IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received March 12, 2020, accepted March 17, 2020, date of publication March 20, 2020, date of current version March 31, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982192

Predictions of USA Presidential Parties From 2021 to 2037 Using Historical Data Through Square Wave-Activated WASD Neural Network

TIANYU ZENG[1], YUNONG ZHANG[1,2], (Member, IEEE), ZHENYU LI[1], BINBIN QIU[1,2], (Member, IEEE), AND CHENGXU YE[3]

[1]School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China
[2]Research Institute of Sun Yat-sen University, Shenzhen 518057, China
[3]School of Computer Science, Qinghai Normal University, Xining 810008, China

Corresponding author: Yunong Zhang (zhynong@mail.sysu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61976230, in part by the Shenzhen Science and Technology Plan Project under Grant JCYJ20170818154936083, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2019A1515012128, in part by the Natural Science Foundation of Qinghai under Grant 2016-ZJ-739, in part by the China Postdoctoral Science Foundation under Grant 2018M643306, and in part by the Fundamental Research Funds for the Central Universities under Grant 19lgpy227.

ABSTRACT The United States of America (USA) has been the most powerful country in the world for decades. The country's global leadership and power are critical for international security and intercountry relationships, in which the political events yield differences. With the deepening development of globalization and the recovery of the global economy, the future policies and presidential parties of the USA should be predicted to avoid potential crises. However, making unbiased and robust predictions is challenging. This study presents a weight-and-structure-determination (WASD) algorithm-based feedforward neural network activated by a set of square wave functions to predict the presidential parties in the USA from 2021 to 2037. The historical data of presidential parties from 1853 to 2017 are used as the training set in the numerical experiments, and the results show that the proposed neural network can predict 9 terms of future presidential parties. Compared with other models, the proposed model is more efficient, robust, purely time-driven, and bias-independent. Through the square wave-activated WASD neural network, the Democratic Party is predicted to win in 2025, 2029, and 2033, whereas the Republican Party might win in 2021 and 2037.

INDEX TERMS Weight-and-structure-determination algorithm, feedforward neural network, USA, presidential parties, prediction.

I. INTRODUCTION

The United States of America (USA) has been the only superpower in the world since the disintegration of the Soviet Union, and this status is foreseen to last for decades [1]. The power and global leadership of the USA affect the world today and in the future. The presidential elections and the international policies generated by the USA's presidential parties always attract global attention and exert positive or negative effects on numerous globally cooperative fields, including technology and economy.

The associate editor coordinating the review of this manuscript and approving it for publication was Derek Abbott.

The political party system in the USA is a two-party system that is led by the Democratic Party (Dem.) and the Republican Party (Rep.) [2], [3]. Both parties have won the presidential elections since 1853 and controlled the Congress of the United States to some extent since 1856 [4]. The domestic and the international policies provided and advocated by the two parties are distinct from each other. The Democratic Party calls for the government's intervention in the economy, technology development, and social equality of the country [5]–[7], whereas the Democrats generate policies for universal health care and minimum wage increase. By contrast, the Republican Party is conservative on the political position of social issues and libertarian on the economy [5], [6], [8]. This party emphasizes lower taxes and the limited-sized government, supports the business development, and pursues close relationships with large corporations in Wall Street [6], [8]. Thus, different parties in power implement different policies in the USA and the world.

The USA is a great engine of the global economy and technology [9], [10]. Historically, the policies presented by the governors are largely affected by the political positions of the presidential parties. With the deepening development of globalization and the recovery of the global economy, the accurate prediction of future presidential parties is crucial for people to acclimatize to new regulations and policy changes and for leaders of foreign governments and international organizations to create strategy substitutions in advance to avoid potential diplomatic and legal problems. A relatively reliable and correct prediction is important for the easily affected fields, such as trading and economy. Numerous factors, such as speeches or debates, personal images, related news, and short-term policies, might affect the final election results [10]. However, these factors cannot be considered all at once because the data collection and quantification of these factors are difficult, and the relationships among them are hard to describe.

Predicting the winning presidential party is a traditional research theme in social science but a new problem in the field of math and computer science [11]. Many political scientists have proposed models to solve this problem. Abramowitz [12] proposed the time-for-change model for the winners of the elections from 1988 to 2012, considering the economy, mid-year support, and incumbency. However, this kind of model is short-time oriented and appears to have biases in the recent election. Thus, making long-time oriented and unbiased models is necessary. Nowadays, predictions can be generated through machine learning frameworks [13], [14] by learning the historical pattern. However, although many time series prediction neural networks are proposed [15], [16], the corresponding interpretability is lacking.

In this study, a weight-and-structure-determination (WASD) algorithm-based feedforward neural network (WASDNN) activated by a set of square wave functions is presented. The WASDNN is a noniterative neural network proposed by Zhang *et al.*[17], Chen *et al.* [18], Li *et al.* [19], Zhang *et al.* [20]; it avoids the regressive calculations in the error in the backpropagation algorithm, accelerates the training, and helps to obtain the most efficient and accurate network structure [21]. Given that the two parties' presidential situation is periodic to some extent, square wave activation functions that can satisfactorily solve periodic prediction problems are creatively proposed. Using the WASDNN and square wave activation functions, this study adopts a proper error evaluation method to achieve the best global bias for the network. Compared with other models, the proposed network is more efficient, purely time-driven, and bias-independent. A robust prediction model, which can precisely predict 9 terms of presidential election results, is obtained by training the neural network with historical data from the 1853–2017 election results. The WASD algorithm simplifies the network into a 14-neuron scale. The numerical experiments are self-supporting and show that the Democratic Party might win in 2025, 2029, and 2033, whereas the Republican Party is foreseen to win in 2021 and 2037.

The remainder of this paper is organized as follows: Section II presents the theoretical basis of the WASD algorithm and corresponding dependent neural network. In Section III, the details of the square wave activation functions and the prediction model are discussed. The historical data learning procedure and tests are presented in Section IV. The final predictions and analysis for the USA presidential parties from 2021 to 2037 are provided in Sections V and VI, respectively. Section VII summarizes the limitations, and Section VIII concludes this study.

II. WASD ALGORITHM AND NEURAL NETWORK

In this section, the WASD algorithm and the construction of the WASDNN are briefly discussed.

Machine learning is a computationally heavy task, and the iterative training error calculation is not a swift procedure. The WASD algorithm for neural network training is proposed to accelerate this procedure and simplify the network structure [17].

In contrast to the iterative weight training in traditional neural networks, the weights in the WASDNN are obtained by the weight-direct-determination (WDD) algorithm [22]. In other words, the weights are learned in a directly pseudo-inverse method. According to previous studies [22], the WDD algorithm can be described as

$$\boldsymbol{w} = [w_0, w_1, w_2, \ldots, w_{N-1}]^{\mathrm{T}} = (A^{\mathrm{T}}A)^{-1}A^{\mathrm{T}}\boldsymbol{y} = A^{+}\boldsymbol{y}, \quad (1)$$

where the vector $\boldsymbol{w}$ is the weight vector for the network, and the vector $\boldsymbol{y}$ is the target output of the neural network. Matrix A refers to the activation matrix with elements that serve as corresponding outputs of the activation functions for the inputs, whereas A^{+} refers to the pseudo-inverse matrix of A.

The procedure and the details of the WASD algorithm are shown in Figure 1 and Algorithm 1, respectively. The WASD algorithm returns the local best neural network structure N_{opt}

Algorithm 1 WASD Algorithm for Neural Network

Require: The input vector $\boldsymbol{x}$, the output vector $\boldsymbol{y}$, the activation matrix function φ and a given step parameter s.

1: Initialize: the network, $N = 0$, and the error $e_{\min} = \infty$.
2: $N \leftarrow N + 1$.
3: Calculate the activation matrix A and its pseudo-inverse A^{+} with the given input $\boldsymbol{x}$ and the activation function φ.
4: Calculate the weight vector $\boldsymbol{w}$ and the error e between $\boldsymbol{y}$ and $\hat{\boldsymbol{y}} = A\boldsymbol{w}$.
5: If $e \leq e_{\min}$, then $e_{\min} \leftarrow e$, $N_{\mathrm{opt}} \leftarrow N$, $\boldsymbol{w}_{\mathrm{opt}} \leftarrow \boldsymbol{w}$, and go to Step 2.
6: Else: if $N < N_{\mathrm{opt}} + s$, then go to Step 2.
7: **return** $e_{\min}$, N_{opt}, and $\boldsymbol{w}_{\mathrm{opt}}$.

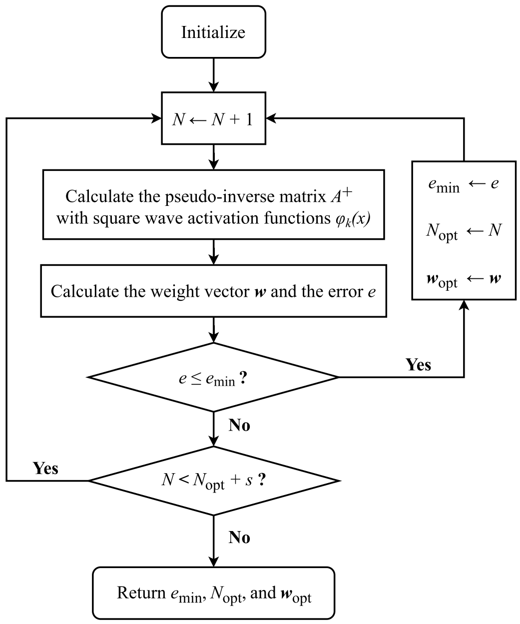


FIGURE 1. Flowchart of WASD algorithm for square wave-activated neural network.

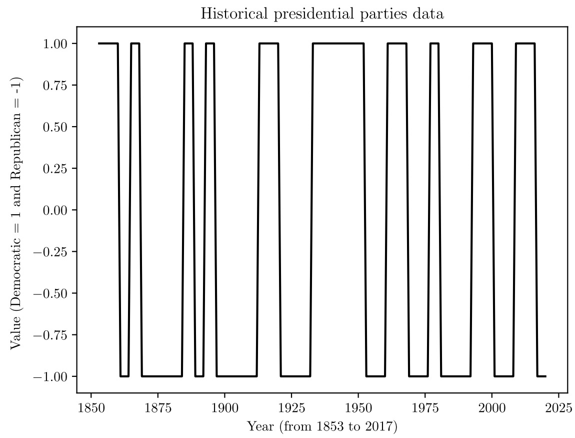


FIGURE 2. Historical data visualization of USA presidential parties.

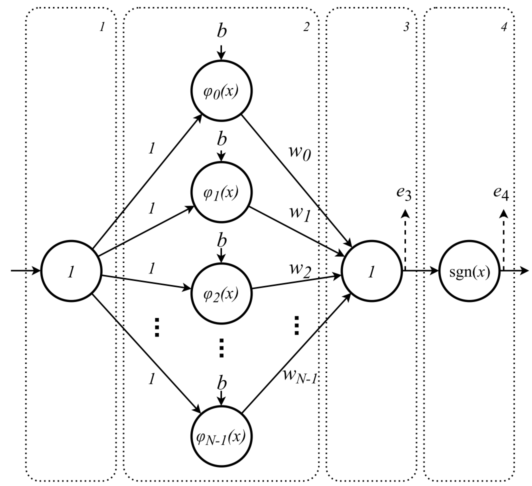


FIGURE 3. Structure of square wave-activated WASDNN.

and the corresponding weight vector $\boldsymbol{w}_{\text{opt}}$. The step parameter s, which is given by the user in terms of minimum errors, is utilized to avoid overfitting problems. Steps 3 and 4 are implemented to accelerate the training procedure, whereas Steps 5 and 6 are designed to determine the optimal parameters for the network. By repeating the algorithm, $\boldsymbol{w}_{\text{opt}}$ automatically fits the target weight.

III. SQUARE WAVE-ACTIVATED WASDNN

In this section, we propose the square wave activation functions and the square wave-activated WASDNN.

A. SQUARE WAVE ACTIVATION FUNCTIONS

The activation functions in traditional neural networks are typically nonperiodic (e.g., Chebyshev polynomials [23], [24], sigmoid functions, and exponential functions). However, for time series prediction tasks, these functions always require normalization on the data sets and the prediction targets. For the prediction task for USA presidential parties and other periodic problems, using periodic activation functions in the neurons is more natural.

For this problem, the historical data of the presidential parties in the USA are encoded and visualized in Figure 2. Let 1 denote the terms when the Democratic Party comes into power, and -1 denote the terms when the Republican Party dominates. The trend suggests that the election results are periodic square waves, and the presidential period is approximately 8–12 years. The proposed square wave function can be expressed as

$$\varphi_k(x) = \begin{cases} 1 & \text{if } k = 0, \\ 1 - 4\left(\frac{\lfloor x/k \rfloor}{2} - \left\lfloor \frac{\lfloor x/k \rfloor}{2} \right\rfloor\right) & \text{if } k \in \mathbb{N}^*, \end{cases} \tag{2}$$

where x refers to the function input, and the parameter k refers to the half period of this function. As sine-similar functions [25], these square wave functions also have a phase parameter for the input x. The other properties of these functions are similar to that of the sine function.

B. SQUARE WAVE-ACTIVATED WASDNN

In this study, a four-layer feedforward neural network, whose structure is shown in Figure 3, is adopted.

The first layer is the input layer, where only one nonactivated neuron is presented. This layer receives and distributes the input vector $\boldsymbol{x}$ to the neuron in the next layer with equal weight 1. The second layer, known as "the N-neuron network", has N square wave-activated neurons. The neurons receive the value from the first layer as the inputs and share the same global bias b. The corresponding activation functions $\varphi_k(x)$ are square wave functions with different periods. The third layer consists of a nonactivated neuron. The weights w_k in the neuron-to-neuron link between the second and the third layers' neurons form the weight vector $\boldsymbol{w}$. Through the WASD and the WDD algorithm, the weight vector $\boldsymbol{w}_{\text{opt}}$ and the structure N_{opt} can be learned in a short time. To fix the potential overfitting problems, the elements in the weight vector $\boldsymbol{w}$ should follow the following principle:

$$\sum_{i=0}^{N-1} |w_i| = 1. \tag{3}$$

An extra neuron with a sign function [26] as the activation function is added as the fourth layer to amplify the election results and fit the data. The sign function is defined as

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ -1 & \text{if } x < 0. \end{cases} \tag{4}$$

For error evaluation, three errors are evaluated and used as the conditions in Step 5 of Algorithm 1. The first error e_3 evaluates the distance between the outputs $\hat{\boldsymbol{y}}$ from the third layer and the real election results $\boldsymbol{y}$, which can reduce the underfitting problems efficiently. The formula of e_3 is

$$e_3 = \left|\left|\boldsymbol{y} - \hat{\boldsymbol{y}}\right|\right|_2^2. \tag{5}$$

The second error is e_4, which evaluates the distance between the outputs $\operatorname{sgn}(\hat{\boldsymbol{y}})$ from the fourth layer and the real election results $\boldsymbol{y}$. Contrary to e_3, e_4 can reduce overfitting problems. The formula of e_4 is

$$e_4 = \sum_{i=1}^{n} \left| y_i - \operatorname{sgn}\left(\hat{y}_i\right) \right|, \tag{6}$$

where n is the number of training terms; and y_i and $\operatorname{sgn}(\hat{y}_i)$ represent the ith element of $\boldsymbol{y}$ and $\operatorname{sgn}(\hat{\boldsymbol{y}})$, respectively. The prediction-type test error e_{p} is also calculated in the same manner as e_4 using the test data.

The global bias b is an important value for the square wave-activated WASDNN. For the data and the target functions, the bias can affect the fitting and the result errors. The bias b can be shared globally by the neurons without making errors due to the network's structure and the data. Moreover, certain conditions can be applied to limit the error and determine the optimal b for certain problems.

IV. HISTORICAL DATA LEARNING AND TESTS

In previous sections, the details of the square wave-activated WASDNN are discussed. In this section, the proposed network is applied to training and tests to determine whether the developed model can learn the pattern and predict reasonably from the historical data.

A. HISTORICAL DATA LEARNING

The newly proposed square wave-activated WASDNN is then trained to learn the pattern of the historical data of the USA presidential parties.

Table 1 displays the historical data, whereas Figure 2 visualizes the coded results. According to previous sections, the election results have an 8-more-year short-term period; thus, the activation functions in the neurons are 8-time periodic, which means that the half periods can be 4, 8, 12, and so on.

TABLE 1. Historical data of USA presidential parties from 1853 to 2017.

Years	Presidential party	Years	Presidential party
1853–1857	Democratic	1937–1941	Democratic
1857–1861	Democratic	1941–1945	Democratic
1861–1865	Republican	1945–1949	Democratic
1865–1869	Democratic	1949–1953	Democratic
1869–1873	Republican	1953–1957	Republican
1873–1877	Republican	1957–1961	Republican
1877–1881	Republican	1961–1963	Democratic
1881–1885	Republican	1963–1969	Democratic
1885–1889	Democratic	1969–1974	Republican
1889–1893	Republican	1974–1977	Republican
1893–1897	Democratic	1977–1981	Democratic
1897–1901	Republican	1981–1985	Republican
1901–1905	Republican	1985–1989	Republican
1905–1909	Republican	1989–1993	Republican
1909–1913	Republican	1993–1997	Democratic
1913–1917	Democratic	1997–2001	Democratic
1917–1921	Democratic	2001–2005	Republican
1921–1923	Republican	2005–2009	Republican
1923–1929	Republican	2009–2013	Democratic
1929–1933	Republican	2013–2017	Democratic
1933–1937	Democratic	2017–	Republican

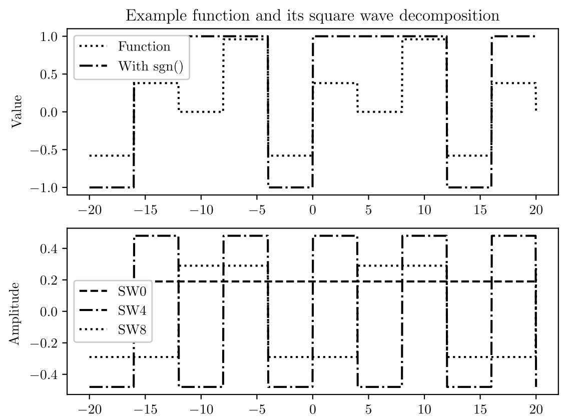


FIGURE 4. Example for square wave decomposition.

An example of a sign-amplified function and its 8-time periodic square wave decomposition are displayed in Figure 4. For the step parameter s, a proper value is set to escape the local minimum of errors and determine the possible global minimum. In the experiments, $s = 10$.

The historical dataset is divided into two parts. The first n-term part is the training set, which contains the presidential records from the first term (1853–1857) to the nth term of the data. The rest serves as the test set. The most efficient and the most stable structures for this problem are obtained by using the WASD algorithm with the errors e_3 and e_4 and the prediction-type test error e_{p}. The training error trends are shown in Figures 5 and 7.

The most efficient network structure contains 14 neurons in the second layer (i.e., the 14-neuron network). The spectrum of the 36-term training results is displayed in Figure 6. From the results listed in Table 2, the 14-neuron network can predict 9 terms of the presidential election precisely. This network generates prediction errors on the test sets when

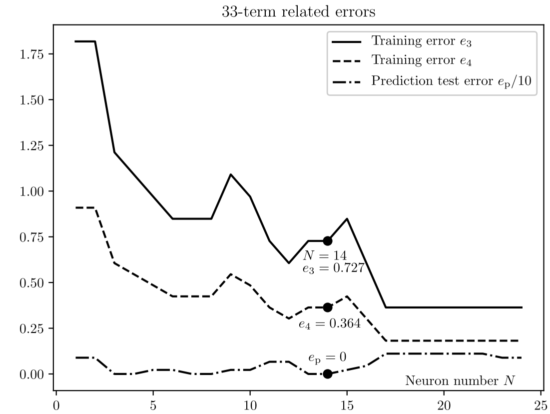


FIGURE 5. 33-term training and prediction-type test errors.

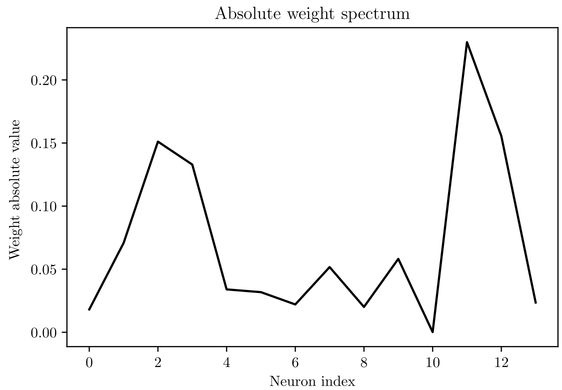


FIGURE 6. Spectrum of 36-term 14-neuron training results in Table 2.

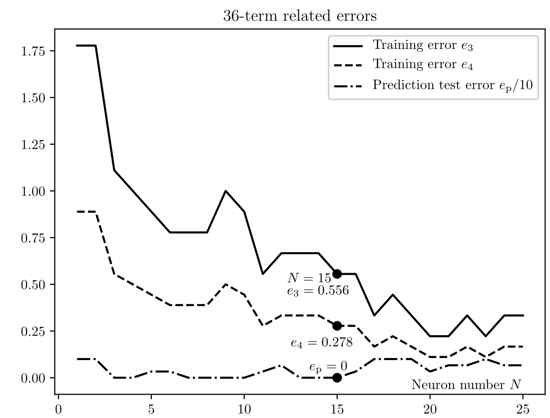


FIGURE 7. 36-term training and prediction-type test errors.

$n \leq 32$. When the training term number $n = 33, 34, 37$, the model's test results are precise, and the far-future predictions are of good stability with some of n, indicating that most of the far-future predictions are consistent with the variation of the training term number but only a few of them are inconsistent. However, the 14-neuron network can generate errorless predictions and stable future results on the test sets when $n = 35, 36$ or $n \geq 38$, suggesting that the results are all consistent with varying n and are of the best stability.

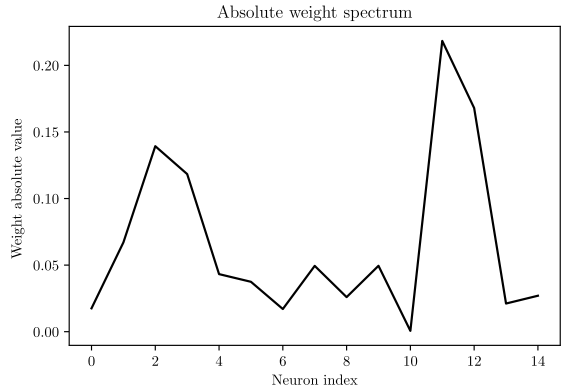


FIGURE 8. Spectrum of 36-term 15-neuron training results in Table 3.

TABLE 2. Relationship between *n*-term training part and its results using 14-neuron network.

No. of training terms n	32	33	34	35	36
Neuron number N	14	14	14	14	14
Best global bias b	32	32	32	32	32
Training error e_3	0.500	0.727	0.706	0.571	0.666
Training error e_4	0.250	0.364	0.353	0.285	0.333
Prediction test error e_p	0.678	0.0	0.0	0.0	0.0
Prediction stability	False	Good	Good	Best	Best
No. of training terms n	37	38	39	40	41
Neuron number N	14	14	14	14	14
Best global bias b	32	32	32	32	32
Training error e_3	0.757	0.737	0.615	0.700	0.683
Training error e_4	0.378	0.368	0.307	0.350	0.341
Prediction test error e_p	0.0	0.0	0.0	0.0	0.0
Prediction stability	Good	Best	Best	Best	Best

The most stable network is the 15-neuron network, which is more accurate in terms of e_3 (Table 3) and has a similar 36-term training result spectrum (Figure 8) with the 14-neuron network. This network can provide predictions with the best stability when the training term number $n \geq 34$. However, this model can only predict eight terms of the presidential election, and to some extent, encounter overfitting.

What is worth noting is that the prediction results of the 14-neuron and the 15-neuron networks are the same when the training term number $n \geq 38$, which implies that the latter can be used as a supportive model for the former and can firmly support the final prediction results.

B. NEURAL NETWORK TESTS

In this section, two kinds of tests are conducted for the new networks. The first is the prediction of the test set. The 14-neuron and the 15-neuron networks are both trained with the 36-term training set and tested with the 6-term test set. The test results, shown in Figures 9 and 10, for the two models are similar to each other. The test results under other situations are listed in Tables 2 and 3. The prediction results from both networks are the same when $n = 35, 36$ and $n \geq 38$.

The other test is the exchanged coding test. As previously stated, 1 denotes the terms when the Democratic Party wins, and -1 denotes the terms when the Republican Party wins. The values are exchanged (i.e., -1 denotes the winning terms

TABLE 3. Relationship between *n*-term training part and its results using 15-neuron network.

No. of training terms n	32	33	34	35	36
Neuron number N	15	15	15	15	15
Best global bias b	32	32	32	32	32
Training error e_3	0.500	0.848	0.588	0.685	0.556
Training error e_4	0.250	0.424	0.294	0.342	0.278
Prediction test error e_p	0.697	0.598	0.0	0.0	0.0
Prediction stability	False	False	Best	Best	Best
No. of training terms n	37	38	39	40	41
Neuron number N	15	15	15	15	15
Best global bias b	32	32	32	32	32
Training error e_3	0.541	0.526	0.513	0.600	0.585
Training error e_4	0.270	0.263	0.256	0.300	0.292
Prediction test error e_p	0.0	0.0	0.0	0.0	0.0
Prediction stability	Best	Best	Best	Best	Best

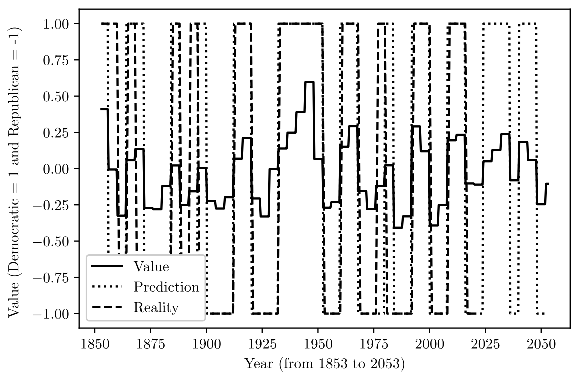


FIGURE 9. Prediction results (including prediction-type test results) after 36-term 14-neuron training.

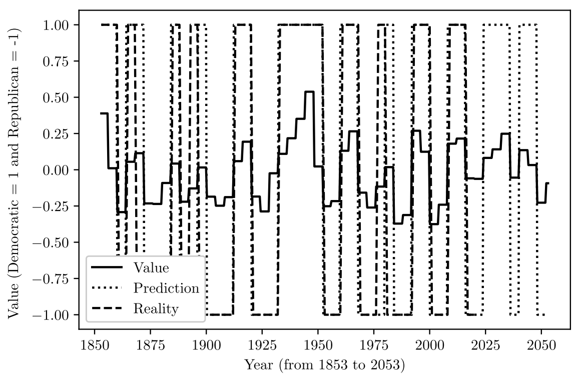


FIGURE 10. Prediction results (including prediction-type test results) after 36-term 15-neuron training.

for the Democratic Party, and 1 denotes the winning terms for the Republican Party) in this test to check whether the initial value affects the predictions and the balance of the neural network. The 36-term training and 6-term prediction results shown in Figures 11 and 12 are the reverse of the original results, which meets the expected output.

V. FINAL PREDICTIONS FOR USA PRESIDENTIAL PARTIES FROM 2021 TO 2037

In previous sections, the most efficient and most stable WASDNN structures are obtained. Tables 2 and 3 indicate

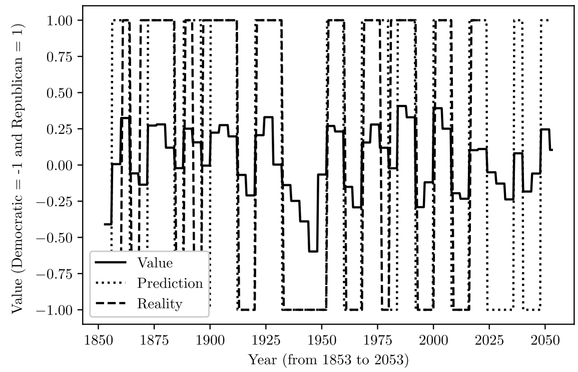


FIGURE 11. Prediction results (including prediction-type test results) after 36-term 14-neuron training with exchanged coding.

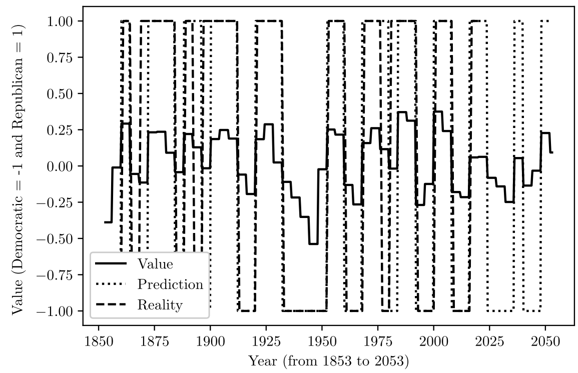


FIGURE 12. Prediction results (including prediction-type test results) after 36-term 15-neuron training with exchanged coding.

TABLE 4. USA presidential party predictions from 2021 to 2037 via 14-neuron network (5 terms).

Year	2021	2025	2029	2033	2037
Party prediction	Rep.	Dem.	Dem.	Dem.	Rep.
Value	-0.162	0.029	0.110	0.208	-0.093

TABLE 5. USA presidential party predictions from 2021 to 2037 via 15-neuron network (5 terms).

Year	2021	2025	2029	2033	2037
Party prediction	Rep.	Dem.	Dem.	Dem.	Rep.
Value	-0.136	0.045	0.114	0.220	-0.073

that the 14-neuron and the 15-neuron networks have the same predictions when training with sufficient data. In this section, the 14-neuron and the 15-neuron networks are trained with all the historical data in Table 1.

After the two square wave-activated WASDNNs training with all the data, the obtained final prediction results are self-supporting and stable. Taking into account that the longer the predictions are, the greater the impact of other factors is, the prediction results for the first 5 presidential terms are listed in this section, although the proposed WASDNN can predict the next 9 terms. The 42-term training and the prediction details are listed in Tables 4 and 5, respectively, and the final predictions are listed in Table 6. The final predictions

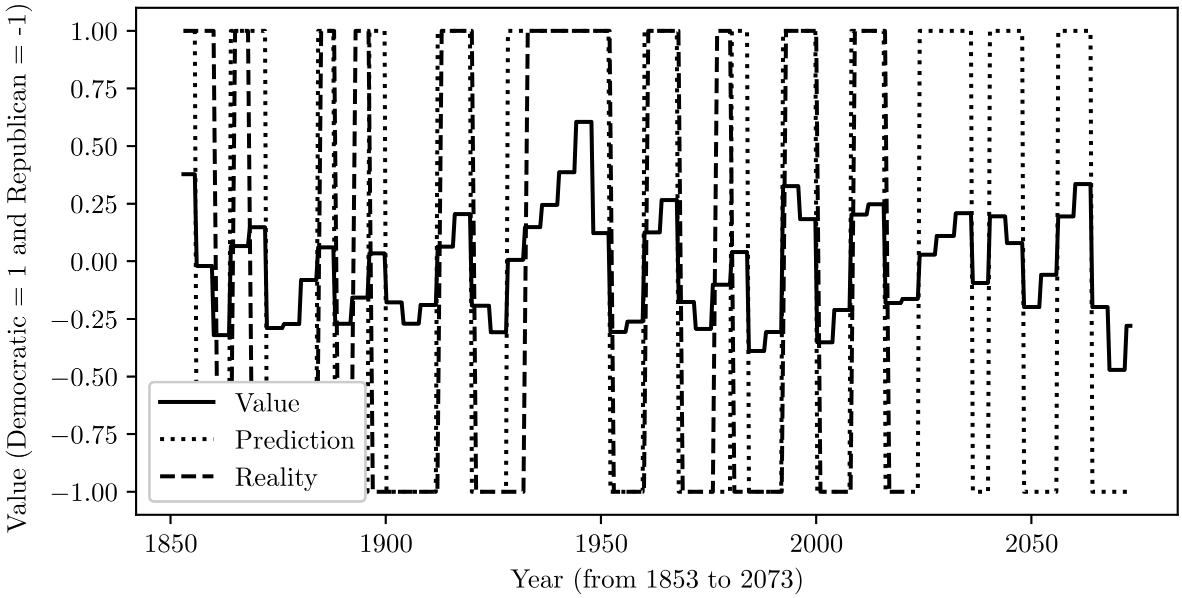


FIGURE 13. USA presidential party predictions via 14-neuron network.

TABLE 6. Final USA presidential party predictions from 2021 to 2037.

Year	Party prediction	Year	Party prediction
2021	Republican Party	2033	Democratic Party
2025	Democratic Party	2037	Republican Party
2029	Democratic Party		

state that the Democratic Party might win in 2025, 2029, and 2033, whereas the Republican Party is foreseen to dominate the 2021 and 2037 terms. The evident consistency of the two networks' predictions can sustain confidence in the final predictions. The results are plotted in Figure 13 to show a long-term waveform of the presidential predictions. The plot then verifies that the proposed neural network is stable and can efficiently learn the pattern from the historical data.

VI. FINAL PREDICTION ANALYSIS

In this section, the final prediction results are analyzed to verify whether the networks have learned the pattern of the historical data or not from the aspect of history.

The respective predictions of the 14-neuron and the 15-neuron networks from 2021 to 2037 are presented in Tables 4 and 5. The historical data and predictions from 2021 to 2073 are then combined and plotted in Figure 13. On the basis of the value trend line in Figure 13, the value within $[-1, 1]$ is balanced and constant in a 4-year time. The predictions can be obtained with the sign function defined in Equation (4).

The prediction line in Figure 13 implies that the results are balanced and reasonable from the aspect of history. For 2021, the Republican Party is predicted to win because the neuron with an index of 2 and a half period of the activation function of 8 years has the greatest weight in the spectrums which are shown in Figures 6 and 8. For 2025–2033, the Democratic Party is predicted to win the elections, because the spectrums shown in Figures 6 and 8 exhibit a strong amplitude on the neuron with an index of 3 and a half period of the activation function of 12 years. The historical election data illustrate the same long-single-party-won situation during 1933–1953. In 2037, the Republican Party might win again. In general, the Democratic Party and the Republican Party alternately dominate in 2025–2037 and 2017–2025, respectively, for approximately 8–12 years, which is balanced, convincing, and reasonable with the 8-year-period common-sense approach on the historical USA presidential elections.

VII. LIMITATIONS OF STUDY

This study introduces the square wave-activated WASDNN and the corresponding 5-term predictions of the USA presidential parties. The proposed method shows the advance in making binary predictions, however, every predictive method has its own shortages. In this section, three limitations are summarized for the proposed method.

The first limitation is the data dependency. The historical data contain the long-term pattern of the elections, and the pattern would last for a long period according to statistics if no political incidences happened. However, the incidences, such as candidates exiting and parties splitting, are hard to be predicted by the proposed method. Thus, the data dependency may be one of the limitations. The second limitation is the data diversity. The historical data imply a part of the social indices but not all. Therefore, the diversity of the data could be a potential limitation. At last, the third limitation is the optimization of the shared bias. Unlike the traditional backpropagation neural networks, the proposed method is based on the WASD feedforward neural network. Thus, the optimization method is a global optimization method, such as gird search and differential evolution, which could be less efficient than the backpropagation algorithm.

VIII. CONCLUSION

This study has introduced the square wave activation functions and proposed the square wave-activated WASDNN. What is more, this new neural network has been utilized to learn and predict the presidential parties of the USA from 2021 to 2037.

As the only superpower in the world, the USA plays an important role in numerous global cooperative fields, and the policies generated by the presidential parties impact domestically and globally. Although the election results can be affected by many sophisticated factors, general predictions can be generated by learning the historical data pattern. Consistent with the saying "Reading history can make people learn wisely from the past and know the future", the proposed square wave-activated WASDNN has achieved satisfactory prediction performance. According to the final predictions, it can be foreseen that the Democratic Party might dominate 2025, 2029, and 2033 terms, whereas the Republican Party might win in 2021 and 2037.

APPENDIXES

APPENDIX A

COMPARISONS BETWEEN MODELS

This section presents the comparisons between the proposed WASDNN and the other models.

A. COMPARED WITH OTHER ACTIVATION FUNCTIONS

First, the proposed square wave activation functions are compared with nonperiodic ones. The comparison targets are the sigmoid functions, the exponential functions, and Chebyshev polynomials [23], [24], all of which require normalized training and prediction sections. The sigmoid and exponential functions can be expressed as

$$\sigma_k(x) = \frac{1}{1+\exp(-x/k)}, \quad \mathrm{Exp}_k(x) = \exp(x/k),$$

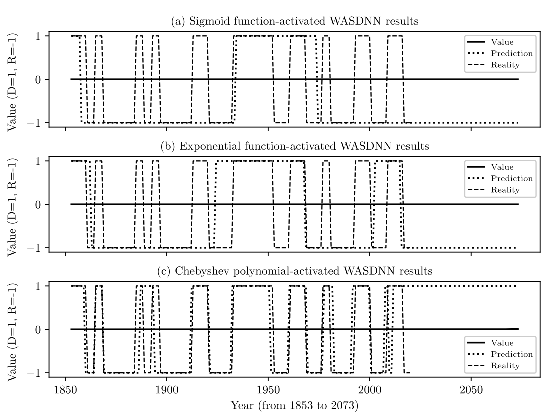


FIGURE 14. Comparison results between the WASDNNs activated by square wave, (a) sigmoid, (b) exponential functions, and (c) Chebyshev polynomials.

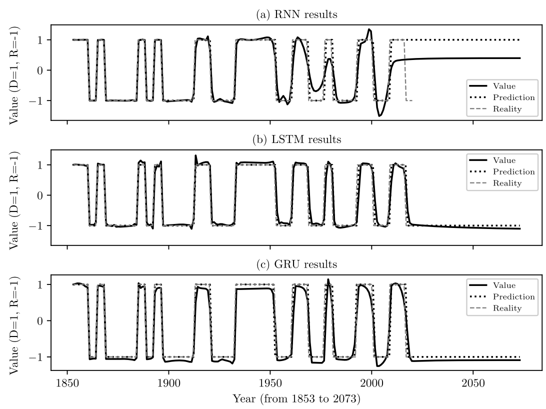


FIGURE 15. Comparison results between square wave-activated WASDNN and (a) RNN, (b) LSTM, (c) GRU networks.

respectively, where k refers to the neuron index. The results are shown in Figure 14. Compared with the results in Figure 13, the prediction values of the traditional methods are close to 0 and suffer from underfitting. The neuron weights are small, and the prediction values are at scales of 10^{-8} due to the up-to-32 neuron counts and Equation (3). The values can be amplified by Equation (4), but the errors and results are not acceptable. Thus, the proposed square wave activation functions are more suitable for this problem.

B. COMPARED WITH OTHER NEURAL NETWORKS

Second, the proposed WASDNN is compared with other neural networks, namely, recurrent neural network (RNN) [15], the long short-term memory (LSTM) network [16], and the gated recurrent unit (GRU) network [27]. These networks are frequently used in prediction tasks. The 8000-time training results of the three networks are shown in Figure 15. The comparison results indicate that it is hard for the existing models to learn the pattern in a short time. In conclusion, the proposed square wave-activated WASDNN is more suitable for this few-shot binary prediction problem, for it is square wave-activated and weight-and-structure-determination.

C. ANALYSIS ON COMPARISONS

Based on the comparative analysis results, the proposed square wave-activated WASDNN is more satisfactory than the WASDNN activated by other activation functions and other neural networks. In addition to these models, the proposed model is compared with other classification methods, such as logistic regression and AdaBoost [28]. However, the results are similarly not satisfying and convincing.

The key feature of the proposed model is the square wave functions that resemble the time-varying properties of the sine function and can efficiently deal with binary series data. The WASDNN is a noniterative feedforward network; thus, the weights are directly obtained and are less likely to suffer from data overfitting. What is more, the data of 168 years are sufficient for the proposed square wave-activated WASDNN to converge in terms of the errors in Tables 2 and 3. Thus, the

TABLE 7. Historical economy-presidential-parties data of USA from 1997 to 2017, where "Unemployment up" checks whether the unemployment rate of the term ending year grows up compared with the start year's one, and "GDP up" checks whether the one-term GDP growth rate grows up compared with the previous one.

Year	Unemployment growth	GDP growth rate	Unemployment up	GDP up	Two-term	Proposed WASDNN	Reality
1997	−2.10%	23.00%	False	False	False	Democratic Party	Democratic Party
2001	−1.41%	25.68%	False	True	True	Republican Party	Republican Party
2005	1.54%	20.57%	True	False	False	Republican Party	Republican Party
2009	0.25%	23.71%	True	True	True	Democratic Party	Democratic Party
2013	2.29%	8.75%	True	False	False	Democratic Party	Democratic Party
2017	−3.20%	16.06%	False	True	True	Republican Party	Republican Party

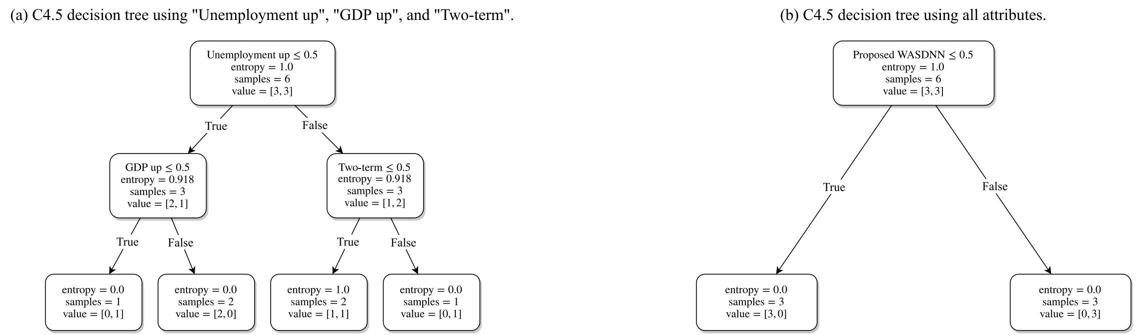


FIGURE 16. C4.5 decision tree classification results using (a) "Unemployment up", "GDP up", and "Two-term", (b) all attributes listed in Table 7.

proposed WASDNN might be more expeditious in predicting this problem.

APPENDIX B
ABOUT ELECTION PREDICTION FACTORS

As previously mentioned, lots of factors might affect the final election results, and collecting them could be difficult. The inputs of the proposed model are the years, and the model is purely time-driven. However, not all factors are difficult to collect and use. For instance, in the time-for-change model [12], the economic indices and the mid-year support rate also can be considered into the models.

However, these kinds of data might possess biases. First, these data only represent the previous work of the current presidential government instead of the party, and collecting these data might cause latency. When it comes to the elections, especially when the economy is growing or during the first term of the candidates from both parties, the models relying on these data will favor the current presidential party. Second, in recent years, the information collection methods vary, and the traditional media is less impactful today compared with before due to the development of the Internet. With the steady recovery of the economy and increasing exposure to the information, people might intend to select the leader that they personally want. Thus, the mid-year support rate and the economic indices may exert less effect on the general-election years compared with that on the re-election years, which can explain the failure of the time-for-change model in predicting the 2016 presidential election.

There is one thing that is sure: Time is fair for all the candidates in the elections. The USA has a history of more than 200 years, and the people have elected their presidents and the presidential parties under various kinds of situations. The historical data pattern contains the people's preferences and can fairly represent the data like the economic indices and the unemployment rate. Moreover, time is also considered in the previous presidential prediction models, which is why time is utilized as the major factor in this research. In the prediction-type tests in Section IV, the 14-neuron network utilizes the historical data of the USA presidential parties from 1853 to 1981 as the training set and precisely predict the election results from 1984 to 2016 (9 terms). This finding shows that the pattern works under 4 unemployment tides and the 2007 international financial crisis [29]. Additionally, another numerical experiment is conducted to illustrate the point more clearly. The data of the experiment consists of the historical data presented in Table 1, the differential growth rates of the one-presidential-term GDP [12] and the unemployment rate [29] in one presidential term. Table 7 lists the data and the corresponding binarizations. By using the C4.5 decision tree [28], the classification results under different attributes are shown in Figure 16, and the information entropy gains of the attributes are calculated and displayed as "entropy" in the corresponding nodes. One can see from the figure that using these three factors to classify the presidential parties would generate errors, whereas classifying the results with the proposed WASDNN's predictions would simplify the tree complexity and generate the predictions without errors. Therefore, it is clear that the time and the corresponding pattern are the two of the main factors and eligible to become the factors of the prediction task.

APPENDIX C
LONG-TERM PREDICTIONS TO 2053

In this study, a 20-year prediction generated by the proposed method is presented. However, according to the prediction

TABLE 8. USA presidential party predictions from 2021 to 2053 via 14-neuron network (9 terms).

Year	2021	2025	2029	2033	2037
Party prediction	Rep.	Dem.	Dem.	Dem.	Rep.
Value	-0.162	0.029	0.110	0.208	-0.093
Year	2041	2045	2049	2053	
Party prediction	Dem.	Dem.	Rep.	Rep.	
Value	0.194	0.078	-0.198	-0.058	

TABLE 9. USA presidential party predictions from 2021 to 2049 via 15-neuron network (8 terms), where "2053*" means that the year 2053 is beyond the prediction capability and the result is only for reference.

Year	2021	2025	2029	2033	2037
Party prediction	Rep.	Dem.	Dem.	Dem.	Rep.
Value	-0.136	0.045	0.114	0.220	-0.073
Year	2041	2045	2049	2053*	
Party prediction	Dem.	Dem.	Rep.	Rep.	
Value	0.164	0.062	-0.187	-0.051	

TABLE 10. Final USA presidential party predictions from 2021 to 2053.

Year	Party prediction	Year	Party prediction
2021	Republican Party	2041	Democratic Party
2025	Democratic Party	2045	Democratic Party
2029	Democratic Party	2049	Republican Party
2033	Democratic Party	2053	Republican Party
2037	Republican Party		

capability mentioned in Section IV, an extra 9-term prediction can be made and is presented in this section. Tables 8 and 9 list the predictions in the next 9 presidential terms generated by the 14-neuron and 15-neuron networks, respectively. From the predictions, one can see that the predictions are consistent and the proposed predictive method are self-supporting, which implies the same situations in Tables 4 and 5. Then, the results are summarized and formed into Table 10. The table indicates that the Democratic Party might dominate 2025, 2029, 2033, 2041, and 2045 terms, whereas the Republican Party might win in 2021, 2037, 2049, and 2053. Generally, besides the results in Table 6, the Democratic Party and the Republican Party will alternately dominate in 2041–2049 and 2049–2057, respectively, which is similar to the 8-year-period common sense. Therefore, the proposed predictive method is effective, efficient and self-supporting for predicting the future presidential parties of the USA.

REFERENCES

[1] L. Bremmer, "These are the 5 reasons why the U.S. Remains the world's only superpower," *Time*, May 2015. [Online]. Available: https://time.com/3899972/us-superpower-status-military/

[2] K. Polina, "Quantum like modelling of the nonseparability of voters preferences in the U.S. political system," *Lect. Notes Comput. Sci.*, vol. 8951, pp. 196–209, Feb. 2015.

[3] D. Dulio and J. Thurber, "America's two-party system: Friend or foe?" *Administ. Law Rev.*, vol. 52, no. 2, pp. 769–792, 2000.

[4] Boundless Political Science contributors. *The Two-Party System*. Accessed: Jul. 23, 2019. [Online]. Available: https://courses.lumenlearning.com/boundless-politicalscience/chapter/the-two-party-system/

[5] W. McWilliams, P. Deneen, and S. McWilliams, *The Democratic Soul: A Wilson Carey McWilliams Reader*. Lexington, KY, USA: Univ. Press Kentucky, 2011.

[6] G. Miller and N. Schofield, "The transformation of the republican and democratic party coalitions in the U.S.," *Perspect. Politics*, vol. 6, no. 3, pp. 433–450, Sep. 2008.

[7] M. Norton, J. Kamensky, C. Sheriff, D. Blight, and H. Chudacoff, *A People and A Nation: A History of the United States*, Toronto, ON, Canada: Nelson Education, 2014.

[8] G. Layman, *The Great Divide: Religious and Cultural Conflict in American Party Politics*, New York, NY, USA: Columbia Univ. Press, 2001.

[9] H. Berghel, "Oh, what a tangled Web: Russian hacking, fake news, and the 2016 US presidential election," *Computer*, vol. 50, no. 9, pp. 87–91, 2017.

[10] Y. Zhang, J. Wu, R. Wang, C. Ye, and Y. Ling, "From historical data through ASF method of equal-type half traversal to predict USA presidential parties from 2021 to 2041," in *Proc. Int. Conf. Ubiquitous Intell. Comput. (UIC)*, Guangzhou, China, 2018, pp. 1402–1407.

[11] M. Ebling, "Could pervasive computing influence democracy," *IEEE Pervas. Comput.*, vol. 16, no. 3, pp. 4–6, Jul. 2017.

[12] A. I. Abramowitz, "Forecasting the 2008 presidential election with the time-for-change model," *Political Sci. Politics*, vol. 41, no. 4, pp. 691–695, Oct. 2008.

[13] A. Shrestha and A. Mahmood, "Review of deep learning algorithms and architectures," *IEEE Access*, vol. 7, pp. 53040–53065, 2019.

[14] C. Bousono-Calzon, J. Bustarviejo-Munoz, P. Aceituno-Aceituno, and J. J. Escudero-Garzas, "On the economic significance of stock market prediction and the no free lunch theorem," *IEEE Access*, vol. 7, pp. 75177–75188, 2019.

[15] D. Rumelhart, G. Hinton, and R. Williams, "Learning representations by back-propagating errors," *Cognit. Model.*, vol. 5, no. 3, pp. 696–699, 1988.

[16] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[17] Y. Zhang, D. Chen, and C. Ye, *Toward Deep Neural Networks: WASD Neuronet Models, Algorithms, and Applications*, Boca Raton, FL, USA: Chapman & Hall, 2019.

[18] L. Chen, Z. Huang, Y. Li, N. Zeng, M. Liu, A. Peng, and L. Jin, "Weight and structure determination neural network aided with double pseudoinversion for diagnosis of flat foot," *IEEE Access*, vol. 7, pp. 33001–33008, 2019.

[19] H. Li, Z. Huang, J. Fu, Y. Li, N. Zeng, J. Zhang, C. Ye, and L. Jin, "Modified Weights-and-Structure-Determination neural network for pattern classification of flatfoot," *IEEE Access*, vol. 7, pp. 63146–63154, 2019.

[20] Y. Zhang, S. Ding, J. Wen, Y. Ding, and M. Mao, "Combining WASP and ASF algorithms to forecast a japan earthquake with mj 7.2 or above," in *Proc. 12th World Congr. Intell. Control Autom. (WCICA)*, Jun. 2016, pp. 2407–2413.

[21] Y. Zhang, P. He, J. Wen, Y. Ding, and Y. Ling, "Potential MIC season of US since Mar. 2017 or so predicted via WASD neuronets and sine wave with historic data and events," in *Proc. Int. Conf. Comput. Sci. Netw. Technol. (ICCSNT)*, Chang China, 2016, pp. 218–223.

[22] Y. Zhang, X. Yu, L. Xiao, W. Li, Z. Fan, and W. Zhang, "Weights and structure determination of artificial neuronets," in *Self-Organization: Theories and Methods*, New York, NY, USA: Nova Science, 2013.

[23] Y. Zhang, D. Guo, Z. Luo, K. Zhai, and H. Tan, "CP-activated WASD neuronet approach to asian population prediction with abundant experimental verification," *Neurocomputing*, vol. 198, pp. 48–57, Jul. 2016.

[24] Y. Zhang, Y. Yin, D. Guo, X. Yu, and L. Xiao, "Cross-validation based weights and structure determination of chebyshev-polynomial neural networks for pattern classification," *Pattern Recognit.*, vol. 47, no. 10, pp. 3414–3428, Oct. 2014.

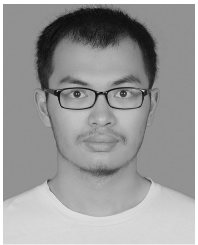
[25] Y. Zhang, Z. Xue, T. Jing, Y. Ling, and C. Ye, "Rough period estimation and peak prediction of stock market based on multiple sine functions extraction," in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, Chengdu, China, Jul. 2018, pp. 311–318.

[26] Y. Zhang, R. Wang, W. Lao, and J. Deng, "Signum-function-activated WASD neuronet and its XOR application," *Acta Scientiarum Naturalium Universitatis Sunyatseni*, vol. 53, no. 1, pp. 1–7, Jan. 2014.

[27] R. Dey and F. M. Salemt, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. IEEE 60th Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Boston, MA, USA, Aug. 2017, pp. 1597–1600.

[28] E. Kabir, S. Guikema, and B. Kane, "Statistical modeling of tree failures during storms," *Rel. Eng. Syst. Saf.*, vol. 177, pp. 68–79, Sep. 2018.

[29] *U.S. Bureau of Labor Statistics*. Accessed: Nov. 29, 2019. [Online]. Available: https://www.bls.gov/



TIANYU ZENG is currently pursuing the B.S. degree in computer science and technology with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. His current research interests include prediction theories and neural networks.

YUNONG ZHANG (Member, IEEE) received the B.S. degree in industrial electrical automation from the Huazhong University of Science and Technology, Wuhan, China, in 1996, the M.S. degree in control theory and control engineering from the South China University of Technology, Guangzhou, China, in 1999, and the Ph.D. degree in mechanical and automation engineering from The Chinese University of Hong Kong, Hong Kong, in 2003.

He is currently a Professor with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China. Before joining Sun Yat-sen University in 2006, he has been with the National University of Singapore, the University of Strathclyde, and the National University of Ireland at Maynooth, since 2003. His main research interests include robotics, neural networks, and computation and optimization. His web-page is now available at http://sdcs.sysu.edu.cn/content/2477.

ZHENYU LI received the B.S. degree in computer science and technology from Sun Yat-sen University, Guangzhou, China, in 2019, where he is currently pursuing the M.S. degree in computer technology. His current research interests include time-varying problems solving, neural networks, and intelligent computing.

BINBIN QIU (Member, IEEE) received the Ph.D. degree in information and communication engineering from Sun Yat-sen University, Guangzhou, China, in 2018. He is currently a Postdoctoral Fellow with the School of Data and Computer Science, Sun Yat-sen University. His main research interests include neural dynamics/neural networks, robotics, numerical computation, and intelligent optimization.

CHENGXU YE received the B.S. degree from Shanxi Normal University, Xi'an, China, in 1991, the M.S. degree from Qinghai Normal University, Xining, China, in 2008, and the Ph.D. degree from Sun Yat-sen University, Guangzhou, China, in 2015. He is currently a Professor with the School of Computer Science, Qinghai Normal University. He has published over 30 scientific articles in journals and conferences. His main research interests include machine learning, neuronets, computation, and optimization.