Assessing Different Bayesian Neural Network Models for Militarized Interstate Dispute: Outcomes and Variable Influences

Monica Lagazio and Tshilidzi Marwala

Social Science Computer Review 2006; 24; 119
DOI: 10.1177/0894439305281512

Social Science Computer Review
Volume 24 Number 1
Spring 2006 119-131
© 2006 Sage Publications
10.1177/0894439305281512
http://ssc.sagepub.com
hosted at
http://online.sagepub.com

Assessing Different Bayesian Neural Network Models for Militarized Interstate Dispute

Outcomes and Variable Influences

Monica Lagazio
University of Kent
Tshilidzi Marwala
University of the Witwatersrand

This article develops and compares two Bayesian neural network models, a more restrictive Bayesian framework using Gaussian approximation and a less restrictive one using a hybrid version of Markov Chain Monte Carlo method (HMC), for the prediction of militarized interstate disputes (MIDs). In addition, to compare and analyze different Bayesian models for international conflict, the authors introduce a new measurement to interpret the relative influence of the model variables on the MIDs. The results indicate that the Gaussian approximation and HMC models are not statistically different in their performance. However HMC correctly recognized a marginally higher number of militarized disputes whose classification is important for policy purpose. On the variable effect, both models indicate similar patter of influences, where the two key liberal variables, democracy and economic interdependence, produce a strong dynamic feedback loop among each other, which greatly increases or decreases the probability of MIDs.

Keywords: *Bayesian; neural network; militarized; interstate dispute; conflict analysis*

Developments in the liberal peace literature increasingly underline the importance of treating international conflicts as complex phenomena often displaying nonlinear and nonmonotonic patterns of interactions. This position challenges the restrictive linear and fixed effect assumptions that have dominated the field by expanding both our theoretical explanations of interstate conflicts and methods commonly applied to conflict data (Beck & Katz, 2001; Bennett & Stam, 2000; Green, Kim, & Yoon, 2001; King, 2001; Oneal & Russett, 2001). From the theoretical side, the existence of interactive relationships among dyadic attributes to influence conflict outcomes has attracted scholars' attention. Beck, King, and Zeng (2000) interpret some key realist variables as creating a prescenario of low or high ex ante probability of military conflict that consequently will or will not trigger the influence of liberal variables. Lagazio and Russett (2004) stress that low levels of key liberal variables

Authors' Note: We thank the National Research Foundation for financial support and Bruce Russett and John Oneal for providing us with the data set and helpful comments. We also thank Evan Govender, Ian Kennedy, and Alex Mintz for further suggestions. Our data, from Russett and Oneal (2001), are available at `http://www.yale.edu/unsy/democ/democ1.htm`.

(economic interdependence, democracy, and shared membership in international organizations) together with distance, relative power, and alliances interact to create multiplicative effects that enhance the likelihood of a dispute. Oneal, Russett, and Berbaum (2003) also find evidence of strong interactions among the liberal variables. In their lag models, trade and peace produce a feedback loop of mutual reinforcement, whereas democracy and shared membership in international organizations increase trade. Pevehouse (2002) reports a link from shared membership in international organizations to democratization. On the interactive relationship between liberal and realist variables, relative power seems also to exert a strong influence on dispute outcomes when nondemocracies are involved, but this influence may be much weaker when democracies settle their disputes (Gelpi & Griesdorf, 2001). Spatial dependence also should be considered. Having states as neighbors that are experiencing conflicts may increase the negative influence of some of the dyadic attributes, whereas having democratic neighbors may reduce it (Gleditsch, 2002). These new insights call for new quantitative techniques for conflict data, and neural network models suggest a robust solution to cope with these complex interactive influences (Beck et al., 2000). However strong doubts still remain on the need to embrace this more complex scenario. Recently, de Marchi, Gelpi, and Grynaviski (2004) find little support for the argument that militarized disputes are generated by a complex and interactive process. This has been rebutted by Beck, King, and Zeng (2004).

To further assess the importance of neural network models to conflict analysis, and to make a contribution to the emerging methodological debate on the need to embrace more complex and interactive models, we focus here on further assessing the Bayesian method put forward by Beck et al. (2000). We do not deal with the issue of neural networks versus statistical analysis here because the Beck et al. (2004) rebuttal to de Marchi et al. (2004) addresses the issue adequately. Our major concern is to address some methodological aspects of Bayesian neural network models that still deserve attention.

First, the Bayesian neural network used by Beck et al. (2000) does not fully untangle the Bayesian formulation. This is because this model still makes some restrictive assumptions. These assumptions may reduce the ability of the Bayesian model to capture the real structure in the data and consequently the complex interactions deemed to exist among the variables (Yoo, Jo, & Jones, 2002). Beck et al. (2000, p. 34) themselves suggest that new approaches should be explored as they become more computationally feasible. To answer this call, we develop a more complex Bayesian neural network model, applied to conflict data, that makes use of the Markov Chain Monte Carlo method (HMC; Neal, 1996). We then compare the HMC model to the Gaussian approximation framework, used by Beck et al. (2000), to assess whether the former can improve model building in conflict analysis.

Finally, concerning issues of interpretation of predictors in neural networks, we investigate the use of a new measurement, called the automatic relevance determination (ARD), that assesses in Bayesian network models the relative influence of the network input variables. Our results once again demonstrate that neural network models can be easily interpreted. Furthermore they strongly support the liberal peace proposition because economic interdependence and democracy emerge as important variables in the model, producing strong dynamic feedback interactions.

The first section of this article will review and critically compare the Gaussian approximation with the HMC model. In sections 2 and 3, we briefly discuss the data utilized and the results of the two Bayesian models. Finally in section 5, we introduce ARD as a measurement to assess variable influences.

Comparing the Bayesian Framework With Gaussian Approximation and HMC

Here we summarize the key ideas of the Gaussian approximation and HMC methods and discuss their advantages and disadvantages for conflict analysis.

Bayesian Framework for Neural Networks

Neural network models implement a nonlinear mapping, or a function approximation, from a set of n input, $x_1, x_2, \ldots, x_n$, to a set of m outputs, $y_1, y_2, \ldots, y_m$. This function approximation is achieved in neural networks using a linear combination of multiple nonlinear functions, represented by neurons. If x is the input and y is the output, then this linear combination of multiple functions may be written as

$$y_k(x) = f_{output}\left(\sum_{j=1}^{m} w_{kj} f_{hidden}\left(\sum_{i=1}^{d} w_{ji} x_i\right)\right) \tag{1}$$

where d is the number of input unit, m is the number of hidden neurons, or functional transformations at the second level of generalization, y_k is the kth output unit, w_{ji} and w_{kj} are the network parameters, in the first and second level of generalization, respectively, going from input i to hidden unit j (see Bishop, 1995, and Ripley, 1994, for a detailed discussion). By increasing the number of functional transformations, the model can approximate any hypothetical relations between the selected explanatory and dependent variables (Hornik & Stinchcombe, 1992). However the major implication of using the flexible functional form provided by the superposition method is that the network functions become nonlinear functions of the network adaptive parameters, the weights (w). Because of the complexity involved in the mathematical structure of the network, the procedure for determining the value of the parameters becomes a nonlinear optimization problem that requires finding efficient learning algorithms to reduce the network overall error function (Bishop, 1995). These algorithms are based on adaptive learning. The idea is to infer the value of the parameters from a training set, D, representing pairs of input-output values with the help of an error function.

When we apply the Bayesian framework to neural networks, instead of searching for a unique optimal value for the unknown weights (as occurs in conventional neural networks), we attempt to model the unknown weights with a probability function, $p(w / D, H)$, which represents the evidence in the data for particular parameter values. According to the Bayesian theorem, given a particular network model, H, a prior distribution for the parameters (w) in the model H, $p(w|H)$, which represents our initial beliefs about the parameters before any data have arrived, and the training data D, then

$$p(w|D,H) = \frac{p(D|w,H)\,p(w|H)}{p(D|H)} \tag{2}$$

In the above expression $p(D|w,H)$ is the likelihood function, which is the probability of the data occurring given the weight parameters, w, and the particular neural network model used, H, whereas the denominator in the expression, $p(D|H)$, is the normalization factor, also called evidence, that guarantees that the total probability is 1. During the network training, the optimal weights, calculated by the network, are the ones that maximize the posterior probability, $p(w|D,H)$. Once the posterior distribution of the weights is known, we can then proceed to calculate the distribution of the output, y, for a given input vector, x, that can be written in the following form (Bishop, 1995; MacKay, 1992a):

$$p(y|x,D) = \int p(y|x,w)p(w|D)dw \tag{3}$$

By applying Equation 3, we can then calculate the output of unseen input data.

The Gaussian Approximation

For complex functions such as neural networks, there are no analytical solutions for the integration in Equation 3. MacKay (1992a, 1992b) has suggested a mathematically convenient Gaussian approximation method to analytically solve the integration in neural networks that conflict scholars have recently applied to conflict data (Beck et al., 2000, 2004; de Marchi et al., 2004). To analytically calculate the values of interest in Equation 3, MacKay approximates the posterior probability of the weights to a Gaussian distribution. If this assumption is made, the posterior probability, $p(w|D,H)$, can be calculated by maximizing the evidence, $p(D|H)$.

What are the implications of using such an approximation? If the approximation is not valid, there are two main implications: The normalization factor or evidence, $p(D|H)$, and the variance (or confidence interval) for the weights and output vector will be inaccurate (Penny & Roberts, 1999). This is because both the calculations of evidence and variance rely on the assumption that the inverse Hessian, A^{-1}, is equal to the covariance matrix of the parameters, which is the case only if the posterior distribution of the parameters is Gaussian (MacKay, 1992a).

In relation to the evidence, incorrect evidence will have an effect on model selection. The reason is that in Bayesian networks, the evidence can be used to select the right level of complexity (or number of functional transformation) for the neural network model (MacKay, 1992a, 1992b; Penny & Roberts, 1999). Neural networks with different levels of complexity are tested, and the ones with the highest evidence are selected as the best models. If the Gaussian approximation is not valid, then using the evidence framework as a solution for model selection in neural networks could result in a model flaw. Instead, in relation to the variance, incorrect calculations of confidence interval will have an indirect effect on model accuracy.

Is the approximation valid for conflict data? For one hand, because of the central limit theorem, Gaussian approximation is thought to perform well in reasonably large sample sizes. This is because the posterior distribution tends to become Gaussian as the sample size increases (Walker, 1969). Therefore the Gaussian assumption should still be valid in conflict analysis because conflict data allow for the large size to be utilized.[1] On the other hand, recent empirical works seem to indicate that the Gaussian approximation cannot handle the presence of multiple modes in the parameters that are likely to be present in complex, real-world data modeling (Yoo et al., 2002). From previous neural network applications to conflict data, it can be derived that the weight space is quite complex and presents multiple local peaks (Garson, 1991; Lagazio & Russett, 2004; Schrodt, 1991). Therefore we can assume that there are indeed multiple modes in the posterior distribution of the parameters of conflict models. Furthermore the more inputs (the explanatory variables) we add to the network, the more multiple modes we should expect in the posterior probability of the parameters as a result of the increasing complexity in the data. These more complex models could present a harder test for the Gaussian approximation.

Bayesian Framework With HMC

A more robust method to solve Equation 3 is through direct integration by HMC (Neal, 1996). Neal has suggested this approach and further improved it by using a hybrid version of HMC that significantly reduces the calculation time of the simulation by efficiently sampling the posterior distribution (Neal, 1992).

The HMC makes no assumption about the form of the posterior distribution because it performs a robust integration based on sampling. The main idea of this method is to approximate the integration (Equation 3) by the mean of the function, $f(w)$, sampled from the posterior distribution of the weights, $p(w|D)$. Because independent sampling from the high dimensional space of the network weight is almost impossible, a dependent variable called Markov chain is created that has $p(w|D)$ as its equilibrium distribution and whose estimate will still converge to the true value as n increases (Neal, 1996). To generate such Markov chain, the HMC algorithm is used. The HMC method uses the gradient of the neural network error to ensure that the network samples throughout regions of higher probabilities. This causes the HMC to avoid the random walk associated with traditional HMC methods. Sampling using the HMC is conducted by taking a series of trajectories and then either accepting or rejecting a resulting state at the end of each trajectory (see Neal, 1992). From the output of the accepted weight vectors, the probability distribution of the final output is obtained. The variance (or confidence interval) of the weight and output is then given by the actual variance of the weight and output vector across the selected samples.

Although the HMC method is a more robust method for integrating Equation 3, some problems still remain. HMC remains more computationally expensive than the Gaussian approach (Marwala, 2001). Another problem lies with assessing when the HMC simulation has reached equilibrium (see Cowles & Carlin, 1996) and how many selected samples, n, are needed to achieve convergence. This is equivalent to asking the question: How long do I need to run the simulation before stopping? Deciding when to stop the simulation is a difficult task, and existing methods are not easy to interpret (Vivarelli & Williams, 2001). However even if the HMC method still presents some problems, it has proven to be a more robust approach than the Gaussian approximation when applied to complex data (Vivarelli & Williams, 2001).

Data and Models

To be consistent with the latest work in conflict studies, we use for our analysis the population of politically relevant dyads for the cold war and immediate post-cold war period, from 1946 to 1992, as described extensively and used by Russett and Oneal (2001). The unit of analysis is the dyad-year. There are a total of 27,737 cases in the population, with 26,845 nondispute dyad-years and 892 dispute dyad-years. We chose the politically relevant population (contiguous dyads plus all dyads containing a major power) because it sets a hard test for prediction.[2]

For explanatory and dependent variables, we follow the theoretical approach of the liberal dyadic program (Russett & Oneal, 2001). Consistent with this dyadic perspective, the analysis includes seven independent variables (network inputs): two liberal and five realist variables (data are available at `http://www.yale.edu/unsy/democ/democ1.htm`). The dependent variable (dispute), or network output, is 1 if a militarized interstate dispute was begun and is 0 otherwise. Only the initial year of the militarized conflict is included because our concern is to predict the onset of a conflict. The realist variables include allies, a binary measure coded 1 if any form of military alliance links the members of a dyad. Contingency is also binary, coded 1 if both states are geographically contiguous. Distance is an interval measure of the log distance between the two states' capitals. Majorpow is a binary variable coded 1 if either one or both states in the dyad are a major power, and capability ratio measures the dyadic balance of power on an interval scale. The first liberal variable, democracy, is a 21-point scale variable measuring the level of democracy in the less democratic state in each dyad. Dependence is a continuous variable measuring the level of economic interdependence (dyadic trade as a portion of a state's gross domestic product) of the less economically dependent state in the dyad. We lag all the independent variables by 1 year to make any inference of causation temporally plausible.[3]

The data set is used to generate two different sets, training and testing, with the training set used only for training and the test set to assess out-of-sample accuracy. The validation set for training is not used because we are pursuing a Bayesian approach to neural network training that does not overfit the model (MacKay, 1992b; Neal, 1996). The size of the training sets consists of 500 dispute and 500 nondispute dyad-years, whereas the test data consist of 392 dispute and 26,345 nondispute dyad-years.

To select the optimal number of hidden neurons (M) and transformation functions for the two models, we used a genetic algorithm (GA).[4] The network was trained several times using scaled conjugate gradient method and GA used to perform a global search in the solution space and select the best solution.[5] Four activation functions were considered (linear, logistic, hyperbolic tangent, and soft-max), and M was restricted to be maximum $\frac{n}{r(i+o)}$, with n being the number of samples in the training, i and o representing the number of neurons in the input and output layer, respectively, and r being a constant set by the noise level of the data (Klimasauskas, 1992). A GA population of 20 was used. The GA identified $M = 10$, logistic function in the output layer (f_{output}in Equation 1), and the hyperbolic function in the hidden layers (f_{hidden} in Equation 1) as the optimal model. The input parameters were normalized, and the training patterns were entered in a shuffled order to avoid bias. Both the Gaussian approximation and the HMC methods use the cross-entropy function as the error function because of its convenience for classification problems (Bishop, 1995) and assume a normal prior distribution for the model parameters. Finally for the HMC approach, the sampling phase was run for 450 iterations. The first third of these were discarded with the aim to allow the network to reach equilibrium (Gelman, Carlin, Stern, & Rubin, 1995), whereas 150 samples were accepted.

Figure 1
The Receiver Operating Characteristics of the Classification of Militarized Interstate Disputes

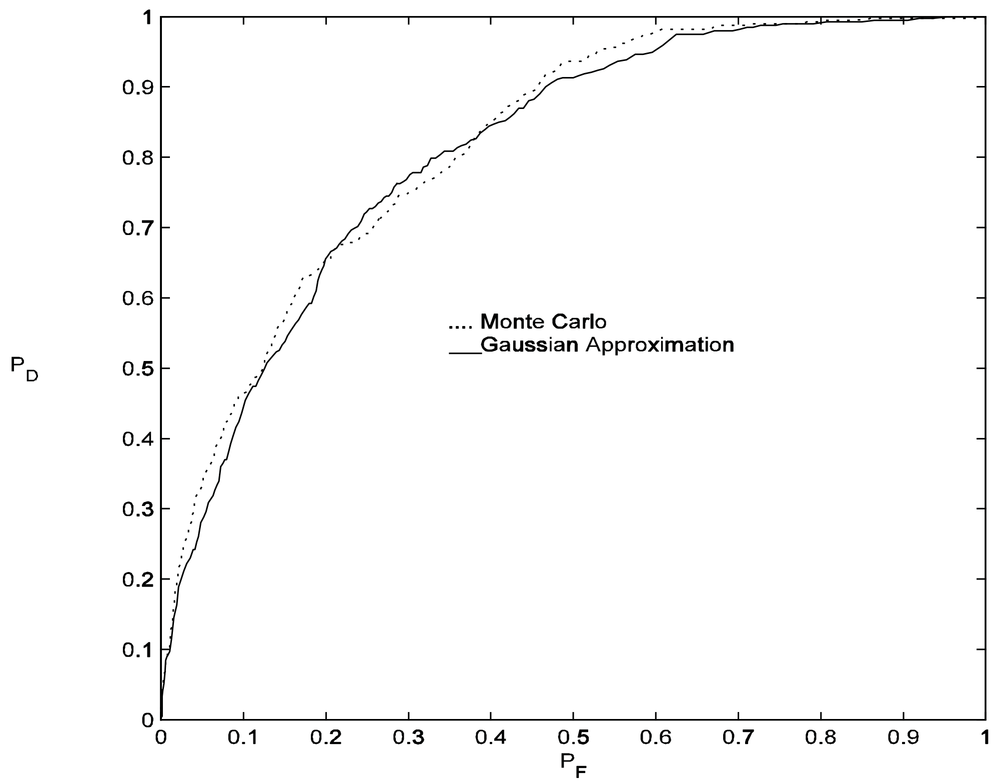


Table 1
Classification Results

Method	True Conflict	False Conflict	True Peace	False Peace
Gaussian Approximation	278	114	19,462	6,883
Hybrid Monte Carlo	286	106	19,494	6,851

Results and Discussion

We now discuss the out-of-sample result of the Gaussian approximation and HMC. To evaluate the two models' performances, we use the receiver operating characteristics (ROC) graphs, whose results are shown in Figure 1. We chose the ROC method because the ROC curves investigate the tradeoff between false positive and false negative for a variety of predictive thresholds and do not penalize models whose prediction is biased too high or too low. In the ROC curves, the x axis gives the proportion of disputes correctly predicted, whereas the y axis provides the proportion of nondisputes correctly predicted for different thresholds. The general idea is that any threshold used as a cutoff value between disputes and nondisputes will correspond to a single point on the ROC curve (Zweig & Campbell, 1993). The area under the ROC curve indicates how good the classifier is. If the area under the ROC curve is 1, then the classifier has classified all cases correctly, whereas if it is 0.5, the classifier is as good as a random guess. Traditional confusion matrices for the models are also provided in Table 1.

In Figure 1, both classifiers give an area under a ROC curve of 0.82, which is a good classification rate. This result indicates that both methods provide similar accuracy. Furthermore because the curves do not present stochastic dominance (i.e., when one curve is always above the other), this result also suggests that neither model is better than the other. It is also worth noticing that the ROC implies a particular normative assumption about the cost of misclassification. Increasing true positive is regarded as producing the same benefit as increasing true negative. But this may not be the case in conflict analysis, where the cost of misclassifying disputes (true positive) can be much higher than nondisputes (true negative). In the absence of stochastic dominance, the result of the confusion matrix in Table 1 can then provide useful insights. This table shows that HMC performs marginally better when recognizing interstate disputes. In fact when the accuracies of the two methods were calculated on the basis of the true positive rate (the proportion of disputes correctly identified), the HMC gives a true positive rate of 73%, whereas the Gaussian approximation gives a rate of 71%. In relation to the true negative rate (the proportion of nondisputes classified correctly), both methods perform the same, with 74% accuracy. This marginal improvement is quite important from a policy view. The HMC model identifies eight more politically costly and dangerous dispute dyads than does the Gaussian approach, therefore allowing eight more dispute cases to be controlled and possibly solved before they actually occur. Furthermore this is done without increasing the rate of false positives. HMC reduces the false positive of 32 cases compared to Gaussian approximation, therefore allowing no waste of valuable resources on controlling disputes that are unlikely to happen.

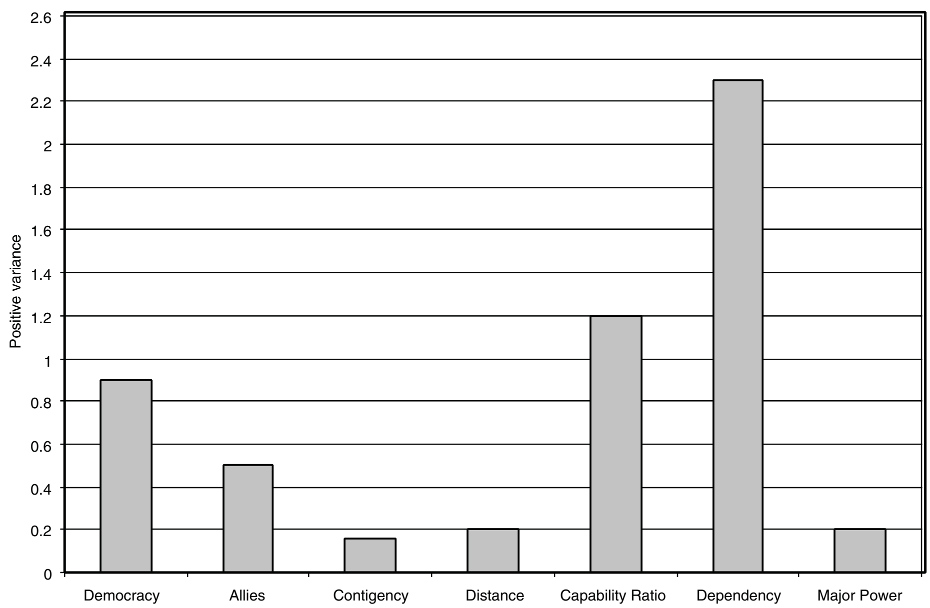
On the basis of these results, the HMC emerges as being marginally more accurate than the Gaussian approximation. This is primarily because of the fact that the Gaussian approximation is generally not as valid as the HMC approach because it is not as efficient in handling multiple modes in the parameters. It is also worth noting that although the Gaussian approximation performs quite closely to the HMC model with these data, this may not be the case if additional explanatory variables are added as input to the Bayesian network. As mentioned earlier, additional explanatory variables could increase the presence of multiple modes in the parameters, thus providing a much harder test for the Gaussian approach.

Model Interpretation: The Liberal Dynamic Feedback Loop

To assess the influences of the model variables on the interstate militarized disputes, we utilize a new measurement, ARD. Contrary to recent critiques that emphasize interpretability issues in neural networks (de Marchi et al., 2004), the ARD method demonstrates once more that neural networks are not black boxes. Indeed the Bayesian framework offers this useful feature, introduced by MacKay (1995) and Neal (1996), to help identify important variables.

ARD provides an estimate of the relative overall influence exerted by each input variable on the output. This value tries to integrate direct and interactive influence, which each input

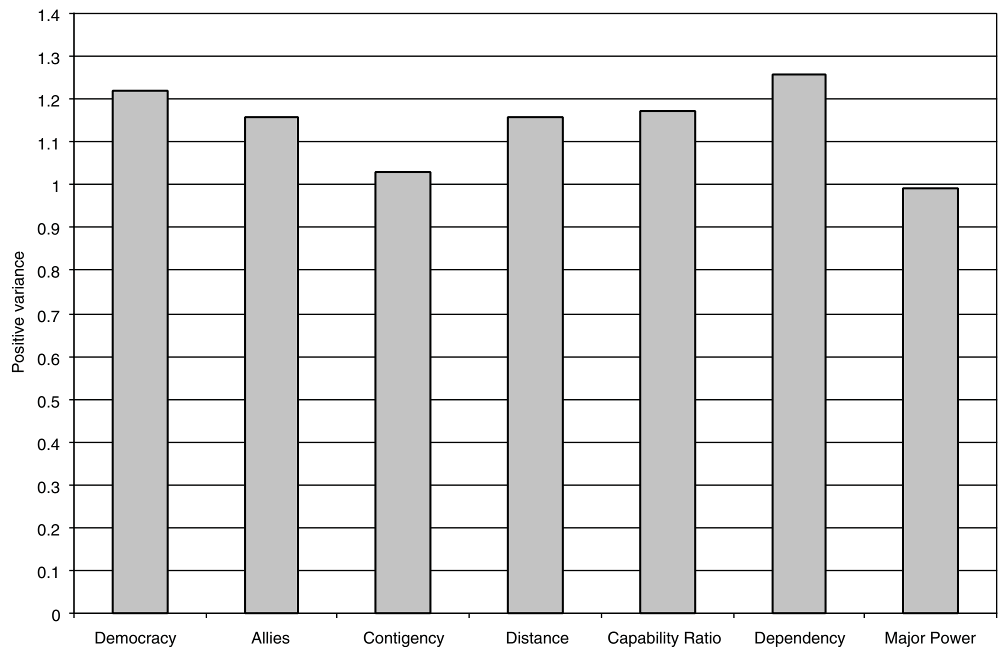
Figure 2
A Graph Showing the Relevance of Each Variable With Regard to the Classification of Militarized Interstate Disputes for the Gaussian Approximation Model



exerts on the output and between each other, into a single measurement. To do that, the ARD method introduces an additional adjustable parameter, α_k, for each group of weights, k, that connects one input unit to all of the units in the next layers. This parameter, α_k, controls the size of the group of weights k: Weights connected to irrelevant inputs are automatically set to small values. The larger α_k, the smaller the value of the weight belonging to group k, thus the less important the connected input. The smaller α_k, the larger the value of the weight group, thus the more important the connected input. During training, α_k is also inferred by maximizing the posterior probability distribution on the basis of the data D and model H similar to the weight calculation in Equation 2. Periodically the training is stopped and α_k updated.

Figures 2 and 3 summarize the ARD results for the Gaussian approximation and HMC method, respectively. In both figures, the inverse of α_k was calculated, therefore higher values indicate more important inputs. Both Bayesian models identify dependence and democracy as key variables. Economic interdependence exerts the greatest influence in both models, whereas democracy is identified as the second and third most important variable in the HMC and Gaussian approximation, respectively. This supports the liberal thesis that the state with the lower level of interdependence and democracy in the dyad has the major effect on dyadic relationships. Furthermore these results also indicate that these two variables exert a strong interactive influence on each other (this is because ARD also captures indirect influence). This confirms recent findings (Lagazio & Russett, 2004; Oneal et al., 2003; Reuveny & Li, 2003) and the theoretical argument that a strong feedback loop exists between the two key liberal variables (Papayoanou, 1997; Weede, 1996). Besides exerting direct influence on the

Figure 3
A Graph Showing the Relevance of Each Variable With Regard to the Classification of Militarized Interstate Disputes for the Hybrid Monte Carlo Model



state dyadic behavior, economic interdependence and democracy interact between each other to increase or decrease the likelihood of militarized conflicts.

Other important variables are capability ratio and to a certain extent allies and distance. Once again this confirms recent positions that see the realist variables as mediating the influence of the liberal variables by providing constraints or opportunities for state action (Beck et al., 2000; Kinsella & Russett, 2002; Lagazio & Russett, 2004; Russett & Oneal, 2001).

Overall the result supports theories of the liberal peace identifying both democracy and economic interdependence as key variables in relation to peace and war. The ARD values also appear to indicate that the two liberal variables have more power, both direct and indirect, than in previous analyses (Lagazio & Russett, 2004). Furthermore the values confirm that a strong feedback loop exists between the liberal variables and between the liberal and realist variables. Once again the need for complexity, both from a methodological and theoretical prospective, is apparent. Relationships across the variables do appear to be highly nonlinear and contingent. Economic interdependence and democracy play important direct and indirect roles in producing war and maintaining peace. Their influence is greatly strengthened by a dynamic feedback loop existing between them and by their interactions with the realist variables of balance of power, geographical proximity, and alliances.

Conclusions

Here we have developed and compared two Bayesian frameworks, the Gaussian approximation and a hybrid version of HMC, for the analysis of interstate militarized disputes. The result shows that neither model is better than the other. However the less restrictive model, HMC, is marginally better in the classification of the politically costly dispute dyads. This result, however marginal, seems to indicate that the HMC model is better equipped to deal with conflict data because it can handle multiple modes likely to be present in the parameters of conflict models.

Concerning the structure of influences, our analysis once again shows that relationships affecting disputes are highly nonlinear and that strong interactions exist among the variables. These results are similar in both models, thus indicating that robust findings on variable influences can be provided by the Bayesian frameworks. Consistent with recent theory and findings, our results show a strong dynamic feedback loop as the two key liberal variables, democracy and economic interdependence, reinforce each other. This feedback loop appears to be even stronger than in previous analyses. A significant interaction effect also exists among these two liberal variables and three key realist variables: capability ratio, allies, and proximity. Even with the limited number of variables employed here, highly interactive relationships already emerge as a consistent feature in conflict analysis. Ignoring this empirical, methodological, and theoretical position could only delay future developments in the field.

Notes

1. Empirical findings suggest that the sample size should exceed the number of model parameters by a factor of at least 5 for the Gaussian approximation to be valid (Penny & Roberts, 1999).

2. By focusing only on dyads that either involve major powers or are contiguous, we test the discriminative power of the Bayesian neural networks on a difficult set of cases. The neural network system is fed with only highly informative data because every dyad can be deemed to be at risk of incurring a dispute, yet it is harder for the network to discriminate between the two classes because the politically relevant group is more homogeneous (e.g., closer, more interdependent) than is the all dyad data set.

3. We have not included the number of years since the last conflict as a variable because neural networks do not require independent observations and thus deal better with the suspected influences that militarized events exercise on each other (Sarle, 1994).

4. We do not rely only on the evidence to select the model for the Gaussian approximation, but instead we utilize the genetic algorithm to assess evidence results. As MacKay (1992b) suggests, the evidence can be tested by looking at the correlation between the evidence and test error. By doing so, we reduce the bias of the Gaussian approach.

5. Genetic algorithm was implemented in this article through performing simple crossover, binary mutation, and roulette wheel reproduction. The details of these can be found in Holland (1975).

References

Beck, N., & Katz, J. (2001). Throwing the baby out with the bathwater. *International Organization*, *55*, 487-475.

Beck, N., King, G., & Zeng, L. (2000). Improving quantitative studies of international conflict: A conjecture. *American Political Science Review*, *94*, 21-35.

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*, 379-389.

Bennett, D. S., & Stam, A. C., III. (2000). Research design and estimator choices for analyzing interstate dyads: When decisions matter. *Journal of Conflict Resolution*, *44*, 653-679.

Bishop, C. (1995). *Neural networks for pattern recognition*. Oxford, UK: Clarendon.

Cowles, M. K., & Carlin, B. P. (1996). Markov Chain Monte Carlo converge diagnostics: A comparative review. *Journal of the American Statistical Association*, *91*, 883-904.

de Marchi, S., Gelpi, C. F., & Grynaviski, J. D. (2004). Untangling neural nets. *American Political Science Review*, *98*, 371-378.

Garson, G. D. (1991). A comparison of neural network and expert systems algorithms with common multivariate procedures for analysis of social science data. *Social Science Computer Review*, *9*, 399-433.

Gelman, A., Carlin, J. B., Stern, H. S., & Rubin, D. B. (1995). *Bayesian data analysis*. London: Chapman and Hall.

Gelpi, C., & Griesdorf, M. (2001). Winners and losers? Democracies in international crisis, 1918-94. *American Political Science Review*, *95*, 633-647.

Gleditsch, K. S. (2002). *All international politics is local: The diffusion of conflict, integration, and democratization*. Ann Arbor: University of Michigan Press.

Green, D., Kim, S. Y., & Yoon, D. (2001). Dirty pool. *International Organization*, *55*, 441-468.

Holland, J. (1975). *Adaptation in natural and artificial systems*. Ann Arbor: University of Michigan Press.

Hornik, K., & Stinchcombe, M. (1992). Multilayer feed forward networks are universal approximators. In H. White (Ed.), *Artificial neural networks, approximation and learning theory* (pp. 12-28). Oxford, UK: Blackwell.

King, G. (2001). Proper nouns and methodological propriety: Pooling dyads in international relations data. *International Organization*, *55*, 497-507.

Kinsella, D., & Russett, B. (2002). Conflict emergence and escalation in interactive international dyads. *Journal of Politics*, *64*, 1045-1068.

Klimasauskas, C. C. (1992). Applying neural networks. In R. R. Trippi & E. Turban (Eds.), *Neural networks in finance and investing* (pp. 47-72). Chicago: Probus.

Lagazio, M., & Russett, B. (2004). A neural network analysis of militarized international disputes, 1885-1992: Temporal stability and causal complexity. In P. Diehl (Ed.), *The scourge of war: New extensions on an old problem* (pp. 28-60). Ann Arbor: University of Michigan Press.

MacKay, D. J. C. (1992a). Bayesian interpolation. *Neural Computation*, *4*, 415-447.

MacKay, D. J. C. (1992b). The evidence framework applied to classification problems. *Neural Computation*, *4*, 720-736.

MacKay, D. J. C. (1995). Bayesian non-linear modelling for the 1993 energy prediction competition. In G. Heidbreder (Ed.), *Maximum entropy and Bayesian methods, Santa Barbara* (pp. 221-234). Dordrecht, the Netherlands: Kluwer.

Marwala, T. (2001). Scaled conjugate gradient and Bayesian training of neural networks for fault identification in cylinders. *Computers and Structures*, *79*, 2793-2803.

Neal, R. M. (1992). *Bayesian training of backpropagation networks by the hybrid Monte Carlo Method* (Tech. Rep. CRG-TR-91-1). Toronto, Canada: University of Toronto.

Neal, R. M. (1996). *Bayesian learning for neural networks. Lecture notes in statistic* (No. 118). Berlin, Germany: Springer-Verlag.

Oneal, J. R., & Russett, B. (2001). Clear and clean: The fixed effects of democracy and economic interdependence. *International Organization*, *52*, 469-486.

Oneal, J. R., Russett, B., Berbaum, M. (2003). Causes of peace: Democracy, interdependence, and international organizations, 1885-1992. *International Studies Quarterly*, *47*, 371-393.

Papayoanou, P. (1997). Economic interdependence and the balance of power. *International Studies Quarterly*, *41*, 113-140.

Penny, W. D., & Roberts, S. J. (1999). Bayesian neural networks for classification: How useful is the evidence framework? *Neural Networks*, *12*, 877-892.

Pevehouse, J. (2002). Democracy from the outside in? International organizations and democratization. *International Organization*, *56*, 515-550.

Reuveny, R., & Li, Q. (2003). The joint democracy-dyadic conflict nexus: A simultaneous equation model. *International Studies Quarterly*, *47*, 325-346.

Ripley, B. D. (1994). Neural network and related methods for classification. *Journal of the Royal Statistical Society*, *56*, 409-456.

Russett, B., & Oneal, J. (2001). *Triangulating peace: Democracy, interdependence, and international organizations*. New York: Norton.
Sarle, W. (1994). Neural network and statistical models. In *Proceedings of the 19th Annual SAS Users Group International Conference* (pp. 1538-1550). Cary, NC: SAS Institute.
Schrodt, P. (1991). Prediction of interstate conflict outcomes using a neural network. *Social Science Computer Review*, *9*, 359-380.
Vivarelli, F., & Williams, C. (2001). Comparing Bayesian neural network algorithms for classifying segmented outdoor images. *Neural Networks*, *14*, 427-437.
Walker, A. M. (1969). On the asymptotic behavior of posterior distribution. *Journal of the Royal Statistical Society B*, *31*, 80-88.
Weede, E. (1996). *Economic development, social order, and world politics*. Boulder, CO: Lynne Rienner.
Yo, Y. S., Jo, C. Y., & Jones, C. N. (2002). Compositional prediction of creep rupture life of single crystal in base superalloy by Bayesian neural network. *Materials Science and Engineering*, *A336*, 22-29.
Zweig, M. H., & Campbell, G. (1993). Receiver-operating characteristic (ROC) plots: A fundamental evaluation tool in clinical medicine. *Clinical Chemistry*, *39*, 561-577.

Monica Lagazio is lecturer in international relations in the Department of Politics and IR at Kent University. She works in conflict research, security, artificial intelligence models, and complexity. She can be reached at *m.lagazio@kent.ac.uk*.

Tshilidzi Marwala is associate professor in computational intelligence in the School of Electrical and Information Engineering at the University of the Witwatersrand. His research interests are neural networks, evolutionary computing, dynamical systems, and fuzzy logic. He can be reached at *t.marwala@ee.wits.ac.za*.