Journal of Forecasting
J. Forecast. **21**, 435–449 (2002)
Published online 29 May 2002 in Wiley InterScience (www.interscience.wiley.com). **DOI**: 10.1002/for.831

Accurate Forecasting of the Undecided Population in a Public Opinion Poll

CHRISTOPHER MONTEROLA,[1] MAY LIM,[1]
JERROLD GARCIA[2] AND CAESAR SALOMA[1]*
[1] *National Institute of Physics, University of the Philippines*
[2] *Department of Physics, Ateneo de Manila University, Philippines*

ABSTRACT

The problem of pollsters is addressed which is to forecast accurately the final answers of the undecided respondents to the primary question in a public opinion poll. The task is viewed as a pattern-recognition problem of correlating the answers of the respondents to the peripheral questions in the survey with their primary answers. The underlying pattern is determined with a supervised artificial neural network that is trained using the peripheral answers of the decided respondents whose primary answers are also known. With peripheral answers as inputs, the trained network outputs the most probable primary response of an undecided respondent. For a poll conducted to determine the approval rating of the (former) Philippine president, J. E. Estrada in December 1999 and March 2000, the trained network predicted with a 95% success rate the direct responses of a test population that consists of 24.57% of the decided population who were excluded in the network training set. For the undecided population (22.67% of December respondents; 23.67% of March respondents), the network predicted a final response distribution that is consistent with the approval/disapproval ratio of the decided population.

KEY WORDS public opinion poll/survey; undecided population; artificial neural network; discriminant analysis

INTRODUCTION

Public opinion polling is conducted over a constituency to test the acceptability of a particular personality, consumer product, or policy. In countries where a citizen's right to choose is guaranteed, polling is commissioned for various political, social, or economic ends. In a poll, a respondent is asked a primary question and three kinds of answers are possible, namely (1) Approved/Favourable (P), (2) Disapproved/Unfavourable (N) and (3) Abstained/Undecided (U). Typically, the same survey also asks indirect questions concerning the views of the respondent on certain peripheral socio-economic and political concerns.

* Correspondence to: Caesar Saloma, National Institute of Physics, University of the Philippines, Diliman, Quezon City, Philippines 1101. E-mail: csaloma@nip.upd.edu.ph
Contract/grant sponsor: Pulse Asia, Inc.

One of the most difficult tasks that a pollster faces with the survey results is to forecast accurately the final primary answers (P or N only) of the undecided respondents (URs)—a precious advantage in a close electoral contest or a highly competitive market with a large number of uncommitted voters or consumers. This paper addresses the task as a pattern-recognition problem of correlating the respondents' primary answer with their answers to the peripheral questions.

We show that the correlation that enables a pollster to predict accurately the final response distribution of the URs can be determined confidently with a supervised artificial neural network (NN). First, the NN is trained using as its inputs the peripheral responses of the decided respondents (DRs) whose primary answers are also known. After an acceptable performance with the DRs, the trained NN is then utilized to predict the final primary response of a UR using as its inputs the UR's known peripheral responses.

Over other standard statistical based (SSB) methods, the NN technique offers the following advantages for solving pattern-recognition problems:

(1) Solution search can proceed without *a priori* knowledge about the statistical behaviour (type of probability distribution function, cost function, etc.) of the data points, unlike in SSBs where predictions are susceptible to errors caused by erroneous assumptions concerning the statistical distribution.
2) Analysis of the trained NN solution and the inherent parallelism of the NN architecture permit an adaptive and fast investigation of the relative impacts of the various socio-economic factors on the primary response of the respondent. For example, by deliberately severing some of the NN connections to a peripheral response input, the robustness or dependence of a DR's primary response can be easily tested with respect to a particular peripheral issue—a versatility that is not generally possible with SSBs.
3) Performance of a trained NN is robust against noise (Soriano and Saloma 1995, 1998; Soriano *et al*., 2001) which in an opinion survey could arise from undersampling and imperfect probe design. Generally, SSBs do not perform well with noisy data sets which makes the determination of the correct *a priori* information more difficult.

Our choice of the NN technique is also based on its previous successes in finding non-linear relationships between variables in a wide variety of complex systems. It has been utilized successfully in noisy image recognition (Soriano and Saloma 1995, 1998; Fu *et al*, 1999) and feature identification in different kinds of signals (Zhang *et al*., 1998; Chen and Ware, 1999; Hoorn *et al*., 1999; Huse and Gjosaeter, 1999; Li and Bridgewater, 2000; Mittal and Zhang, 2000; Qureshi *et al*., 2000; Shimada *et al*., 2000), stock trading and securities (Anders *et al*., 1998; Gençay and Stengos, 1998; Moody *et al*., 1998; Shin and Kil, 1998; Garcia and Gençay, 2000. Hamm and Brorsen, 2000; Lam and Lam, 2000), financial management fraud (Fanning and Cogger 1998), inflation (Moshiri and Cameron, 2000), risk assessment in auditing (Ramamoorti *et al*., 1999), and financial earnings (Charitou and Charalambous, 1996). To our knowledge, the NN method has not yet been applied to forecasting the tendencies of the undecided population in a public opinion poll.

The presentation of our paper proceeds as follows. This next section describes the materials and methods utilized in our work. The experimental results are given and analysed in the third section. Comparison of the NN technique with a conventional statistical method called predictive discriminant analysis is provided in the fourth section. Discussion regarding how to use the NN technique for new sets of survey data is also presented in this section followed by a summary in the final section. The NN architecture and its supervised training are described in detail in Appendices A and B while Appendix C provides a survey methodology.

PRELIMINARIES

Pattern-recognition problem

Accurately determining the final primary response distribution of the URs is a difficult task when correlating the peripheral answers with the P or N response. The underlying patterns are mined from the data provided by the DRs but the search is not straightforward due to noise and the possibility that the peripheral questions are irrelevant, or overlapping in information content. Uncertainties are also inevitable when dealing with a large number of (independent-minded) respondents who are probed with a wide variety of questions. Determining the irrelevance of an indirect question or its dependence on the others is an equally difficult task for a complex and non-linear system such as the one at hand.

In our work, the performance of the trained NN is first assessed by applying it to a test set that consists of the remaining DRs (in our case, about 24% of the decided population) that were not part of the training set. With an acceptable performance on the test set, the trained NN is then used to predict the probable primary answer of a UR based on his or her peripheral responses. The NN outputs a '+1' for a 'P' response and '−1' for an 'N' response.

Survey data

Our data were taken from a poll that was conducted by a survey organization (Pulse Asia, Inc., 1999) to determine how the electorate rates the performance of then Philippine President Joseph Estrada. The poll questionnaire consists of one primary (direct) question and 291 (pro-rated) peripheral ones that probe into the respondent's opinions on specific socio-economic and political issues, and his or her approval ratings of other government executives, legislators, and institutions. A total of 1200 respondents participated from various parts of the country which allows a 95% confidence level with error margin of ±3% at the national level and ±6% at the sub-national level. Details of the survey methodology are given in Appendix C.

The December 1999 survey revealed that out of 1200 respondents, 596 (49.67%) were approving (P) of Estrada, 332 (27.67%), disapproved (N), and 272 (22.67%), were undecided (U). The total number of DRs was 928. The March 2000 survey results showed that out of 1200 respondents, 559 (46.58%) approved (P) of Estrada, 357 (29.75%), disapproved (N), and 284 (23.67%) were undecided (U). The total number of DRs was 916.

We also tested whether the URs belong, in some way, to the same statistical population as the DRs in the December 1999 survey. Their responses to the peripheral questions of each of the 1200 respondents were plotted in a 20-dimensional Euclidean feature space $\{\mathbf{X}_k\}$ where $k = 1, \ldots, 20$. Each axis corresponds to a peripheral question and every response is described by $\{\mathbf{x}_k\}$ where x_k can range from 1 (strong disapproval) to 5 (strong approval).

If there are patterns to the responses of the DRs then two clusters are formed corresponding to the P-respondents and the N-respondents, with their respective centroids. If the responses of the URs also form another separate class then three distinct clusters are found in $\{\mathbf{X}_k\}$. In this instance, the final outcomes of the URs cannot be predicted confidently with any SSB or NN technique.

In each $\mathbf{X}_k$ there corresponds a standard deviation σ_i brought about by the distribution of the x_k values in that axis for the 1200 respondents. Using the $\pm\sigma_k$'s as diameters centred at the mean $< x_k >$ of each x_k, we constructed two 20-dimensional ellipsoids: one centred at the centroid of the P-cluster (P-ellipsoid), and the other at the centroid of the N-cluster (N-ellipsoid). The size of each ellipsoid was progressively increased and the percentage of the U-points falling within either ellipsoid was determined as its size increased. We found that 97% of the U-points fell within the P-ellipsoid when its diameters reached the $\pm 3\sigma$ level, while 99% of the U-points could be found within the N-ellipsoid when its diameters reached the $\pm 3\sigma$ level. Hence, the set of U-points overlapped the sets of P-points and N-points, respectively, in the feature space and did not form a third separate cluster (Monterola *et al.*, 2002).

RESULTS

Network optimization

With reference to the December 1999 survey data, our experiments indicate that the best one-output, three-layer NN architecture is one with input node number I = 21 (M + bias = 21), hidden node number $H = 30$, and a hidden activation function $f_H(x) = 1.7159 \tanh(2x/3)$ where M is the number of peripheral questions that are used as NN inputs. Details of the NN architecture and its supervised training are given in Appendices A and B. The trained NN yields an 86.84% success rate for the test set as shown in Table I, and predicts a P : N value of 171 : 101 for the UR's. Hence, 62.87% of the URs will eventually support Estrada which is remarkably close with that (64.22%) of the DRs.

Table I. Performance of the NN as a function of the number of input questions M and number of hidden nodes H (data source = December 1999 poll) where $f_H(x) = 1.7159 \tanh(2x/3)$, $H = 30$, $M = 20$, η = adaptive (starting at 0.001), $D = 700$, $T = 228$, $\alpha = 0.1$, and $q = 10,000$

M	H	% success training set	% success test set	Predicted P : N URs
5	10	73.43	69.30	185 : 87
5	20	73.86	71.05	179 : 93
5	25	74.57	74.56	199 : 73
5	30	74.43	75.00	198 : 74
5	40	74.57	75.00	188 : 84
5	60	74.00	75.44	189 : 83
8	10	81.00	78.51	192 : 80
8	20	80.86	76.32	199 : 73
8	25	80.57	82.02	199 : 73
8	30	81.71	81.58	196 : 76
8	40	83.00	82.46	189 : 83
8	60	82.86	79.39	199 : 73
10	10	87.57	78.95	188 : 84
10	20	88.86	82.02	178 : 94
10	25	88.86	84.21	187 : 85
10	30	90.14	84.21	188 : 84
10	40	90.57	81.14	185 : 87
10	60	90.14	82.89	190 : 82
20	10	91.86	85.53	165 : 107
20	20	92.14	85.53	170 : 102
20	25	93.14	85.96	172 : 100
20	30	94.00	86.84	169 : 103
20	40	93.14	86.84	171 : 101
20	60	92.14	85.53	169 : 103

The performance of the trained NN ($M = 20$, $H = 30$) may be enhanced by removing possible redundancies in the 20 indirect questions that are used as NN inputs. We examine this possibility by training the NN where the inputs are introduced in a sequential manner:

- *Step 1*. Train the NN until the error function E^q stabilizes to a (small) value, with only one varying input ($r1$) and with the remaining nineteen ($r2$, $r3$, ..., $r20$) all held fixed to one,
- *Step 2*. Continue training with two varying inputs ($r1$, $r2$) until E^q stabilizes to a (small) value,
- *Step 3*. Repeat Step 2 with three varying inputs ($r1$, $r2$, $r3$), then four, and so on; until the maximum of 20 questions are all utilized as varying inputs.

The 20 questions were selected and ranked according to their degree of correlation with respect to the primary question. The degree of correlation is determined using a one-input/one-output feedforward NN with $H = 30$, $f_H(x) = f_o(x) = 1.7159\tanh(2x/3)$. In $\{r_i\}$ where $i = 1, 2, \ldots, I - 1$, $r1$ and $r2$ denote the questions with the highest and second highest correlation values, respectively.

Table II shows that the sequential mode of training results in a trained NN with a higher success rate of 93.86% (an increase of more than 7 percentage points) for the test set in the December 1999 survey. It also reveals that the best-performing NN is one with 20 varying inputs. NN performance is not enhanced by including in the training the respondents' impressions on the Philippine National Police ($r15$) possibly because they introduce only noise into the patterns and can be attributed to the redundancy of the police performance with $r2$ (curbing widespread criminality) and $r3$ (effective

Table II. Learning ability of NN that is trained via sequential presentation of the questions that are most correlated with the approval rating of Estrada (data source = December 1999 poll) where $f_H(x) = 1.7159\tanh(2x/3)$, $H = 30$, $M = 20$, η = adaptive (starting at 0.001), $D = 700$, $T = 228$, $\alpha = 0.1$, and $q = 10,000$

No. of inputs varied	% success training set	% success test set	Predicted P : N URs
1	66.71	63.16	232 : 40
2	65.71	65.35	246 : 26
3	68.00	67.54	221 : 51
4	69.86	72.37	210 : 62
5	76.00	76.75	200 : 7
6	77.86	81.14	199 : 73
7	79.71	82.46	199 : 73
8	82.57	82.89	187 : 85
9	88.29	83.77	186 : 86
10	92.00	85.53	189 : 83
11	92.14	85.53	179 : 93
12	93.14	86.40	179 : 93
13	93.14	86.84	168 : 104
14	94.57	88.59	165 : 107
15	94.42	87.28	164 : 108
16	96.00	88.16	168 : 104
17	96.57	90.79	168 : 104
18	96.86	92.98	169 : 103
19	97.29	92.11	171 : 101
20	97.71	93.86	172 : 100

Table III. Learning ability of NN for five different training sets ($D = 700$) that are randomly selected from the decided population (source data = December 1999 poll) where $f_H(x) = 1.7159 \tanh(2x/3)$, $H = 30$, $M = 20$, η = adaptive, $D = 700$, $T = 228$, $\alpha = 0.1$, and $q = 10,000$

Training set	% success training set	Success test set (%)	Predicted P : N URs
1	97.71	93.86	172 : 100
2	97.00	93.42	168 : 104
3	96.86	93.42	170 : 102
4	96.71	92.98	168 : 104
5	96.29	92.10	170 : 102

discipline of police). After determining the optimal NN architecture, we found a P-to-N ratio of 172 : 100 for the undecided population, i.e. 63.23% of the URs in the December 1999 survey will eventually support Estrada.

Five different training sets were utilized—each with 700 randomly selected URs. Table III shows an (average) test set performance of 93.16%—changes in the composition of the training set did not affect significantly the generalization performance of the 21-input NN. The small standard deviation (±0.67%) yielded by the five training sets attests to the randomness of the data collection process in the survey.

Similar experiments were performed in the March 2000 poll with the same sets of inputs. In this survey, the approval rating of Estrada decreased by about 3.5% while the URs increased by 1% (P : N ratio = 559 : 357, with UR = 284). The success rate of the trained NN was 97.71% for the training sets and 93.52% for the test sets. The trained NN predicted a P : N ratio of 175 : 109 for the URs which implies that 61.62 % will eventually support Estrada as summarized in Table IV.

Table IV. Comparison of three different strategies for predicting the final distribution of the undecided respondents where $f_H(x) = 1.7159 \tanh(2x/3)$, $H = 30$, $M = 20$, η = adaptive, $D = 700$, $\alpha = 0.1$, and $q = 10,000$. Net approval(%) = 100(P − N)/(P + N)

(A) December 1999 data ($T = 228$)

Prediction tool	Predicted P : N UR = 272	Accuracy (%)	Net Approval (%)
DR-based projection	175 : 97	NA	28.50%
Disc. analysis	136 : 136	69.87	24.17%
Neural networks	172 : 100	93.86	28.00%

(B) March 2000 data ($T = 216$)

Prediction tool	Predicted P : N UR = 284	Accuracy (%)	Net Approval (%)
DR-based projection	173 : 111	NA	22
Disc. analysis	146 : 138	70.39	14.83
Neural networks	175 : 109	93.52	22.17

The comparable accuracy (in both the training and test sets) obtained using the data of both polls (December 1999 and March 2000) indicates the repeatability of the procedure and the similarity of the dynamics of the sample populations. Interestingly, the predicted splitting of the URs by the NN technique is very close (average deviation of 0.24%) to the P : N ratio of the decided population.

Socio-demographic probe

We also study the influence of demography on the final primary answer of a UR by replacing the following peripheral questions in the original set: three questions on political personalities ($r16$, $r17$, $r18$), and one that is concerned with police performance ($r15$) with demographic questions on gender, age group, civil status and educational attainment. They are pro-rated as follows: (a) gender $r = 1$ (male), $r = 2$ (female); (b) age group $r = 1$ (18–19), $r = 2$ (20–24), $r = 3$ (25–29), $r = 4$ (30–34), $r = 5$ (35–39), $r = 6$ (40–44), $r = 7$ (45–49), $r = 8$ (50–54), $r = 9$ (55–59), $r = 10$ (60–70), $r = 11$ (71–75), $r = 12$ (75 and over); (c) civil status $r = 1$ (married), $r = 2$ (widowed), $r = 3$ (divorced), $r = 4$ (separated/married but separated/not living with legal spouse), $r = 5$ (single/never married), $r = 6$ (living-in as married); and (d) educational attainment $r = 1$ (no formal education), $r = 2$ (incomplete elementary education), $r = 3$ (elementary), $r = 4$ (incomplete high school education), $r = 5$ (high school), $r = 6$ (incomplete vocational), $r = 7$ (vocational course), $r = 8$ (incomplete college), $r = 9$ (college), $r = 10$ (post college).

With the new set of peripheral questions, the NN is trained sequentially using the March 2000 data in a manner similar to the one employed in Table II. Table V shows that the trained NN exhibits a slightly higher (about 1.5%) success rate of 95.37% for the test set. Because the difference is more than twice the standard deviation of different sets of randomized data shown in Table III, the improvement is probably not due to mere statistical fluctuations.

There are two possible reasons for the improvement of the success rate. One is that the questions related to $r15$ up to $r18$ contain only noise and redundant information that degrade the information presented by the other r_i's. The other possibility is that support for Estrada is also determined by socio-demographic conditions, and that the replacement questions contain additional information which were not present in the previous inquiries.

We experiment by randomly replacing any four of the original 20 questions with demographic information of the DR. A success rate of 95.18% with a very small standard deviation of $\pm 0.38\%$, is obtained for the test set after performing 50 sets of such randomization procedures which shows the robustness of the trained NN to the redundancy and noise that may be present in the original set of 20 peripheral questions. The results indicate that the dominant cause of the increase of the NN prediction accuracy is not the removal of $r15$ to $r18$ but rather the inclusion of the demographic questions.

The best-performing NN predicts a 95.37% accuracy, a P:N ratio of 176 : 108 for the undecided population which implies that 61.97% of the UR's will become favourable to Estrada. Again, the

Table V. Effect of socio-demographic inputs to NN performance (data source: March 2000) where $f_H(x) = 1.7159 \tanh(2x/3)$, $T = 216$, $H = 30$, $M = 20$, $\eta =$ adaptive, $D = 700$, $\alpha = 0.1$, and $q = 10,000$

Prediction tool	Predicted P : N UR = 284	Accuracy (%)	Net approval (%)
DR-based projection	173 : 111	NA	22
Neural Networks	176 : 108	95.37	22.5

ratio is almost equal to the previous prediction using the December 1999 results that the URs follow the proportion of the DR.

DISCUSSION

Comparative performance

Two assumptions may be immediately used to predict the final outcomes of the URs: (1) the P : N ratio displayed by the DR's also holds for the URs, or (2) the P : N ratio for the URs splits evenly. The first scenario implies that in the December 1999 survey, 175 out of the 272 URs will eventually support Estrada in accordance with the P : N ratio of 64.34 : 35.66 that is displayed by the decided population. For the latter scenario, the eventual P : N ratio of the undecided population would be 50 : 50. The first assumption is possible for URs who are incapable of deciding on their own, and had to rely on the decisions of the DRs. However, the assumption is tenuous for a population of independent-minded individuals in the absence of the bandwagon effect and/or massive coercion.

To highlight further the advantage of the NN technique, we compare its performance with an SSB method called predictive discriminant analysis (DA) (Hair *et al.*, 1998; Huberty, 1994), where M indirect responses are used as predictor variables to form the discriminant function Z. The best scenario is for the two populations to group separately and without overlaps. A UR's final answer (P or N) is then determined from his or her relative position with respect to the cut-off between the P and N populations.

For DA to be optimally effective, the M indirect questions must be all relevant and independent of each other. However, establishing that the M questions generate non-redundant information about the respondent is difficult in a large distribution of electorates from various cultural, socio-political, and personal backgrounds. In applying DA, we have to assume (1) multivariate normality of the predictor variables, (2) equal (but unknown) dispersion and covariance matrices for the P and N groups, (3) low correlation between two or more predictor variables, (4) all relationships are linear and (5) number of outliers is negligible.

We consider $M = 20$ peripheral questions that concerns the following issues: $r1$—economic recovery, $r2$—curbing widespread criminality, $r3$—effective discipline of police, $r4$—increasing the pay of workers, $r5$—fighting rapid price increases, $r6$—eradicating graft and corruption, $r7$—reducing the great poverty, $r8$—non-coddling of well-known Marcos cronies, $r9$—curbing widespread sale and use of illegal drugs, $r10$—stopping the widespread destruction of the environment, $r11$—protecting the human rights of citizens, $r12$—protecting the freedom of the press, $r13$—protecting territories against alien claims of aliens, $r14$—increasing peace in the country, $r15$—approval rating of Philippine National Police, $r16$—support for elected executive **Z** who is from opposition party, $r17$—support for elected legislative **Z** who is from an opposition party, $r18$—support for cabinet secretary **Y** who is from the same political party, $r19$—approval rating of the Bureau of Customs, and $r20$—approval rating of the Armed Forces of the Philippines. These peripheral questions are chosen arbitrarily and are also used as NN inputs.

Figure 1 presents the histogram $f(Z)$ plot (bin width $\Delta = 0.2$) for the discriminant scores Z for the P (solid circle), N (circle), and U (square) respondents. Although the plots hint a dependence between the peripheral and the primary answers, the grouping is far from ideal. The $f(Z)$ distributions for both P and N are not localized and they overlap in the Z-range considered. The overlaps which could arise from the partial dependency of the peripheral questions with each other represent uncertainties in the behaviour of the P and N populations.

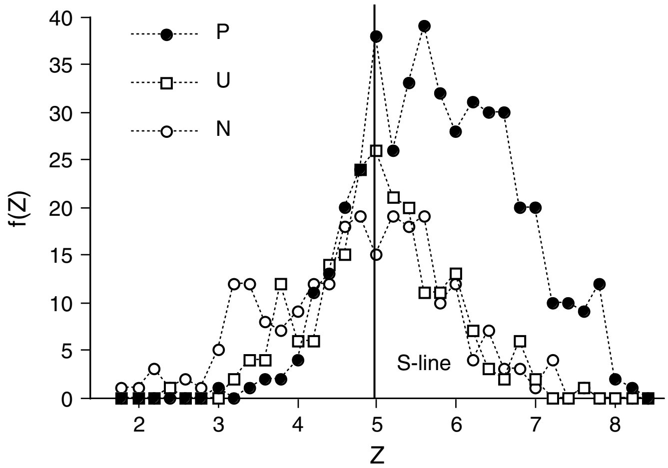

Figure 1. Distribution of the primary responses of the DRs in the December 1999 survey by discriminant analysis. Histogram $f(Z)$ of the discriminant scores Z for the P (solid circle), N (circle) and U (square) respondents that shows the separation line S = 4.95, that partitions the P and N distributions of the URs in the survey

The URs could be grouped into either the P or N state by segregating them (with the S-line) along a Z-score that minimizes the difference between the actual net approval rating and the calculated net approval (NA) rating for the data set used in the calculation, where NA = 100(P − N)/(P + N). It happens with the S-line situated in the range $4.94 < Z < 4.96$. With a prediction accuracy of 69.87%, a P : N ratio of 136 : 136 is obtained for the URs, i.e. 136 (50%) out of the 272 URs will support Estrada. Prediction accuracy is given by the ratio of the correctly determined respondents and the total number of respondents in the cross-validation set.

The prediction accuracy of DA is 24% lower than the NN method. Mostly we attribute the error of DA to the initial assumptions and bias used to established the P and N clusterings of the URs. The linearity and low correlation of predictor variables assumption, for instance, can never take into account the possibility of interdependence of the input parameters.

Efficacy of the trained NN

To ensure optimal performance, the NN is retrained with each new set of survey data. However, the NN architecture and method of training both remains the same with each retraining. We found that an NN trained only with the December 1999 survey data would perform poorly (the success rate is only 73% on average) in predicting the behaviour of the test set that belongs to the March 2000 survey. With the NN parameters already established, retraining proceeds rapidly because no *a priori* knowledge is required for the statistics of the respondents.

SUMMARY

We have established that the task of predicting the final primary answer of a UR in a public opinion poll is a pattern-recognition problem that is best solved via the NN technique. Our experiments with the test set showed that a trained NN could map accurately the correlation pattern between the peripheral responses and the primary response. The trained NN yields a 95% success rate for the DRs in the test set—a high degree of success that provides a sound basis for an accurate forecast of the final distribution of URs with regard to the primary question.

The trained NN predicted a P : N value of 171 : 101 (62.87%) for the URs in the December 1999 survey which is also remarkably similar to that (64.22%) of the DRs. It also predicted the answers of the URs with a greater level of confidence (as high as 95%) than the conventional DA technique which is only 70%. The generalizing capability of the trained NN as measured by its success in determining the correct P : N distribution of the test set is enhanced (from 87% to 94%) when the NN input information is sequentially presented according to the degree of its correlation with the popularity of Estrada.

The high success rate and degree of accuracy are achieved even with only 20 probe questions ($M = 20$) used as NN inputs which is much less than the total number (291) utilized in the questionnaire. Although the performance of the NN generally increases with larger M, the utilization of all the remaining 271 questions as additional NN inputs can improve the generalization success rate by 5% at best. Also, larger M leads to longer training times—a large NN normally takes more time to learn than a smaller one.

Aside from the use of orthogonal inputs, better generalization performance could be achieved if the patterns that are generated by the URs are less contaminated by noise. Poll data with a better signal-to-noise ratio may be obtained by employing a smaller number of probes in the questionnaire to avoid fatigue and loss of interest on the part of the respondent and the pollster.

APPENDIX A. NETWORK TRAINING

We utilized a single-output, three-layer feedforward NN consists of I inputs and H hidden units $\{y_{h;}h = 1, 2, \ldots, H\}$. It is trained in a supervised manner using the gradient descent backpropagation (GDB) method (Haykin, 1999; Khanna, 1990). The inputs $\{r_i\}$ consist of the M indirect (pro-rated) responses (r_1, r_2, …, r_M) and a bias $r_{M+1} = 1.0$. Initially $\{r_i\}$ are preprocessed by undergoing sequentially three normalization steps to hasten the reduction of error E^q, namely: (1) removal of the mean, (2) decorrelation, and (3) covariance equalization (LeCun, 1993). The entire process involves normalization of $\{r_i\text{'s}\}$ within the interval $(-1, 1)$ while simultaneously forcing its mean to be zero.

The training set consists of a D-number of respondents who are randomly selected from the DRs in the survey. Each respondent has a known direct response (either P or N), and a set of M indirect responses. Different sets of M indirect questions were also considered. The test set T consists of the remaining decided respondents ($T = DR - D$). The goal of training is to develop an NN with the highest success rate for the respondents in the test set.

Our goal is to find the weights $\{w_{ih}\}$ and $\{d_h\}$ that minimize the error function E^q, where w_{ih} represents the connection between the ith input and then hth hidden unit, where d_h represents the connection between the hth hidden unit and the NN output. Initially, the weights $\{w_{ih}, d_h\}$ are randomly chosen within the interval $\{-1/\chi^{1/2}, 1/\chi^{1/2}\}$ to avoid the probability of driving the elements $\{w_{ih}\}$ and $\{d_h\}$ to saturation values at an early stage of training, where χ is the number of synaptic connections in a particular neuron (Le Cun, 1993; Yam and Chow, 2000). The weight element w_{ih} is updated according to:

$$w_{ih}^{q+1} = w_{ih}^{q} - \eta(\partial E^q/\partial w_{ih}^{q}) - \alpha\Delta w_{ih}^{q-1} \tag{A1}$$

where η is the learning rate, α is the momentum term, and Δw_{ih}^{q-1} is the change in w_{ih} after $q-1$ iterations. Similar relations hold for d_h^{q+1} and d_h^{q+1}.

The η value is adaptively varied per iteration to account for the variation of the error surface along different regions of weight dimension (Lou, 1991). Its value is increased if the gradient of E^q maintains its algebraic sign after several (five) iterations and decreased if the gradient of E^q fluctuates. These rules are incorporated to enhance the minimization of E^q and prevent unnecessary oscillations in the search space (Haykin, 1999; Lou, 1991).

E^q is defined as $E^q = (1/2)\Sigma_i\ [\Psi(r_i)$ - $\Psi^q(r_i)]^2$, where $\Psi(r_i) = \Psi$, is the true solution and $\Psi^q(r_i) = \Psi^q$, is the NN solution after the qth iteration. The true solution has two possible values: $\Psi = +1$ for a P response and $\Psi = -1$ for an N response. On the other hand, the NN output state is determined on the winner-take-all principle. The NN output is rounded-off to -1 when $\Psi^q \quad \leq 0$ and $+1$ when $\Psi^q > 0$. The NN output is given by $\Psi^q = f_o(\Sigma_h d_{hR}^q y_h^q)$, where $y_h^q = f_H(\Sigma_k w_{ih}^q r_i)$ and $f_H(x)$ and $f_o(x)$ are hidden and output activation functions, respectively.

For pattern-recognition tasks, the best choice for $f_o(x)$ is the non-linear antisymmetric sigmoidal function (1, 2, 34): $f_o(x) = 1.7159$ tanh(2x/3). For the hidden nodes, we compared the performance of the bounded functions of $f_H(x) = \sin(x)$ and $f_H(x) = f_o(x)$. Our experiments showed that $f_H(x) = \sin(x)$ converged faster but at the expense of poorer generalization when compared with $f_H(x) = f_o(x)$. To establish the best NN architecture, the December 1999 poll results were utilized to determine the optimal NN parameter values.

Figure 1 plots the NN performance as a function of q for $f_H(x) = f_o(x)$ (circles) and $f_H(x) = \sin(x)$ (solid line), where $H = 30$ and $M = 20$. The indirect questions are the same as those employed in the discriminant analysis experiments (see earlier in the paper). The 20 questions for the NN inputs were chosen randomly from the 271 indirect questions in the questionnaire. Figure A1(a) shows than at $q = 1000$, E^q is reduced to 0.70 [$f_H(x) = \sin(x)$], and 0.66 [$f_H(x) = f_o(x)$], respectively. Figure A1(b) reveals that a lower E^q decrease does not necessarily lead to better prediction of the respondents in the training set. Note that the number of responses correctly learned by the NN for both types of $f_H(x)$ saturates at 656 out of the 700 patterns in the training set (93.70% classification rate). Figure A1(c) indicates that $f_H(x) = f_o(x)$ yields a better prediction for the test sets. Although the use of $f_H(x) = \sin(x)$ results in a lower E^q value during training, the trained NN has a poorer generalization performance, which is a much more important consideration.

We also studied the performance of the NN with $f_H(x) = a\tanh(bx)$ as a function of H, and M with $D = 700$. For the December 1999 poll $DR = 928$ and thus $T = 228$. Table II shows that the test set performance (after 10,000 iterations) of the NN does not depend significantly on H. The highest success rate for the test set is 87.72% and it is obtained with a 21-input NN ($M = 20$) with 30 hidden units. Generally, the performance of the trained NN on the test set improves non-linearly with the size of M.

APPENDIX B: ADVANTAGES OF GDB TRAINING METHOD

Our choice of the GDB method are based on the following observations:

(a) In higher-order optimization methods like the limited memory quasi-Newton algorithm (LMNA) or the conjugate gradient (CG), two quantities are important—the Hessian matrix (HM) and the Jacobian matrix or gradient vector (JM), respectively. The HM elements are second-order partial derivatives of the error surface with respect to the weights, while JM contains first-order

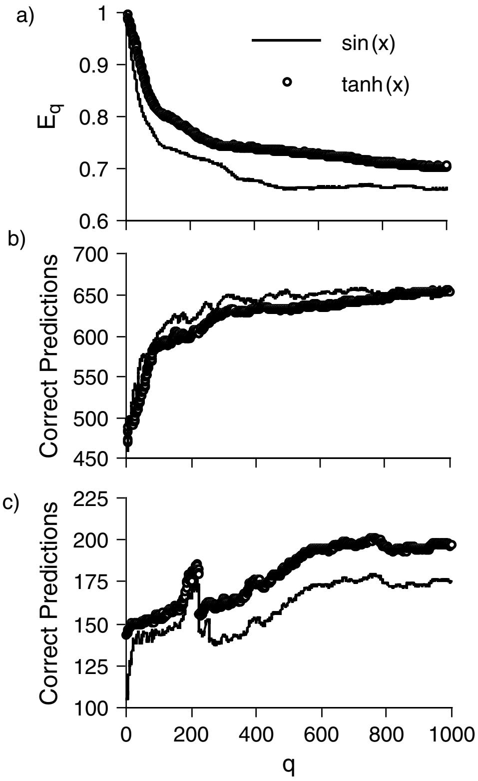

Figure A1. NN performance versus iteration number q which is proportional to training time. (a) Average error E^q versus q; (b) number of correctly predicted respondents in a training set of 700 elements as function of q; (c) number of correctly predicted respondents in a test set of 228 elements as a function of q

partial derivatives. Empirical studies (Saarinen *et al.*, 1991a,b) have shown that both matrices are almost rank-deficient when used in multilayer perceptron NN, which is a consequence of the intrinsically ill-conditioned nature of the NN training process. Their conclusion has established the advantage of GDB over higher-order methods which are computationally complex and may not converge more rapidly.

(b) Both LMNA and CG involve the calculation of HM and JM and their respective inverses. Though in terms of iterations both algorithms could converge faster than GDB, this advantage is outweighed by the fact that HM and JM are not guaranteed to be non-singular and their inverses are not be computable. Both matrices and their inverses are also expensive to compute (Saarinen *et al.*, 1991a,b, Battiti, 1992).

(c) Faster training is not the primary concern but the accurate determination by the NN of the mapping between the peripheral and primary questions (generalization performance). The GDB has a proven record of success to many real-world problem, and is the most commonly used technique in NN learning (Haykin, 1999, Quito *et al.*, 2001; Monterola and Saloma 2001; Saad and Rattray, 1997, 1998).

The following issues have remained essentially unresolved in NN applications (Huberty, 1994; Saad and Rattray, 1998): (1) optimum number of hidden layers and H for a given task, and (2) optimum learning rate that prevents saturation. The successful design and training of an NN still remains to be a trial- and-error method that depends highly on the skills of the designer.

APPENDIX C: SURVEY METHODOLOGY

A multistage probability sample of 1200 respondents 18 years old (the voting age for Filipino citizens) and above is employed defining the Philippines into four major geographic areas namely (Pulse Asia, Inc 1999): (1) National Capital Region also known as Metropolitan Manila, (2) Rest of Luzon, (3) Visayan islands in Central Philippines, and (4) Mindanao. Each area is assigned 300 randomly drawn respondents and two independent surveys were conducted on 9–20 December 1999 and on 15–29 March 2000. Sex and urban-rural controls were provided for in accordance with demographic projections for 1999 and 2000 based on the 1995 Census of Population by the National Statistics Office—Philippines. The questionnaire was originally formulated in the Filipino language and translated into five other languages and backtranslated into Filipino for accuracy and reliability.

The survey was not commissioned by any single individual or corporate entity. Its costs were defrayed primarily through contributions of survey subscribers from both the private sector and the Philippine government, neither one of which had any proprietary right to the survey data generated. The questionnaire was designed without political or partisan bias and the ensuing fieldwork were all completely under the management and control of Pulse Asia, Inc.

Each questionnaire contained one direct question and 291 indirect ones. Inquiries regarding the socio-demographic conditions (age, educational attainment, civil status, etc.) of respondents form a separate part of the questionnaire. The answers (r) to the indirect questions related to approval ratings are pro-rated numerically in the following manner: $r = 5$ (strongly supportive), $r = 4$ (mildly supportive), $r = 3$ (no opinion), $r = 2$ (mildly unsupportive), $r = 1$ (strongly unsupportive). For questions dealing with socio-demographic data, the corresponding numerical ratings are based on the actual number of possible choices of the respondents.

ACKNOWLEDGMENTS

The authors are grateful for stimulating discussions with F. Miranda of the Department of Political Science, University of the Philippines. The project is partially supported by Pulse Asia, Inc.

REFERENCES

Anders U, Korn O, Schmitt C. 1998. Improving the pricing of options: a neural network approach. *J Forecasting* **17**: 369–388.

Battiti R. 1992. First- and second-order methods for learning: Between steepest descent and Newton's method. *Neural Computation* **4**: 141–166.

Charitou A, Charalambous C. 1996. The prediction of earnings using financial statement information: empirical evidence with logit models and artificial neural networks. *Int. J. of Intelligent Systems in Accounting, Finance & Management* **5**: 199–215.

Chen D, Ware D. 1999. A neural network model for forecasting fish stock recruitment. *Canadian J. Fisheries & Aquatic Science* **56**: 2385–2391.

Fanning K, Cogger K. 1998. Neural network detection of management fraud using published financial data. *Int. J. Intelligent Systems in Accounting, Finance & Management* **7**: 21–41.
Fu H, Xu Y, Chang H. 1999. Recognition of handwritten similar Chinese characters by self-growing probabilistic decision-based neural network. *Int. J. Neural Systems* **9**: 545–562.
Garcia R, Gencay R. 2000. Pricing and hedging derivative securities with neural networks and a homogeneity hint. *J. Econometrics* **94**: 93–101.
Gençay R, Stengos T. 1998. Moving average rules, volume and the predictability of security returns with feedforward networks. *J. Forecasting* **17**: 401–414.
Hair J, Anderson R, Tatham R, Black W. 1998. *Multivariate Data Analysis* (5th edn). Prentice-Hall: New York.
Hamm L, Brorsen B. 2000. Trading futures markets based on signals from a neural network. *Appl. Economics Lett.* **7**: 137–140.
Haykin S. 1999. *Neural Networks: A Comprehensive Foundation* (2nd edn). Prentice-Hall: New York.
Hoorn J, Frank S, van der Ham F. 1999. Neural network identification of poets using letter sequences. *Literary & Linguistic Computing* **14**: 311–318.
Huberty C. 1994. *Applied Discriminant Analysis*. John Wiley: New York.
Huse G, Gjosaeter H. 1999. A neural network approach for predicting stock abundance of the Barents Sea capelin. *Sarsia* **84**: 457–460.
Khanna T. 1990. *Foundations of Neural Networks*. Addison-Wesley: Reading, MA.
Lam K, Lam K. 2000. Forecasting for the generation of trading signals in financial markets. *J. Forecasting* **19**: 39–52.
LeCun Y. 1989. Generalization and network design strategies. Technical Report CRG-TR-89-4. Department of Computer Science, Toronto Canada.
LeCun Y. 1993. Efficient learning and second-order methods. A tutorial at NIPS 93, Denver.
Li Y, Bridgewater J. 2000. Prediction of extrusion pressure using an artificial neural network. *Powder Technology* **108**: 65–70.
Lou Z. 1991. On the Convergence of the LMS algorithm with adaptive learning rate for linear feedforward networks. *Neural Computation* **3**: 226–245.
Mittal G, Zhang J. 2000. Prediction of temperature and moisture content of frankfurters during thermal processing using neural network. *Meat Science* **55**: 13–19.
Monterola C, Lim M, Garcia J, Saloma C. 2002. Feasibility of a neural network as classifier of undecided respondents in a public opinion survey. *Int. J. Public Opinion Research* (forthcoming).
Monterola C, Saloma C. 2001. Solving the nonlinear Schrodinger equation with an unsupervised neural network. *Opt. Express* **9**: 72–84.
Moody J, Wu L, Liao Y, Saffell M. 1998. Performance functions and reinforcement learning for trading systems and portfolios. *J. Forecasting* **17**: 441–470.
Moshiri S, Cameron N. 2000. Neural network versus econometric models in forecasting inflation. *J. Forecasting* **19**: 201–217.
Pulse Asia, Inc. 1999. Ulat ng Bayan Survey (9–20 December 1999). Philippine Social Science Center Building, Commonwealth Ave, Diliman, Quezon City, Philippines.
Quito M, Monterola C, Saloma C. 2001. Solving *N*-body problems with neural networks. *Phys. Rev. Lett.* **86**: 4741–4744.
Qureshi K, Naguib R, Mellon J. 2000. Neural network analysis of clinicopathological and molecular markers in bladder cancer. *J. Urology* **163**: 630–633.
Ramamoorti S, Bailey A, Traver R. 1999. Risk assessment in internal auditing: a neural network approach. *Int. J. Intelligent Systs in Accounting, Finance & Management* **8**: 159–180.
Saad D, Rattray M. 1997. Globally optimal parameters for on-line learning in multilayer neural networks. *Phys. Rev. Lett.* **79**: 2578–2581.
Saad D, Rattray M. 1998. Analysis of on-line training with optimal learning rates. *Phys. Rev. A* **58**: 6379–6391.
Saarinen S, Bramley RB, Cybenko G. 1991a. CRSD Report no. 1089. Center for Supercomputing Research and Development, University of Illinois, Urbana.
Saarinen S, Bramley RB, Cybenko G. 1991b. In *Automatic Differentiation of Algorithms: Theory, Implementation, and Application*, 31–42. Philadelphia, PA: SIAM.
Shimada T, Shiina T, Saito Y. 2000. Detection of characteristic waves of sleep EEG by neural network analysis. *IEEE Trans. Biomed. Engineering* **47**: 369–375.

Shin F, Kil D. 1998. Classification Cramer–Rao bounds on stock price prediction. *J. Forecasting* **17**: 389–399.
Soriano M, Saloma C. 1995. Cell classification by a learning principal component analyzer and a backpropagation neural network. *Bioimaging* **3**: 168–176.
Soriano M, Saloma C. 1998. Improved classification robustness for noisy cell images represented as principal-component projections in a hybrid recognition system. *Appl. Opt.* **37**: 3628–3639.
Soriano M, Garcia ML, Saloma C. 2001. Fluorescent image classification by major color histograms and a neural network. *Opt. Express* **8**: 271–277.
Yam JYF, Chow TWS. 2000. A weight initialization method for improving training speed in feedforward neural networks. *Neurocomputing* **30**: 219–232.
Zhang G, Patuwo B, Hu M. 1998. Forecasting with artificial neural networks: The state of the art. *Int. J. Forecasting* **14**: 35–62.

Authors' addresses:

Christopher Monterola, **May Lim** and **Caesar Saloma**, National Institute of Physics, University of the Philippines, Dillman, Quezon City, Philippines 1101.

Jerrold Garcia, Department of Physics, Ateneo de Manila University, Loyola Heights, Quezon City, Philippines.

