.)

Definition: let dij be the minimum number of edge steps between vertices i and j in **M**n, where each vertex corresponds to the alternatives ai and aj respectively.

Then a pairwise Condorcet score Sij between alternatives ai and aj is given by

$$S_{ij} = \Sigma_k \, v_k \, \mathrm{sgn}[\, d_{jk} - d_{ik}] \quad [5]$$

with the sign positive for the alternative ai or aj closer to ak.
A Condorcet winner is then

$$\text{winner\_ Condorcet} = \textit{ForAll}_{i=!=j} \; S_{ij} > 0 \,. \quad [6]$$

Although a Condorcet winner is a true majority outcome, it comes at a computational cost. For n alternatives, a complete pair-wise comparison would require (n | 2) or O(n^2) separate tallies. Hence a neural network that calculates the Condorcet winner is more complex than that for the Borda winner. However, if the voting is constrained by a shared model **M**n, or its equivalent **M**ij, simulations using a Borda bias vector of {1, 0.5, 0} show that about 90% of the time, the Borda and Condorcet winners will agree if **M**n resembles a random graph.

6.0 Discussion

Biological neural networks are not arbitrary, and presumably the form of their organization incorporates knowledge about the domain of interest. When information about the choice domain is available and used, then significant improvements in performance can be achieved with networks that implement a simple version of the Borda method. The Borda network's resistance to perturbation in the weights on inputs is demonstrated here. Preliminary studies show that a Borda network will also be robust to small inconsistencies in the shared model **M**n.

References

Arrow, K.J. (1963) Social choice and Individual Values. Wiley, NY

Borda, J-C. (1784) Memoire sur les elections au Scrutin. Histoire de l'Academie Royal des Sciences.

Condorcet, Marquis de (1785) Essai sur l'application de l'analyse a la probabilite des decisions rendue a la pluralite des voix, Paris (See Arrow, 1963)

Hopfield, J.J. and Tank, D. (1986) Computing with neural circuits: a model. Science 233, 625-633.

Maass, W. (2000) On the computational power of winner-take-all. Neural Computing. 12, 2579-35.

Richards, D., McKay, B. and Richards, W. (1998) Collective Choice and mutual knowledge structures. Adv. in Complex Systems. 1, 221-36.

Richards, W., McKay, B. and Richards, D. (2002) Probability of Collective Choice with shared knowledge structures. Jrl Math. Psychol. 46, 338-351.

Runkel, P.J. (1956) Cognitive similarity in facilitating communication. Sociometry, 19, 178-91.

Saari, D. G. (1994) Geomtry of Voting. Springer-Verlag,, Berlin

Saari, D. and Haunsberger, D. (1991) The lack of consistency for statistical decision procedures. The Amer. Statistician, 45, 252-55.

Xie, X-H, Hahnloser, R. and Seung, H.S. (2001) Leanring winner-take-all competitions between grouips of neurons in lateral inhibiting networks. Adv. Neural. Info. Proc. 13, 350-6.

Young, H.P. (1995) Optimal Voting Rules. Jrl. Econ. Perspectives 9, 51-64

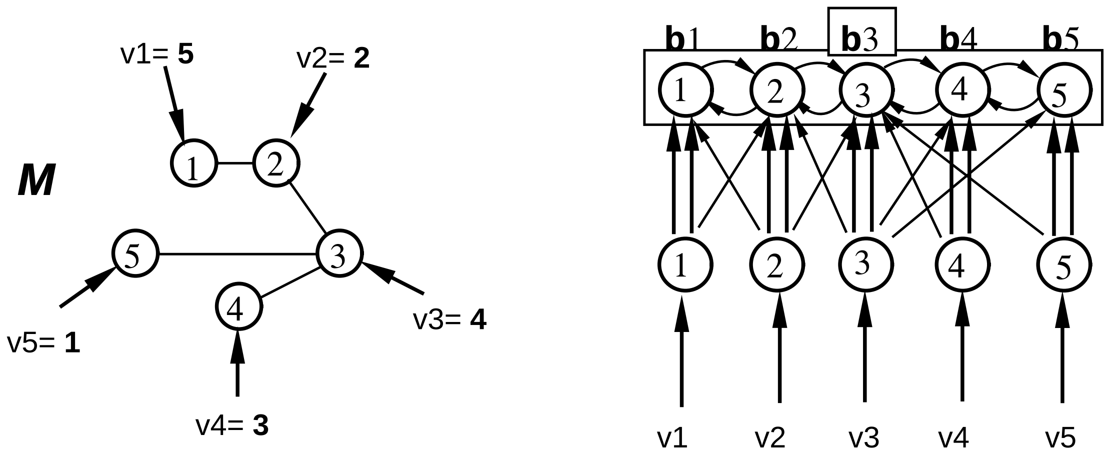

Fig. 1: Borda count network for shared model M

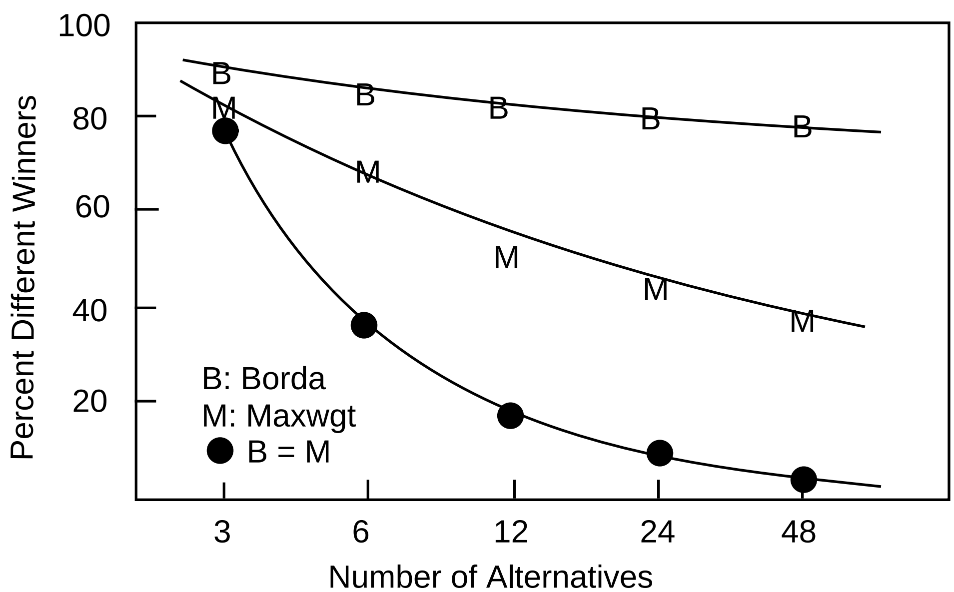

Fig. 2 Percent different Winners with mean weight variation of 50%