Learning to Do Better

Using Boosted Neural Networks and Leader Experience Measures to Improve the Accuracy of Conflict Prediction Models

Patrick James Rogers

December 9, 2008

1 Introduction

It has become almost cliche to bemoan the sorry state of quantitative conflict research. This paper will be no exception. Indeed, most of the assumptions necessary in order to analyze conflict using statistical methods are rather egregiously violated by the data. The purpose of this paper is to propose and test (a) the use of boosted, single hidden layer, feed-forward perceptron neural networks to predict conflict likelihood, (b) the addition of an instrumentalization of state learning into the predictive model. I find that while standard neural network regression does offer an improvement over logistic regression methods, the advantage of neural networks can be significantly extended by using boosting methods to transform the "weak" neural net into a "strong" learner. The predictive advantage of boosted neural networks is further enhanced by the inclusion of measures of leader tenure as proxy measures of state learning.

2 Neural Network Regression

Conflict studies suffer from a number of challenges intrinsic to the research question. Conflict is a rare event — the overwhelming number of possible observations include a zero for the dependent variable. It is without a doubt proximately

multi-causal — there appears to be at least one unique triggering event for every conflict on record. It is very likely generally multi-causal as well — the most widely accepted model of conflict onset points to information asymmetry, commitment problems and issue indivisibility as the three root causes of conflict (Fearon, 1995; Fey & Ramsay, 2006; Filson & Werner, 2002; Garfinkel & Skaperdas, 2000; Gartzke, 2001; Powell, 2002; Reiter, 2003; Slantchev, 2003; Smith & Stam, 2004; Wagner, 2000). Data collection is a difficult prospect. Even when government reports can be credibly trusted as representing elite opinion, data on battle deaths are often inconsistent and untrustworthy. Years can pass before a final tally of casualties is finalized. Other information can be impossible to get, such as information on who talked to whom behind closed doors, or what factors most influenced decision-makers towards conflict rather than settlement. If private information — and the incentive to misrepresent — is legitimately an ultimate cause of war, the predictive power of any model is severely curtailed. Private information is definitionally ex ante unknown, leaving researchers no data for prospective modeling (Gartzke, 1999). Data collection problems notwithstanding, much of the classification decisions are made by human experts, and therefore exhibit noise, subjectivity and sometimes contradictory classifications[1]. Levels of democracy, levels of hostility, and discerning the identity of initiator all require the researcher to make a decision, and each decision represents an opportunity influence the data — either consciously or unconsciously. Coding rules can reduce the ad hoc nature of such intervention, but even under the best of circumstances errors will get made and bias will be introduced.

For the most part this paper will not be dealing with the problems outlined above. This is not to ignore the importance of addressing these issues, nor to denigrate the efforts of those researcher developing techniques to overcome the limitations of the data just described. The methodological approach discussed in this paper is not in competition with, but rather complements their work. Integra-

[1]For a discussion on how one of the central databases in conflict research, the Correlates of War Project, was initially coded, see Singer and Small (1972).

tion is the ultimate goal.

In addition to the problems discussed above, conflict studies suffer from a misapplication of statistical method. I adopt here the viewpoint of (Beck, King, & Zeng, 2000). Conflict data is complex, non-linear and strongly interacted. The importance of any single variable on the probability of conflict is strongly dependent on the ex ante likelihood of war — that is, one variable having an impact is predicated on other variables also pointing strongly in the direction of conflict as well. This level of interaction makes linear regression models poor estimators. Most general linear models require assumptions of near-linearity in the data, independence of observations, and identical effects of IVs over all observations. The poor match between statistical assumptions and the empirical data leads to predictive models with statistically weak relationships, large confidence intervals, poor forecasting performance, and a lack of robustness.

One technique in the literature is to relax the assumption of strict linearity and instead adopt non-linear distributions as the link function to the general linear model. Logit and probit functions are examples of this approach. Although this improves estimator efficiency for non-linear data, it still assume a monotonic effect for each variable. That is, if the fitted coefficient for a given independent variable is positive, then an increase observed increase in that variable should always lead to an increase in the observed value of the dependent variable. these techniques essentially represent a mechanism to rescale non-linear data so that it can be treated linearly. If the variables are non-monotonic — or are so heavily interacted that the interaction effects are effectively non-monotone — then rescaling will not significant improve the model's fit. Another possible solution is the inclusion of interaction terms in the regression model. Interaction terms — by allowing two terms with differing signs to be interacted — raise the possibility of fitting a non-monotonic function using a linear regression model. One problem with interaction terms is that there remains a measure of collinearity between the interaction term and its constitutive variables. It is suspected that conflict data is very strongly interacted, which would require the inclusion of a large num-

ber of interaction terms, multiplying collinearity and making the predictive model highly unstable. Another problem with the use of interaction terms is that the researcher must explicitly include them in the regression model. As the number of independent variables increases, the number of possible interactions increases at a much higher rate. Sorting through the possible interaction terms, theorizing effectively on which to include and which to discard, and then coding the model represent a significant drain on the researcher's time. If the outcome of the analysis represented useful results, the time could be considered wisely spent; however, given the problems discussed thus far, any inferences drawn from such an analysis would be largely suspect.

Instead, a regression technique that can handle the high degree of interaction is necessary. A particularly effective technique for pattern recognition under such circumstances is the use of neural networks. In particular, a type of multi-layer perceptron (MLP) called a single hidden layer, feed-forward perceptron neural network. The use of such networks in the analysis of political science data is hardly novel (Beck et al., 2000; Lagazio & Russett, 2004; Zeng, 2000, 1999); however, its use has remained on the sidelines. This avoidance is likely due to the perception of neural networks as a "black box" which prevents researchers from having effective control over their analyses, and makes interpretation of outputs difficult (Marchi, Gelpi, & Grynaviski, 2004). However, the black box phenomenon is more likely due to unfamiliarity with the technique on the part of researchers, rather than due to the mysterious nature of neural nets. Beck, King, and Zeng (2004) point out that the criticisms being leveled against neural network regression techniques are the same complaints initially leveled against logit and probit models when they were first applied within the discipline.

Beck et al. (2000) (hereafter referred to as BKZ) provide an excellent overview of the statistical basis of neural network regression. Instead of repeating their work here, I offer a much shorter introduction to neural networks. Whereas standard regression models can be thought of as the application of some linear function to a matrix of independent variables X_i:

$$\text{linear}(X_i)$$

and logistic regression models can be thought of as applying a linear function to a matrix of independent variables X_i, then rescaling the output according to a non-linear distribution:

$$\text{logit}(\text{linear}(X_i))$$

Neural network regression consists of a set of nested logistic regressions:

$$\text{logit}(\text{linear}(\text{logit}(\text{linear}(X_i))))$$

Equation 3 above represents 1 pathway through the neural network. Each logit function represents a node in the network where the data is individually fit according to a logistic regression. In the special case where there is only one pathway through the network (i.e. the network links straight from the inputs into a single hidden layer consisting of a single node, which then links into a single output node), then the equation above represents the entire model. Adding more nodes to the hidden layer increases the number of pathways from input to output, therefore increasing the number of linear(logit(X_i)) functions fitting the data.

The network is trained by adjusting by weighting the links from the input layer to the hidden layer and from the hidden layer to the output layer. These weights represent the linear(x) in Equation 3 above. Back-propogation is one of the most popular techniques for fitting neural networks. Training via back-propogation starts with randomized weights, typically drawn from a uniform distribution $U(-1, 1)$. Observations from the training set are run through this "null" network, and the model's predicted outcome is then checked against the measured outcome. The

matrix of weights is then adjusted to minimize a penalty function, such as the sum of squares. This new matrix of weights is then re-tested against the training set until the penalty function converges to a stable value. It should be clear that the efficiency of the model is largely dependent on the selection of the number of hidden-layer nodes. Too few hidden layer nodes will prevent the neural network from fitting the true distribution of the data — i.e the model has "underfit" the data . Conversely, too many hidden layer nodes will cause the neural network to begin fitting to noise in the training set, leading to "overfitting". There are techniques to use information from the dataset itself to estimate the correct number of hidden layer nodes, but the most straightforward method is to use the training set to train multiple networks, varying the number of hidden layer nodes, then pick the network with the lowest residual sum of squares.

3 Boosting

In their use of neural network regression to test the impact of various IVs on conflict likelihood, BKZ adopt the technique described in the previous section. Their results (Table 1) are based on a study of Politically Relevant Dyads (PRDs) from 1947 to 1989. Training their network on a subset of the data from 1947 to 1985, then running a test on the years 1986 to 1989, the trained neural network was able to correctly predict 16.7% of out of sample conflict episodes and 99.42% of out of sample peace episodes.

Table 1: Prediction Results of Single Neural Network on Conflict Outcomes 1986 to 1989 (Beck, King and Zeng 2000)

	Conflict (Predicted)	Peace (Predicted)
Conflict (Observed)	28	140
Peace (Observed)	28	4768

This represents a substantial improvement over forecasting via logistic re-

gression. Single logit models aren't able to separate dyads where the influence of explanatory variables are significant from dyads where the explanatory variables are weak. The result is a dilution of the estimates for the coefficients, such that the predicted outcome is always peace. A standard logic model will therefore correctly predict 0% of conflict episodes, but 100% of peace episodes. Given this, BKZ's 16.7% accuracy is important; however, it still misses the majority of conflict episodes. Fortunately, there is much room for improvement in the technique.

It has long been recognized that neural networks suffer from biasing predictions in favor of the modal outcome. This drawback is extremely relevant in conflict data, where the outcome of interest is particularly rare. Given this bias, a single neural network represents a "weak" learner. Boosting is a machine learning technique to leverage multiple weak learners into a single strong learner. Essentially, multiple neural networks are trained on subsets of the training data, and then linked into an ensemble predictor. This process is illustrated in Figure 1[2]:

Additionally, I propose adding an additional set of independent variables to test. Specifically, the tenure length of each state's executive leader. The purpose of these variables is to instrumentalize the ability of the state to learn. Experience offers leaders an additional resource with which to overcome the bounded rationality of uncertainty and time constraints to better perceive the bargaining range and find a negotiated outcome that is mutually preferable to war. Theories of private information–and the incentive to misrepresent–typically ignore the ability of leaders to learn, either directly from their prior experience, or vicariously, through the behavior of other states. Essentially, an experienced leader will be better able to infer the content of private information from secondary information sources, making the likelihood of successful bargaining more likely[3].

[2]Taken From "Enhancing Boosting by Feature Non-Replacement for Microarray Data Analysis" (Guile & Wang, 2007)

[3]Experienced leaders may also be better equipped to overcome commitment problems and issue indivisibility. Familiarity with mechanism to control for time inconsistency helps avoid credibility traps. Likewise familiarity with potential side payments helps avoid issue indivisibility. These concept has been more fully theorized in previous papers by the author. In the interest of space, I am leaving the learning concept under-theorized here.

4 Research Design

For the purposes of this paper, four different neural network training algorithms are proposed:

- A single neural network trained a single time
- 10 individual neural networks trained on the same training set, with a committee vote threshold of one
- An boosted ensemble neural network consisting of three differentially trained neural networks, with a committee of two and a third tie-breaker.
- 10 boosted neural networks trained as in 3 above, with an overall committee vote threshold of one.

Algorithm 1 is essentially identical to the method described in the previous sections, and is analogous to that used by BKZ. Algorithm 2 trains 10 networks individually (using Algorithm 1). For any given observation in the test set, if a single network predicts conflict, then conflict is predicted by the committee of networks. Algorithm 3 is adapted from Schapire (1990):

- The training set is subdivided into 3 subsets
- Neural Network 1 (N_1) is trained identically to Algorithm 1 on an initial training set subset
- The training set for N_2 is formed in the following way: (1) A random number is drawn from the uniform distribution U(o,1). (2) If it is greater than 0.5, then new training examples from the second subset are passed through N_1 until it misclassifies an example. That example is then added to a new training set. (3) If it is less than 0.5, then new training examples new from the second subset are passed through N_1 until it correctly classifies an example. That example is then added to the new training set. (4) Once all examples from

the second subset have been passed through N_1, N_2 is trained on the newly formed training set.

- The training set for N_3 is formed by passing the third training subset through both N_1 and N_2: (1) If they disagree on the predicted outcome, add that example to another new training set. (2) If they agree of the predicted outcome, ignore that example and go on to the next. (3) Once all examples from the second subset have been passed through N_1 and N_2, N_3 is trained on the newly formed training set.

N_1, N_2 and N_3 form a ensemble network N_e. The out of sample test is then passed through N_e. For each observation, if N_1 and N_2 agree, use their predicted outcome. If they disagree, pass the example through N_3 as a tie-breaker.

Finally, algorithm four trains three ensemble networks. For any given observation in the test set, if any of the ensemble networks predicts conflict, then the committee predicts conflict.

These four algorithms are used trained and tested twice. Once each using the same variables as BKZ use, then again with the addition of tenure measures for the leaders of each state in the dyad. Tenure data is calculated by subtracting the year the leader attained executive office from the current year of the dyad. Accession dates are taken from Mesquita, Smith, Siverson, and Morrow (2003).

5 Results

In all, each data set was fitted 4 times, once for each training algorithm. Table 2 through 9 below give the results

These initial results provide strong evidence that boosted neural networks represent a significant improvement over stand alone neural networks for forecasting and predicting the likelihood of conflict. Furthermore, the results are highly significant. In the case of Table 9, the likelihood of predicting at least 42 conflicts correctly out of 138 random guesses is 5.193296×10^{-33}.

Similar to the criticism of BKZ's 16.7% result, 52.5% accuracy leaves a lot of conflict episodes incorrectly predicted as peace episodes. A caveat however applies. Running the analysis at the dyad-year level, with the outcome variable representing the initiation of conflict that year places all predictive models at a disadvantage. It is extremely likely that a dyad that goes to conflict in a given years was very similar to itself in prior years. It is possible that the false positives represent dyads that are approaching the "threshold" for conflict. If that is the case, we may be predicting conflicts for dyads that will go to war, just not this year. In other words, we may be calling some of the conflict prediction false positives, when in fact they are just examples of the network doing a good job predicting further out than the arbitrary 1-year cutoff.

In any case, the gap between 52.5% prediction accuracy and 100% represents new opportunities to refine the predictive models, and further research into causal variables. One possibility is to recode the dependent variable such that it captures the presence of an imminent conflict. For instance by measuring the DV not along a dichotomous conflict/peace dimensions, but by counting the number of years until a conflict will occur, or coding the 2-3 years just prior to a conflict with a discounted value between 0 and 1.

6 Conclusion

Neural networks represent a particularly fruitful opportunity to improve predictive models in many aspects of international relations, and political science more generally. Conflict onset is one of the most intractable problems in IR. The rarity of the event, inherent noise in the coding, and strongly interacted explanatory variables all combine to make it extremely difficult to tease out the causal effects using standard regression models. The boosting model proposed here shows the proof of concept, but it is exceedingly simple. Much more complex and powerful algorithms have already been developed for other data management problems, such as in particle physics, biological systems, and applied engineering. These

methods are essentially ready off the shelf for the researcher willing to spend a little time learning the technique. Also interesting but of secondary importance is the evidence in favor of learning by state leaders, and the ability of leaders to translate experience into improved bargaining outcomes.

Figure 1: Boosting Flowchart

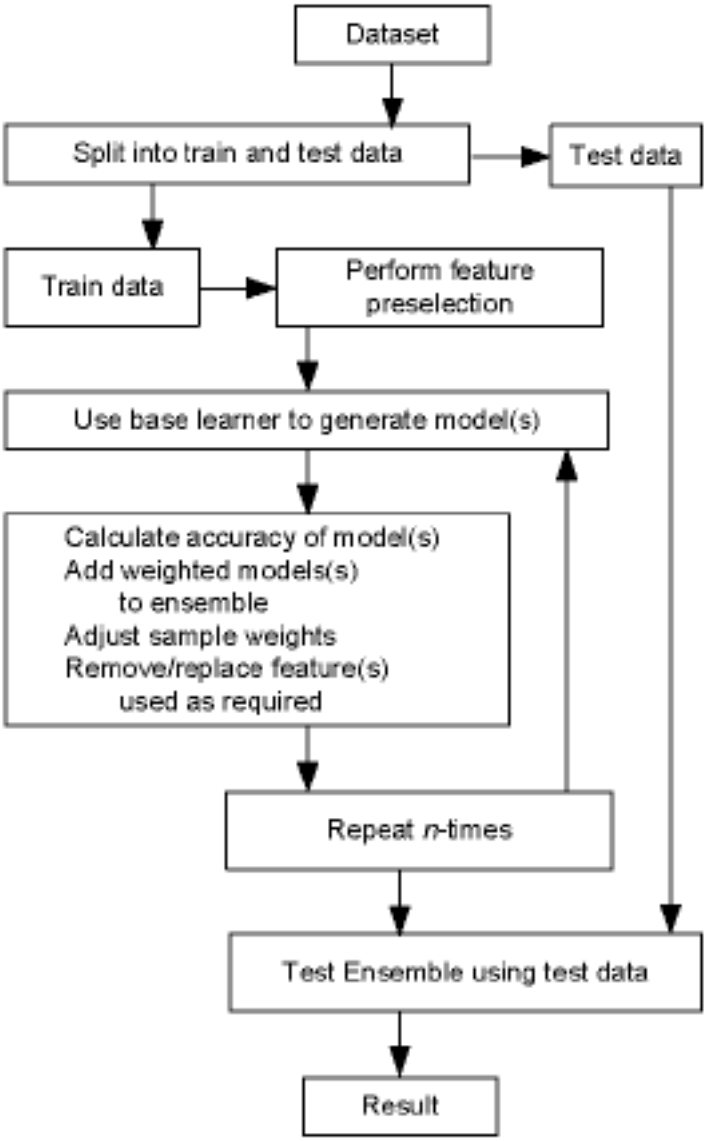

Table 2: Prediction Results of Single Neural Network on Conflict 1986 to 1989

	Conflict (Predicted)	Peace (Predicted)
Conflict (Observed)	13 (16.25%)	67
Peace (Observed)	23	2244 (98.99%)

Table 3: Prediction Results of Committee of 10 Neural Networks on Conflict 1986 to 1989

	Conflict (Predicted)	Peace (Predicted)
Conflict (Observed)	15 (18.7%)	65
Peace (Observed)	29	2238 (98.72%)

Table 4: Prediction Results of Boosted Neural Network on Conflict 1986 to 1989

	Conflict (Predicted)	Peace (Predicted)
Conflict (Observed)	17 (21.25%)	63
Peace (Observed)	28	2240 (98.77%)

Table 5: Prediction Results of Committee of 10 Boosted Neural Networks on Conflict 1986 to 1989

	Conflict (Predicted)	Peace (Predicted)
Conflict (Observed)	32 (40.00%)	48
Peace (Observed)	57	2211 (99.07%)

Table 6: Prediction Results of Single Neural Network on Conflict 1986 to 1989 (With Tenure Data)

	Conflict (Predicted)	Peace (Predicted)
Conflict (Observed)	14 (17.50%)	66
Peace (Observed)	21	2246 (99.07%)

Table 7: Prediction Results of Committee of 10 Neural Networks on Conflict 1986 to 1989 (With Tenure Data)

	Conflict (Predicted)	Peace (Predicted)
Conflict (Observed)	30 (37.5%)	50
Peace (Observed)	48	2219 (97.88%)

Table 8: Prediction Results of Boosted Neural Network on Conflict 1986 to 1989 (With Tenure Data)

	Conflict (Predicted)	Peace (Predicted)
Conflict (Observed)	18 (22.5%)	62
Peace (Observed)	36	2232 (99.07%)

Table 9: Prediction Results of Single Neural Network on Conflict 1986 to 1989 (With Tenure Data)

	Conflict (Predicted)	Peace (Predicted)
Conflict (Observed)	42 (52.5%)	38
Peace (Observed)	96	2172 (98.77%)

References

Beck, N., King, G., & Zeng, L. (2000). Improving quantitative studies of international conflict: A conjecture. *American Political Science Review*, *94*(1), 21–36.

Beck, N., King, G., & Zeng, L. (2004). Theory and evidence in international conflict: A response to de marchi, gelpi, and grynaviski. *American Political Science Review*, *98*(02), 379–389.

Fearon, J. D. (1995, Summer). Rationalist explanations for war. *International Organization*, *49*(3), 379-414.

Fey, M., & Ramsay, K. W. (2006, August). The common priors assumption: A comment on "bargaining and the nature of war". *Journal of Conflict Resolution*, *50*(4), 607-613.

Filson, D., & Werner, S. (2002, October). A bargaining model of war and peace: Anticipating the onset, duration, and outcome of war. *American Journal of Political Science*, *46*(4), 819-837.

Garfinkel, M. R., & Skaperdas, S. (2000, December). Conflict without misperceptions or incomplete information: How the future matters. *Journal of Conflict Resolution*, *44*(6), 793-807.

Gartzke, E. (1999, Summer). War is in the error term. *International Organization*, *53*(3), 567-587.

Gartzke, E. (2001, March 15). War, bargaining, and the military commitment problem. In *Political economy of conflict conference.* New Haven, CT: Yale University.

Guile, G., & Wang, W. (2007). *Enhancing boosting by feature non-replacement for microarray data analysis.*

Lagazio, M., & Russett, B. (2004). A Neural Network Analysis of Militarized Disputes, 1885-1992: Temporal Stability and Causal Complexity. In P. Diehl (Ed.), *The scourge of war: New extentions on an old problem* (pp. 28–62). Ann Arbor, MI: Univ of Michigan Press.

Marchi, S. de, Gelpi, C., & Grynaviski, J. (2004). Untangling neural nets. *American*

Political Science Review, *98*(02), 371–378.

Mesquita, B. Bueno de, Smith, A., Siverson, R., & Morrow, J. (2003). *The logic of political survival*. Cambridge, MA: MIT Press.

Powell, R. (2002). Bargaining theory and international conflict. *Annual Review of Political Science*, *5*, 1-30.

Reiter, D. (2003, March). Exploring the bargaining model of war. *Perspectives on Politics*, *1*(1), 27-43.

Schapire, R. (1990). The strength of weak learnability. *Machine Learning*, *5*(2), 197–227.

Singer, J., & Small, M.(1972). *The wages of war, 1816-1965: A statistical handbook*. San Francisco, CA: Wiley.

Slantchev, B. L.(2003, February). The power to hurt: Costly conflict with completely informed states. *American Political Science Review*, *97*(1), 123-133.

Smith, A., & Stam, A.(2004, December). Bargaining and the nature of war. *Journal of Conflict Resolution*, *48*(6), 783-813.

Wagner, R. H.(2000, July). Bargaining and war. *American Journal of Political Science*, *44*(3), 469-484.

Zeng, L. (1999). Prediction and classification with neural network models. *Sociological Methods & Research*, *27*(4), 499.

Zeng, L.(2000). Neural network models and political data analysis. In D. Richards & D. Doyle (Eds.), *Political complexity: Nonlinear models of politics* (pp. 239–268). Ann Arbor, MI: Univ of Michigan Press.