ELECTION RESULT FORECASTING USING TWO LAYER PERCEPTRON NETWORK

G. S. Gill
Lecturer, Institute of Instrumentation Engineering,
Kurukshetra University Kurukshetra, India
E-mail: gaganssgill@yahoo.co.in

ABSTRACT

Neural networks are increasingly used to solve highly non linear control problems. The current paper addresses the problem of forecasting the result of general elections in India. The neural network is first made to learn and then the trained network is made to forecast the result of the election. While training the network minimal disturbance principle was followed, which suggests that during training it is advisable to inject new information into the network in such a manner that it disturbs the stored information to the smallest extent possible.

Keywords: *Election Result, Neural Network, Forecasting.*

1. INTRODUCTION

Election forecasting is one of the most interesting challenges for political scientists. The current paper describes artificial neural network model for forecasting election results. An artificial neural network mimics the ways in which biological nervous systems process the information. The brain contains a complex set of interconnected "processors" or neurons that work together towards problem solving. Experience in problem solving is vital to how well this network works – the brain works more efficiently and effectively as it learns to recognize familiar patterns. In a similar way neural networks can be trained to recognize patterns in the data that other computing and statistical techniques might fail to identify. Once trained the network can then be used to make 'forecasts'. Although now widely used in a growing number of fields there is little evidence of their application in political science [1],[2]. A paper presented by Gorban and Waxman at a meeting of the World Congress on Neural Networks reported a neural network model for forecasting the outcome of US Presidential elections[3]. It correctly predicted that Clinton would beat George Bush senior in the 1992 elections.

The current problem aims to develop a model that can forecast the outcome of Lok Sabha elections in the India. The model developed is still in preliminary stage and could not take into account the all peculiarities of the Indian voting system, particularly the operation of the Electoral College. Since in India do not have two party system, so we could not account the factors such as vote splitting. Also, the local factors could not be fully accounted for.

2. EXISTING MODELS

Political scientists working globally have been active in developing models for forecasting elections. Existing forecasting models fall into one of four categories. First, there is the evidence gathered from opinion polls although some analysts believe that an opinion poll is merely a snapshot of opinion frozen in time and it may or may not carry over onto election day itself and it seldom makes allowance for the particular voting situation confronted by each respondent in his or her constituency. This is just as well because the recent track record of the opinion polls, both in the UK [4],[5] and USA is not particularly good – even at the level of predicting the correct winner.

Another approach uses a set of indicators to model how voters might behave at a future election. Many, of these indicators are associated with economic circumstances and reforms either at the national and/or personal level. Within the British context Sanders has developed a Personal Expectations model which uses an individual's perception of their own economic well-being and associates that with the level of Conservative party support as measured by the opinion polls [6]. Sanders' model has met with considerable success. Responding to Sanders' work Clarke and Stewart [7] believe that national economic expectations may work as well if not better than the personal expectation approach. Many of the models developed by US forecasters adopt a similar approach, employing a battery of economic indicators as well as measures of party popularity [8],[9].

A third approach, developed relatively recently, eschews opinion poll data entirely and takes as its basic resource the results from local government by-elections [10]. In addition to this the results from the provincial assembly elections can serve as a useful indicator. In such a case, the local issues which directly or indirectly govern the choice of the voter are also adequately addressed. The basic premise behind this model is that voting in local elections requires people to make real choices involving some personal cost in voting contrasted with the decision without consequences of responding to a survey questionnaire. The fact that such elections are frequent and spread relatively evenly throughout the year means that some estimate can be made about the dynamics of electoral opinion. This model has met with mixed success.

One of the difficulties in the first two approaches described above is that reliable polling data on the specified indicators is sometimes unavailable and in any case polling information dates from the post independence period only. Moreover till the year 1977, there was more or less one party culture only with no other party posing sufficient threat to form the government instead of Indian National Congress. Only in late seventies India had seen an emergence of any major opposition party. So the availability of data for the training the network is for just a short span of time. Even the by-election model suffers equally from the unavailability of data. Consequently all three approaches described thus far suffer from the same difficulty. One cannot single out as to which approach would yield the best results. One can only expect to find a correlation between indicators of economic prosperity (national or individual based) and perceptions of the governing party's performance.

A fourth approach avoids such difficulties by using prior election results to generate forecasts. Writing some time before the 1996 presidential election Helmut Norpoth published his forecast that Clinton would be re-elected [11]. At the time he made his forecast Clinton's popularity was sufficiently low that it was unclear whether he would actually be chosen as the Democrat candidate. Norpoth's approach was to use an autoregressive model that depicts the present as a weighted sum of certain previous election outcomes. As Norpoth notes, his model has two distinct advantages over others. First, it provides a long-range forecast, one that can be made as soon as the presidential race has finished. Second, it is cheap since no evidence regarding voters' prospective choices is required. For the 2000 election Norpoth introduced a variation. In addition to past elections he also concentrated on presidential primaries and discovered that if the incumbent party's eventual candidate performed well at these initial contests then that was a vital indicator in predicting whether that party would succeed at the following presidential election. Retrospectively, the model forecast correctly all presidential elections since 1912 with a single exception – the extremely close Kennedy-Nixon race in 1960. However this modified model did not provide an accurate forecast of the 2000 election, predicting that Gore would defeat Bush taking 55% to 45% of the major party vote [12].

3. NEURAL NETWORKS APPROACH

Mathematical explanations can be provided for the difficulties encountered in forecasting. One factor is the dynamical unpredictability of complex system behaviour over a given time period. A second factor is the probabilistic and non-deterministic nature of the subject where it is not known which of many variables exert the biggest influence upon the system under review. New mathematical and computational methods known as Neural Network (NN) techniques, based on ideas and principles arising from the neurosciences, have been developed precisely to deal with such difficult problems [13]-[16]. These techniques can be used to improve forecasting. The technique is essentially based on the idea of learning by example. A Neural Network called a Multi-Layer Perceptron, developed by Rosenblatt in the 1950s [17] is one of the most promising and widely used. The perceptron itself consists of several layers of elements or artificial neurons (Figure 1). The first layer of neurons forms the perceptron's input. The connections from the input layer deliver the information to the internal, hidden layer, and connections from neurons of the hidden layer send signals to the neurons of the output layer. Each connection between two neurons has specific connection strength, which is a measure of its efficiency. The connection strengths are adjusted during the learning process in such a way as to provide the correct relationship between input and output signals. The number of neurons in each layer varies according to the problem under consideration.

There are two main stages in the development of a Neural Network. The first stage of the process is called training. Usually, the set of available data is split into two independent (non-overlapping) subsets. One subset is called the training set and with this the NN effectively

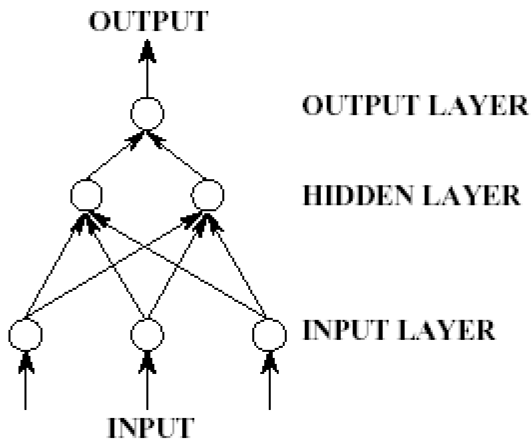

Figure 1. *Perceptron*

learns by example. As we sequentially present the elements of the training set to the input of the NN, we adapt the connection weights between neurons in such a way that the corresponding output signal becomes the desired target. The training set, employed so that the perceptron can learn from examples, contains pairs of signals: an input signal and the target desired output for this input signal. During the learning process, connection strengths are modified such that a given set of inputs result in a known output. At the beginning of the training, the output is very different from the desired output because the training starts with random connection strengths but, sequentially, presentations of the training set improve the perceptron's performance. The adaptation of connection strengths increases the number of correct outputs and successful training can give a zero error, i.e. the presentation of any input from the training set results in the appearance of the desired output signal. This approach is termed 'supervised learning'.

The training process is rather long and requires many repetitions of elements within the training set. The mean least squares error for the whole training set effectively regulates this training process. The sum of squares for differences between the actual NN output and the desired output for the training set is calculated for each step of the training procedure. There are many different training algorithms - learning rule/strength adaptation rule, including the back-propagation rule where the error information propagates back to make an adaptation/modification of connection strengths. The error of the back propagation algorithm is used to update connection strengths in such a way that the neural network, now equipped with updated connection strengths produces a smaller error for subsequent presentations of the same training data. These algorithms are in effect based on the minimization of the least squares error. The training process stops if this sum is small enough. Occasionally, it becomes impossible to reduce the error value further, although the required threshold for that value has not been reached. In such cases the learning is acknowledged as partial and imperfect. However, if the NN has a small mean least squares error we can proceed to the second stage of testing the network.

Recall, that the set of data for testing the network is independent of the training set. During this testing phase the connection weights of the various elements comprising the input signal are fixed and unchangeable. Once again, we sequentially present the elements of the testing set to the input of the NN and compare the output value with the known value. At this stage we discover whether the training set was representative of the entire data set and that the NN has learnt sufficiently well that it is able to match its output signal to the known output. Information generalization is an important feature of any neural network. The training set, with a very limited number of examples, is used to train the neural network. We anticipate that after training, the neural network will also deal correctly with other data (testing set), which is not included in the training. Due to the generalization property, the neural network will respond correctly to the presentation of new input signals from the testing set. In cases with a good testing result we can conclude that the neural network has the generalization property and, if we continue to present new input signals to the input, then the output hopefully will be correct.

Clearly, training the network is a critical part in developing an effective forecasting tool. The result of training depends on the choice of the network parameters. The number of input and output neurons is fixed according to the considered problem, while both the number of neurons in each hidden layer and indeed the number of hidden layers is variable. Increasing the number of hidden neurons means an increase in the number of connection weights and, hence, the number of free parameters. If the number of parameters is large then the over fitting problem can appear. This means that the NN delivers a good performance for the training set but that the generalization is rather poor and, as result of that, the testing results are also poor. Choosing the correct number of parameters requires experience of dealing with NN and many trials should be undertaken to choose the appropriate number of elements which optimize NN performance (see, for example, Beck et al. 2000 where the special in sample learning procedure for the best choice of hidden neuron number has been developed).

The data that we are using to forecast the result of a Lok Sabha election is based on information regarding all previous general elections from. Normally, when developing Neural Networks it is usual for the number of cases to be in the hundreds, if not thousands, and so our data set is atypical for this particular method. This meant that some modifications to the training and testing process were required.

As stated earlier, a neural network requires two basic types of data – input and output signals. For our purpose the output signal is the party winning the election. For all general elections from 1977 onwards the network was told whether the P (power) party had been returned or whether the O (opposition) had been successful. The corresponding result was coded as zero if the (O) opposition party won or 1 if the (P) Party in power had been re-elected. Of course, during certain periods of our electoral history such neat answers are somewhat arbitrary, particularly during periods of dislocation to the existing party system and/or hung parliaments, but after consultation we felt that our coding of these events was at least defensible.

The input signal consisted of ‘expert’ responses to a battery of the following ten questions that related to the period between two general elections:

1. Do you think that the government kept the promises it made before the polls?
2. In your view has the economy improved under this government?
3. Do you think that this government has been associated with political scandal, scams and controversial judgments?
4. Has peace, law and order prevailed during the regime of this government?
5. Has the government spent sufficient funds for welfare and development?
6. Has there been significant social tension, e.g. industrial strikes, during this parliament?
7. Overall, has the media been hostile to the government?
8. Overall, has the media been hostile to the official opposition?
9. Overall, in your view has the Prime Minister performed well during this parliament?
10. Overall, in your view has the Leader of the Opposition performed well during this parliament?

The questions were designed to identify a range of issues that might or might not become salient at election times. An expert was requested to generate answers to our set of questions. The expert was only given the choice of answering ‘Yes’ or ‘No’, although a more sophisticated NN could permit a greater range of responses. In effect, we have a data set corresponding to the responses from our expert consisting of 9 cases based on the general elections between 1977 and 1999 and where each case has an input vector of 10 binary numbers and the target output (0 or 1). In next section the experimental results of the network training are reported.

4. TESTING AND COMPUTATION

The data from the expert was used to train the Neural Network. A two layer network- comprising of input and output layers only, with fixed rate was used for this purpose. We decided that the training set would comprise data for 8 general elections with a single election comprising the set for testing. For example, all general elections between 1977-2004 become the set for training with the 2004 result the set for testing. An alternative training set could be created by selecting all general elections between 1977-2004, with the 1999 excluded and used as the set for testing. In fact, we repeated the process of selecting a different election for inclusion as the testing set a total of 9 times. With each trial a separate NN is created, trained and tested using its own unique sets. Of course, we should recognize that despite the uniqueness of each training set, they are not entirely independent of one another, because the pair-wise overlap is significant. After that we repeat the process a total of 9 times, using for testing a different general election each time. Finally, we used each of these trained NN to forecast the result of the next general election. The test results are shown in Table 1.

Table 1. *Test results*

Total number of neural networks subjected to testing	No of neural networks with correct testing result	No of neural networks with incorrect testing result
9	6	3

In the above case the number of NN that forecast the output correctly was 6, with a further 3 NNs producing incorrect test results. Elections which consistently proved to be the most difficult to predict and whose results were least compatible with the views of our experts were 1977, 1984 and 1989. We suspected that part of the problem was at least associated with the prior circumstances relating to these particular elections. The phase saw alarming growth of terrorism in the country, along with the assassination of the sitting prime minister, application of president's rule in certain provinces, host of parties boycotting the general elections. These all parameters couldn't be accounted for when we are analyzing the things in terms of yes/no. moreover the 1989 elections saw the beginning of the era or hung parliaments, where no single party formed the government, the phase which has continued since then. Moreover, a small political party or a splinter group which supported the party in power may form the post poll or pre poll alliance with the other party in the next elections; this cannot be accounted for by our network.

5. CONCLUSIONS

Since this study is based on the data for only 9 general elections, so a little claim for generality can be made for any conclusions. Before any generalization could be drawn from this sort of study a much larger amount of would have to be examined on similar terms. This paper may be considered as a first attempt to use neural network for the purpose of forecasting general election results and that the future work will include more data and would take into account the constitution of Electoral College more precisely. In addition to this, data from local by-election and provincial assembly elections can significantly improve the accuracy of forecasting. Moreover, data set can be increased, by increasing the number of experts used to generate the data. Obtaining the views of different experts to help train and test different networks clearly demonstrates that the subsequent network will reflect the characteristics of each expert. There is a clear danger that if those views are idiosyncratic then the NN will reflect that particular trait.

REFERENCES

[1] Beck N, King G, & Zeng L. (2000). Improving Quantitative Studies of International Conflict: A Conjecture. *American Political Science Review, 94,* 21-35.

[2] Eisinga, R., Franses, P., & Dijk, v.D. (1998). Timing of Vote Decisions in First and Second Order Dutch Elections 1978-1995: Evidence from Artificial Neural Networks. *Political Analysis, 7,* 117-142.

[3] Gorban, A. & Waxman C. Neural Networks for Political Forecast. Washington DC: 179p.

[4] Curtice, J. (1995). What Future for the Opinion Polls? The Lessons of the MRS Inquiry. In C. Rallings, D. Farrell, D. Denver & D. Broughton (eds.) *British Elections and Parties Yearbook 1995* (pp. 1396-156). London: Frank Cass.

[5] Curtice, J. (1997). So how well did they do? The polls in the 1997 election. *Journal of the Market Research Society, 39,* 449-461.

[6] Sanders, D. (1991). Government Popularity and the Next General Election. *Political Quarterly, 62,* 235-261.

[7] Clarke, H.D., & Stewart, M.C. (1995). Economic Evaluations and Election Outcomes: An Analysis of Alternative Forecasting Models. In D. Broughton, D.M. Farrell, D. Denver, & Rallings C. (Eds.), *British Elections and Parties Yearbook 1994.* (pp. 73-91). London: Frank Cass.

[8] Lewis-Beck, M.S., & Rice, T.W. (1992). *Forecasting Elections.* Washington D.C.: Congressional Quarterly Press.

[9] Campbell, J.E., & Garand, J.C. (2000). *Before the Vote: Forecasting American National Elections.* Thousand Oaks, CA, Sage.

[10] Rallings C., & Thrasher, M. (1999). Local Votes, national forecasts - using local

government by-elections in Britain to estimate party support. *International Journal of Forecasting, 15,* 153-162.

[11] Norpoth, H. (1995). Is Clinton Doomed? An Early Forecast for 1996. *PS: Political Science and Politics, 8,* 201-207.

[12] Norpoth, H. (2001). Primary Colors: A Mixed Blessing for Al Gore. *PS: Political Science and Politics, 34,* 45-48.

[13] Lisboa, P.G.J. (ed) (1992). *Neural Networks: Current Applications.* London: Chapman Hall.

[14] Picton, P. (1994). *Introduction to Neural Networks.* London: Macmillan.

[15] Swingler, K. (1996). *Applying Neural Networks: A Practical Guide.* London: Academic Press.

[16] Abdi, H., Valentin, D., & Edelman, B. (1999). *Neural Networks.* London: Sage.

Roman Borisyuk, Galina Borisyuk, Colin Rallings and Michael Thrasher. *Forecasting the 2001 General Election Result: A Neural Network Approach.*

[17] Rosenblatt, F. (2001). *Principles of Neurodynamics.* Washington D.C.: Spartan Books.