Volume 4, Issue 3
June 2015

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

Introduction of a New Method to Improve the Performance of Bayesian Neural Network

[1] Mahdi Sabri, [2] Roshanak Rezaeipour*

[1] Department of Electrical Engineering, East Azarbaijan Science and Research Branch, Islamic Azad University, Tabriz, Iran.
[2] Department of Electrical Engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran.
E-mail: [1]stu.mahdi.sabri@iaut.ac.ir, [2]rezaeipour@iaut.ac.ir

Abstract

In this paper, bootstrap method re-sampling is a family of methods that can be observed without assuming normality in the use of statistical quality control. In statistics, computer bootstrapping a method for assigning a measure on the accuracy of the estimates. This technique can be estimated almost any data with a very simple method of distributing data. It is generally considered sampling procedure again. Bootstrapping estimate of the characteristics of an estimate (i.e., variance) is measured using the same characteristics in an approximate distribution to the data. In the case with a series of observations on a population can be assumed to be independent and evenly distributed, the bootstrapping can be implemented by making some sample reviewed. Review each of these instances, in fact, set random samples with replacement from the original data. Network training is done separately for each of the samples.
Keywords: estimation, performance, Bayesian Neural Network (BNN), bootstrap

1. INTRODUCTION

In the late eighteenth, century has been bootstrapped words written by Radolf Eriek Raspe. When he finds himself in the bottom of the lake without the aid and any tools, using a strap to take up his boots. When we have the least information, namely distribution among the population is not known, nor is the data sufficiently. We try to offer a small sample that we have good and valid inference. A re-sampling method, the bootstrap. Bootstrap sampling was explained by Efron [1] and was working on it for a long time. Liang and et al [2] used to bootstrap the partial linear regression model. In this method, the analysis is done on real data. Small samples and authentic part of an investigation and in practice, based on simulation. However, in some instances do not need to simulate. The simulation is a community distribution but if the distribution is unknown, a taken arbitrary number of core samples and consider them out. So there is no need to distribute community.

Several articles have suggested bootstrap method for the design of statistical process control when the distribution process is unknown. Bajgier [3] used bootstrap percentile confidence interval to determine the structure of control. Franklin and Wasserman [4] reviewed bootstrap confidence for process performance. Seppala and et al [5] examined the statistical process control through the bootstrap groups. Liu and Tang [6] used control charts for dependent and independent observations based on bootstrap. Wu and Wang [7], Wood and et al [8] examined the bootstrap technique dependent model for the analysis of control. Teyarachakul and et al offered a more advanced bootstrap method based on analysis of statistical process control charts remaining. Park [10] based on Bootstrap Method of control provided to control the middle of the chart. In all bootstrap confidence intervals used in the above references, only the percentile bootstrap confidence intervals were used to determine the control limits.

* Corresponding author. Tel.: +98 9144113847;
E-mail address: r.a.rezaei@gmail.com (R. Rezaeipour)

This paper is structured as follows. In Section II, we are discussed Bayesian Neural Network (BNN) analysis using the equations of this type of education for review relationship between the input and aims .Section III describes bootstrapping method. In addition, the bootstrapping is introduced to select the most effective simultaneous inputs and parameters of the BNN. Therefore, bootstrapping a method for assigning a measure on the accuracy of the estimates. Furthermore, in Section IV, introduced bootstrapped Bayesian Neural Network (BNN) . This new method introduced to improve performance in BNN. Furthermore, bootstrapping has been shown to provide very good estimates of error for statistical models. Section V concludes the paper with a summary of results.

2. BAYESIAN NEURAL NETWORK ANALYSIS

During routine training artificial neural networks, create a relationship between the input and aims. If the input data "x" and "y" is taught objective data, the relationship between "x" and "y" can be explained as follows.

$$y = \int (x|w) + E \tag{1}$$

$\int (x|w)$ is a function that estimates of the relationship between input data and target. "w" is the weight vector and bias layer neural network neurons also "E" is the error. The purpose of education is to find an optimum weight vector for a common network while network was trained to input data to produce objective data with minimal error. Unlike conventional neural networks, Bayesian neural network target is not to find the optimal weight vector. It aims to provide clear uncertainty with regard to the amount of weight by a probability secondary dis-

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

tribution. The weight of this layer has a probability distribution that the distribution is obtained from the output data. According to the law Bayesian [11] have:

$$P(w|y, X) = \frac{P(y|w, X)P(w)}{P(y|X)} \quad (2)$$

"X" is input vector and consists of (x_1, x_2, ... x_n), "Y" is the target vector and consists of (y_1, y_2, ... y_n), also P (w) is primary of the weight distribution, P (Y | w, X) is the conditional maximum likelihood [12]. Border distribution, is called the normalized fixed proportionality constant truth. The initial distributions of weight are actually aware of the weight of the inputs for the network. Here, it is assumed that there is no information in the record (optional values). As in [13] followed the predicted distribution of "Yn + 1" is obtained from the following equation.

$$P(y_{n+1}|x_{n+1}, y, X) = \int P(y_{n+1}|x_{n+1}, w)\, P(w|y, X) dw \quad (3)$$

Index "n + 1" represents a new range of input and output variables that are not part of the "X", "Y". The integration cannot solve by analytical and numerical methods. To solve this integral used Markov Chain Monte Carlo (MCMC) numerical method to achieve in 1992. MCMC method is always the target to aim at multiple samplings of the masses [14]. Mainly primary distribution is a complex multi-layer neural network weight. Since the sample directly is a problem from the primary distribution complex. It is used to make it easier to distribute the weight vector to dispatch a proposal called the Gaussian distribution. The proposed distribution is dependent only on the initial weight by Markov random survey on the implementation of the loop and Start with optional quantities of primary distribution, a series of values of "W *" given by Markov ring. Proposed amounts likely to be accepted from the following equation.

$$\alpha = min\left\{\begin{matrix} 1 \\ \frac{P(y|X, w*)P(w*)}{P(y|X, w*)P(w_{prev})} \end{matrix}\right\} \quad (4)$$

Where "W_{prev}" is the initial weight. If shall be "W *", quantity of "W *" replaced by the "W_{prev}" and the process is repeated from the beginning. The optimal amount of between 30 and 70 percent intended to evaluate the acceptance of the screw [15]. The use of a large number of iterations to the convergence of Markov ring ensures a constant distribution. So we try to take the weight of its primary distribution.

3. BOOTSTRAPPING

In statistics, computer bootstrapping a method for assigning a measure on the accuracy of the estimates. This technique can be estimated almost any data with a very simple method of distributing data. It is generally considered sampling procedure again. In other words, the bootstrapping estimate feature (i.e., variance) is measured using the same characteristics in an approximate distribution of the data. In the case with a set of observations are assumed to be independent of the population is evenly distributed. Bootstrapping can be re-implemented with the construction of a number of samples, each of these subjects again, in fact, random samples with replacement from the original data set. Bootstrapping can also be used for the production of statistical hypothesis testing. This method is commonly used as an alternative to analytical methods based on parametric assumptions about the assumptions when they have doubts. In cases, where the use of the bootstrapping, parametric inference is impossible or complicated mathematical formula, used to calculate the standard error. A major advantage of bootstrapping is its simplicity. The method used to estimate the standard errors and confidence intervals to estimate the distribution of complex parameters such as percentile points, compared to the fraction of excellence and correlation coefficients, straightforward. In addition, a good way to check the stability of results. Because the bootstrapping asymptotically stable under some conditions, general finite sample does not provide a guarantee. In addition, there tends to act very optimistic. This method is most simplest default bootstrap analysis (i.e., assuming the independence of the sample) while in other hidden assumptions in drawing expressed [16]. Bootstrap methods are recommended for use in the following cases:

- When a statistic distribution of the unknown or complicated.
- When the sample size is insufficient for a straightforward statistical inference.
- When the calculations can be done, but we have a small pilot sample.

4. BOOTSTRAPPED BAYESIAN NEURAL NETWORK

A model to estimate the probability distribution of the forecast should provide appropriate and accurate, which is obvious. Distributed forecasts predict when exactly the amount expected to be close to the true value. The closing of a distribution predicted by the measured range. Bayesian Neural Network (BNN) is an acceptable tool for the development of transition functions, but is not able to estimate the uncertainty of output parameters. To ensure data created by the transition functions, uncertainty quantification should be expected. So, we need to calculate that the various sources of uncertainty that affect weight "$\hat{w}$" [17]. To achieve this purpose, to concerns about the estimated output of "f (x_n; w)" of the μ_y(x), should it be considered that the data series "T≡ {(x_n, y_n), n = 1, 2, ..., n_p}" that are used in network training is one of the infinite numbers of possible data from the data on the input "v_x" and according to the distribution of derived statistical error. In fact, diversity training data set for both the input sampling frequency "x_n;" and correspondence outputs with the inputs. Training with the "T" of the data series is different from the weight of the network. So we will have a distribution of the variance of the error function is calculated as follows (according to the data set "T"):

$$\sigma_f^2(x) = E\{[f(x_n; \hat{w}) - E[f(x_n; \hat{w})]]2\} \quad (5)$$

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

Where it is practically the Neural Network algorithm may be more or less complete and systematic survey results. The expected rate "E[f (x_n; $\widehat{w}$)]" is not equal to the actual final performance "μ_y (x)". This difference is called bias. The bias is zero in the entire neural network. Given all the possible variance of the error training "f (x; $\widehat{w}$) -μ_y (x)" is equal to:

$$E\left\{\left(f(x;\widehat{w})-\mu_y(x)\right)^2\right\}=$$

$$E\left\{\left[f(x;\widehat{w})-E[f(x;\widehat{w})]\right]^2\right\}+\left\{E\left[f(x;w^{\wedge})\right]-\mu_y(x)\right\}^2 \quad (6)$$

The first dimension is equal to the variance, distribution function "f (x; $\widehat{w}$)", the second term is the square of the bias. If "$E\left\{\left[f(x;\widehat{w})-E\left[f(x;w^{\wedge})\right]\right]^2\right\}$" equal to zero there will be no additional production. Another source of error into the estimates "μ_y (x)" arises from improper network structure. In addition to the number of nodes on the network is very low since the distribution function "f (x; $\widehat{w}$)" flexible enough to model data, the bias is high. The flexibilities of the model by increasing the number of nodes (the number of input parameters) are increasingly increases. The increase in the variance of the "equation (6)," increases bias and therefore, more training in the network. Another source of uncertainty comes from the fact that the network carried out where the error at least minimized the algorithm itself. In this case, training may be prematurely stopped before reaching the minimum. Quantify the precision of the estimates "f (x; $\widehat{w}$)" of a certain function properly, "μ_y (x)" (from the perspective of Confidence interval) requires the assumption of a distribution error "$\left(f(x;w^{\wedge})-\mu_y(x)\right)^2$" in "equation (6)" and it is estimated variance. The usual practice is to assume that the bias of the second term in "equation (6)" according to the first clause is ignored. The neural network training data set is skewed estimates of things (for example, they always want to peak too sharp). However, in many applications the variance is so dominant, in fact, the so-called bias [18]. Furthermore, if it was possible to calculate the bias components could be used for early reform in order to achieve a more accurate function. That is why we focused around the issue of variance estimation "$\sigma_f^2(x)$" using the bootstrapping technique. Bootstrapping estimate of the characteristics of an estimate (i.e., variance) is measured using the same characteristics in an approximate distribution of the data. In the case with a series of observations on a population can be assumed to be independent and evenly distributed, the bootstrapping can be implemented by making some sample reviewed. Review each of these instances, in fact, set random samples with replacement from the original data. Network training is done separately for each of the samples. Suppose the data from a series of random data "$T\equiv\{(x_1,y_1),(x_2,y_2),\ldots,(x_n,y_n)\}$", "n" is the size of the population with an unknown probability distribution "F". Where "$t_i=\{(x_i,y_i)\}$" is a random sample distribution of the "F", includes the same independent predictor vector "x_i" and "y_i" output variable is corresponding. "F" is empirical distribution function of "T_n" and "T^n" with the random sample of size "n". That use the same and independent distribution by the placement of "F". While "t_i" is a single random observation of "$t_i=\{(x_i,y_i)\}$". "B" series provided of samples bootstrapped as "T_1, T_2, ..., T_B" that the "B" sample is bootstrapped by its range of 50 to 200 [19]. "T_B" runs for each separate network once. It is shown that the output "$f_{BNN}\left(x_i;{}^{w_i}/_{T^B}\right)$". Network is used for in evaluation of "y" that are not within the "B" sample. To calculate the average error is used to generalize the results of from implementation of the model for bootstrapped samples. The generalization error "E_0"is obtained for the neural network of the following equation [17]:

$$E_\circ=\frac{\sum_{b=1}^{B}\sum_{i\epsilon A_b}\left(y_i-f_{BNN}\left(x_i;{}^{w_i}/_{T^B}\right)\right)^2}{\sum_{b=1}^{B}\neq(A_b)} \quad (7)$$

The output from the bootstrap samples is shown marked with "$f_{BNN}\left(x_i;{}^{w_i}/_{T^B}\right)$". "$A_b$" consists of a series of indicators for pairs has been observed that are not "T^B" bootstrap examples. "x_i" and "w_i" respectively represent the input vector and weight, "$\neq(A_b)$" number of pairs found in "A_b" for a new input "x". BNN bootstrapped estimates "$\hat{\theta}_{(X)}$" (which is the average of estimates from bootstrap "B") is obtained from the following equation.

$$\hat{\theta}_{(X)}=\frac{1}{B}\sum_{b=1}^{B}f_{BNN}\left(x_i;{}^{w_i}/_{T^B}\right) \quad (8)$$

And the "equation (9)," is used to estimate "$\hat{\delta}_{boot}^2(x)$".

$$\hat{\delta}_{boot}^2(x)=\frac{\sum_{b=1}^{B}\sum_{i=A_b}\left(y_i-f_{BNN}\left(x_i;{}^{w_i}/_{T^B}\right)\right)^2}{B-1} \quad (9)$$

The significant level "α%" confidence intervals indicate repeat the procedure. Typical value for "α", 0.05, which is evident to the "100(1-α%)" are equal to about 95%. In order to determine the level of the confidence intervals used in "equation (10)".

$$\hat{\theta}_{(X)}\pm t_{n-p}^{\alpha/2}\delta(x) \quad (10)$$

"$\delta(x)$" standard deviation, and "B" are the number of estimates from bootstrap. "T" value from Table "t – student" and "$t_{n-p}^{\alpha/2}$" percent "α/2" to distribute "t – student" with degrees of freedom is "n_p" ("p" the number of parameters in the neural network and "n" is the number of observations). Examples bootstrapped Bayesian Neural Network (BNN) in [20] have been examined. Bayesian neural network output has shown excellent performance of the bootstrapping method because the bootstrap corrects for the bias of the resubstitution error. Bootstrapping has been shown to provide very good estimates of error for statistical models, there are few instances of its use in the neural network literature.

Volume 4, Issue 3
June 2015

ITEE Journal
Information Technology & Electrical Engineering

ISSN: - 2306-708X

5. CONCLUSION

Bayesian Neural Network (BNN) to predict with bootstrapped do fewer errors than using artificial neural network. In general, neural networks and ensure higher credit bootstrapped Bayesian estimation compared with conventional artificial neural network. While the method can estimate the uncertainty of estimates as well. Choosing the number of validation networks for the bootstrap is a decision that can be made a priori or iteratively by the analyst.

REFERENCES

[1] B. Efron, "Bootstrap methods: another look at the jackknife," The annals of Statistics, pp. 1-26, 1979.
[2] H. Liang, W. Härdle, and V. Sommerfeld, "Bootstrap approximation in a partially linear regression model," Journal of statistical planning and inference, vol. 91, pp. 413-426, 2000.
[3] S. Bajgier, "The use of bootstrapping to construct limits on control charts," in Proceedings of the decision Science Institute, 1992, pp. 1611-1613.
[4] L. A. Franklin and G. S. WASSERMAN, "Bootstrap lower confidence limits for capability indices," Journal of Quality Technology, vol. 24, pp. 196-210, 1992.
[5] T. Seppala, H. Moskowitz, R. Plante, and J. Tang, "Statistical process control via the subgroup bootstrap," Journal of Quality Technology, vol. 27, pp. 139-153, 1995.
[6] R. Y. Liu and J. Tang, "Control charts for dependent and independent measurements based on bootstrap methods," Journal of the American Statistical Association, vol. 91, pp. 1694-1700, 1996.
[7] Z. Wu and Q. Wang, "Bootstrap control charts," Quality Engineering, vol. 9, pp. 143-150, 1996.
[8] M. Wood, M. Kaye, and N. Capon, "The use of resampleing for estimating control chart limits," Journal of the Operational Research Society, vol. 50, pp. 651-659, 1999.
[9] S. Teyarachakul, S. Chand, and J. Tang, "Estimating the limits for statistical process control charts: A direct method improving upon the bootstrap," European journal of operational research, vol. 178, pp. 472-481, 2007.
[10] H.-I. Park, "Median control charts based on bootstrap method," Communications in Statistics—Simulation and Computation®, vol. 38, pp. 558-570, 2009.
[11] R. B. Jana and B. P. Mohanty, "Enhancing PTFs with remotely sensed data for multi-scale soil water retention estimation," Journal of hydrology, vol. 399, pp. 201-211, 2011.
[12] A. Gelman and K. Shirley, "Inference from simulations and monitoring convergence," Handbook of Markov chain Monte Carlo, pp. 163-174, 2011.
[13] G. B. Kingston, M. F. Lambert, and H. R. Maier, "Bayesian training of artificial neural networks used for water resources [1]modeling," Water Resources Research, vol. 41, 2005.
[14] A. Goh and C. Chua, "Nonlinear modeling with confidence estimation using Bayesian neural networks, " Electronic Journal of Structural Engineering, vol. 1, pp. 108-118, 2004.
[15] J. C. Lemm, "Bayesian field theory: Technical Report MS-TP1-99-1," Universität Münster, Institut für Theoretische Physik, 1999.
[16] D. N. Politis and J. P. Romano, "The stationary bootstrap ," Journal of the American Statistical association, vol. 89, pp. 1303-1313, 1994.
[17] J. M. Twomey and A. E. Smith, "Bias and variance of validation methods for function approximation neural networks under conditions of sparse data," Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, vol. 28, pp. 417-430, 1998.
[18] W. E. Stewart, M. Caracotsios, and J. P. Sørensen, "Parameter estimation from multiresponse data," AIChE Journal, vol. 38, pp. 641-650, 1992.
[19] B. Efron and R. J. Tibshirani, An introduction to the bootstrap: CRC press, 1994.
[20] M. Sabri, R. Rezaeipour, " Improvement Estimation Power Flow Using Bayesian Neural Network." International Journal of Information Technology and Electrical Engineering, vol 4, no. 2, pp. 32-39, April 2015.

AUTHOR PROFILES

Mahdi Sabri was born in Tabriz, Iran, in 1990. He received his B.Sc. degree in electrical engineering from the Islamic Azad University, Tabriz branch, Tabriz, Iran in 2013.He is currently a M.Sc. student in department of electrical engineering in East Azarbaijan Science and Research Branch, Islamic Azad University, Tabriz, Iran.

He is a member of Young Researchers and Elite club in Islamic Azad University, Tabriz Branch. His research interests are power system analysis, neural networks, application of fuzzy control in power system and FACTS devices.

Roshanak Rezaeipour was born in Tabriz, Iran, in 1979. She received her B.Sc. degree in electrical engineering from the Islamic Azad University, Tabriz branch, Tabriz Iran in 2001. She received her M.Sc. and Ph.D. in electrical engineering from the University of Science and Technology, Tehran, Iran in 2008 and 2011, respectively.

She is an Associated Professor in department of electrical engineering, Tabriz Branch, Islamic Azad University, Tabriz, Iran. Her research interests are power system analysis, application of fuzzy control in power system and FACTS devices.