

Article

# Stacked-GRU Based Power System Transient Stability Assessment Method

Feilai Pan <sup>1</sup>, Jun Li <sup>1</sup>, Bendong Tan <sup>2,\*</sup>, Ciling Zeng <sup>1</sup>, Xinfan Jiang <sup>1</sup>, Li Liu <sup>1</sup> and Jun Yang <sup>2</sup>

<sup>1</sup> State Grid Hunan Electric Power Company, Changsha 410000, China; panfl@hn.sgcc.com.cn (F.P.); lijun@hn.sgcc.com.cn (J.L.); zengcl@hn.sgcc.com.cn (C.Z.); jiangxf@hn.sgcc.com.cn (X.J.); liul56@hn.sgcc.com.cn (L.L.)

<sup>2</sup> School of Electrical Engineering, Wuhan University, Wuhan 430000, China; JYang@whu.edu.cn

\* Correspondence: 2013302540270@whu.edu.cn; Tel.: +86-27-68776346

Received: 24 July 2018; Accepted: 7 August 2018; Published: 9 August 2018



**Abstract:** With the interconnection between large power grids, the issue of security and stability has become increasingly prominent. At present, data-driven power system adaptive transient stability assessment methods have achieved excellent performances by balancing speed and accuracy, but the complicated construction and parameters are difficult to obtain. This paper proposes a stacked-GRU (Gated Recurrent Unit)-based transient stability intelligent assessment method, which builds a stacked-GRU model based on time-dependent parameter sharing and spatial stacking. By using the time series data after power system failure, the offline training is performed to obtain the optimal parameters of stacked-GRU. When the application is online, it is assessed by framework of confidence. Basing on New England power system, the performance of proposed adaptive transient stability assessment method is investigated. Simulation results show that the proposed model realizes reliable and accurate assessment of transient stability and it has the advantages of short assessment time with less complex model structure to leave time for emergency control.

**Keywords:** data-driven; adaptive transient stability assessment; stacked-GRU; time series; intelligent assessment system

## 1. Introduction

With the development of the economy, load demand is increasing, and the power system is getting closer to the transmission capacity limit. In addition, the pace of nationwide networking is gradually advancing, the security and stability characteristics of the power grid are increasingly complex, and the risk to get an unstable power system after a fault is higher and higher [1], so quickly and accurately determining the stability of the power system is an urgent problem to be solved to allow for the safe operation of the power grid.

At present, the time domain simulation [2], energy function method [3,4], machine learning method [5,6], and dynamic security domain [7] are studied in the research of power system transient stability assessment (TSA) methods. The time domain simulation method and the energy function method still have great limitations after many years of development and have not effectively solved the TSA problem; with the breakthrough of machine learning in image processing, speech processing, and other fields [8], TSA based on data-driven methods has begun to attract the attention of many scholars. The use of machine learning methods for TSA is also known as an intelligent system (IS) [9–11], IS can find the relationship between power system features and stability; IS has the advantages of fast assessment speed, strong generalization ability, and ability to estimate dynamic security domains for control decisions compared with the time domain simulation and energy functional methods [9].

In order to achieve rapid TSA, a large number of machine learning algorithms such as the BP neural network [12], support vector machine [13,14], decision tree [15], random forest [16], and K-nearest neighbor [17] are applied in intelligent systems, such algorithms mainly use the key features of the power system after feature selection to establish functional relationship with transient stability. Once the fault is cleared, the transient stability assessment can be performed, but there is no guarantee of 100% accuracy of assessment. Therefore, an adaptive assessment method [9,18] has emerged in recent years. By determining the time window of the fixed length after the fault,  $m$  data sampling points in the time window is used for training; the assessment is performed in chronological order until the assessment confidence meets the requirements. Such approaches guarantee the correctness of assessment, but the assessment time is long and the construction of the model is very cumbersome:  $m$  data sampling point means that  $m$  classifiers need to be trained. Although the authors of a past paper have [18] considered the influence of historical information on the current time assessment in the time series, it reduces the assessment time to a certain extent, but still needs to train a large number of classifiers.

Aiming at solving the problems existing in the above adaptive assessment method, the main contribution of this paper is to propose a stacked-GRU based intelligent TSA model, which realizes the parameter sharing at each moment through stacked-GRU and utilizes the time series data after the power system failure to train with less complex model structure. The offline training is performed to obtain the optimal parameters. When the application is online, it is judged whether the current assessment result satisfies the confidence level at the time of fault clearing, otherwise measurements will continue to be provided to the next moment for assessment until the confidence level is satisfied.

On the one hand, it can perform TSA in the current moment, and on the other hand, can provide historical information for the assessment at the next moment when the assessment result at the current moment is unreliable, so there is no need to train multiple classifiers, and the problem of cumbersome training is solved. In addition, the method draws on the idea of deep learning [19] to use “stacking” to learn representational features from the input can further improve the accuracy of the classifier.

The remainder of the paper is organized as follows. Section 2 introduces the structure of the proposed stacked-GRU. Section 3 proposes the transient stability intelligent assessment method. Simulations and analysis are carried out in Section 4. Finally, a conclusion is drawn in Section 5.

## 2. Methodology

### 2.1. Gated Recurrent Unit

As shown in Figure 1 [20], GRU (Gated Recurrent Unit) [21] is a kind of recurrent neural network (RNN) used to solve the problem of gradient explosion and gradient disappearance [22]. Traditional neural networks such as the BP neural network [23] and convolutional neural networks [19] are not good at processing time series information, and GRU, as a variant of RNN, can combine historical information with current time to predict further information, so it has been widely used in speech recognition, language modeling, translation, picture description, and other issues [24].

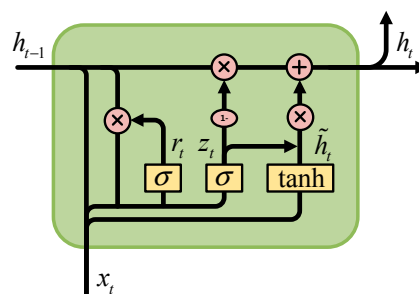


Figure 1. Gated Recurrent Unit.

The forward propagation of the GRU unit is shown in the Equations (1)–(5), where “ $\cdot$ ” is matrix multiplication and “ $\odot$ ” represents matrix wise element multiplication:

The GRU’s update gate  $z_t$  and reset gate  $r_t$  are used to control the direction of the data stream at time  $t$ . The calculation method is as follows:

$$z_t = \sigma(W_z \cdot [h_{t-1}, x_t]) \quad (1)$$

$$r_t = \sigma(W_r \cdot [h_{t-1}, x_t]) \quad (2)$$

where  $x_t$  is the input at time  $t$ ,  $W_z$  is the weight of update gate,  $W_r$  is the weight of reset gate, and  $h_{t-1}$  is the output of hidden layer.

When calculating  $\tilde{h}_t$  which is the output of candidate hidden layer at the time  $t$ , the information of the historical time which means the output of hidden layer is retained, and the historical time information is controlled by adjusting the value of  $r_t$ , as shown by:

$$\tilde{h}_t = \tanh(W \cdot [r_t \odot h_{t-1}, x_t]) \quad (3)$$

where  $\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$ .

Finally, by using  $z_t$  to control the information of the hidden layer that how much is forgotten and the how much information in the candidate hidden layer is retained, the final calculated output of hidden layer is as shown in the equation:

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \tilde{h}_t \quad (4)$$

Then the output of GRU is:

$$\tilde{y}_{last} = \sigma(W_o h_{last} + b_o) \quad (5)$$

where  $\tilde{y}_{last}$  is the output of last output layer in the time series,  $W_o$  is weight of output layer, and  $b_o$  is bias of output layer.

If the reset gate is close to 0, then the output of hidden layer at the moment will not be preserved, and the GRU unit will discard some historical information unrelated to the future. The update gate determines how much information the hidden layer at time  $t - 1$  needs to be saved in the output of hidden layer at time  $t$ . If the element is close to 1, the corresponding information of the hidden layer at the moment will be copied to the time  $t$ . At this time, the update gate unit with long distance dependence is active, therefore long-distance information can be learned; if the element is close to 0, it is equivalent to the standard RNN, and the short-distance information can be processed. At this time, the update gate unit with short-distance dependence is active. Under this mechanism, historical information can flexibly help predict future information.

When training the model, the GRU uses the backpropagation through time (BPTT) [25] for network optimization. The target loss function to optimize is:

$$L = \sum_i^N y^{(i)} \ln(\tilde{y}_{last}^{(i)}) \quad (6)$$

In the formula,  $y^{(i)}$  is the real label of  $i$ -th sample, and  $\tilde{y}_{last}^{(i)}$  is predicted label at the last moment of  $i$ -th sample, and  $N$  is the total number of samples.

## 2.2. Stacked-GRU

GRU is a shallow model with weak capability of feature extraction, and the stacked-GRU is composed of several GRU units, as shown in Figure 2. Specifically, the input of first layer in stacked-GRU is the original data, and the formulas is the same as the GRU unit in Section 2.1.

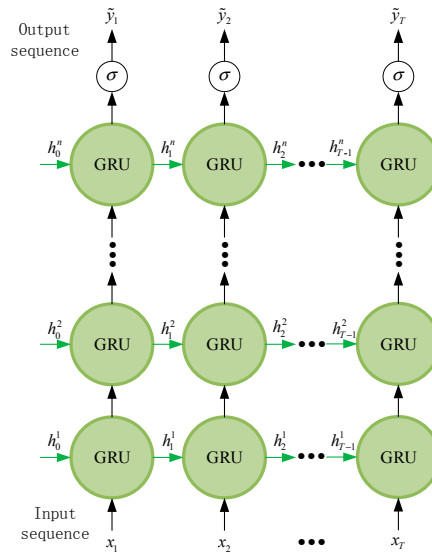


Figure 2. The structure of stacked-GRU.

The input of each GRU unit in the middle is the output of the hidden layer of the upper layer GRU unit:

$$\begin{cases} z_t^i = \sigma\left(W_z^i \cdot [h_{t-1}^i, h_t^{i-1}]\right) \\ r_t^i = \sigma\left(W_r^i \cdot [h_{t-1}^i, h_t^{i-1}]\right) \\ \tilde{h}_t^i = \tanh\left(W^i \cdot [r_t^i \odot h_{t-1}^i, h_t^{i-1}]\right) \\ h_t^i = z_t^i \odot h_{t-1}^i + (1 - z_t^i) \odot h_t^{i-1} \end{cases} \quad (7)$$

Among them, the superscript represents the  $i$ -th GRU unit, and the subscript represents the moment  $t$ . The hidden layer of the last layer in GRU unit adds a layer of sigmoid as the classifier to provide output:

$$\tilde{y}_{last} = \sigma(W_o^n h_o^n + b_o^n) \quad (8)$$

In the formula,  $\tilde{y}_{last}^{(i)}$  is predicted label at the last moment of  $i$ -th sample,  $W_o^n$  is weight of output layer, and  $b_o^n$  is the bias of the  $n$ -th GRU unit. The training method of stacked-GRU and the optimized objective function are the same as a single GRU. The objective function is:

$$L = \sum_i^N y^{(i)} \ln(\tilde{y}_{last}^{(i)}) \quad (9)$$

On the one hand, such a deep structure can efficiently discover high-level feature from limited data. On the other hand, it can more fully utilize information in the time series, which can effectively improve the performance of the classifier. In addition, the stacked-GRU model parameters are independent of time, so the trade-off between time and precision in reference [18] can be avoided.

### 3. Transient Stability Intelligent Assessment Method

#### 3.1. Offline Training

In order to train model effectively, the angle of generators, active power of generators, reactive output of generators, the bus amplitude, and phase, active, and reactive of line, active and reactive of load are selected as input features. The total dimension of the input features is 368. The feature vectors are denoted as  $x_t$ , it represents  $t$ -th sampling point in the time series after fault-clearing, then the input of model can be denoted as  $X = [x_1, x_2, \dots, x_t, \dots, x_T]$ ,  $T$  is the length of timing observation window selected after the fault is cleared.

In offline training, time domain simulation is used to generate a fixed time length fault sequence data set under various operating conditions, and the labels are set as a corresponding stable state, and the stable state is defined as:

$$y = \begin{cases} 1 & \eta < 0 \\ 0 & \eta \geq 0 \end{cases} \quad (10)$$

where  $\eta = \frac{360 - \delta_{max}}{360 + \delta_{max}}$  [18],  $\delta_{max}$  is the maximum value of the rotor angle deviation between any two generators in the power system at the end of the time domain simulation, “1” represents instability and “0” represents stability. The dataset is used as input to train the stacked-GRU model. The training model is shown in Figure 2, and the model optimal parameters are obtained by grid search.

### 3.2. Online Application

In the online application, the confidence framework is used to evaluate the feature data obtained in the power system. Input  $x_t$  at the time  $t$  and specify:

$$assessment\ result = \begin{cases} stable & \tilde{y}_t \subseteq [0, \alpha] \\ unstable & \tilde{y}_t \subseteq [1 - \alpha] \\ unknown & others \end{cases} \quad (11)$$

Among them,  $\tilde{y}_t$  is the output of the time  $t$ . If the intelligent method assessment result is in the confidence interval, then the assessment result is considered to be reliable; otherwise, the input is further evaluated until the assessment time reaches the maximum size of the observation window; if the assessment time exceeds the length of the observation window, the transient instability is considered and emergent control measures are taken. The model performance measurement is defined as the average response time (ART) [9]:

$$ART = \frac{\sum_i^N C(X_i)}{N} \quad (12)$$

where  $C(X_i)$  is the time that sample  $X_i$  is assessed, so the smaller the value of ART, the faster the model is assessed. A schematic diagram of the transient stability intelligent assessment model is shown in Figure 3. It can be found that although the model structure is roughly the same, the model proposed in this paper is more flexible than the model proposed in reference [18] because of the form of parameter sharing at each moment.

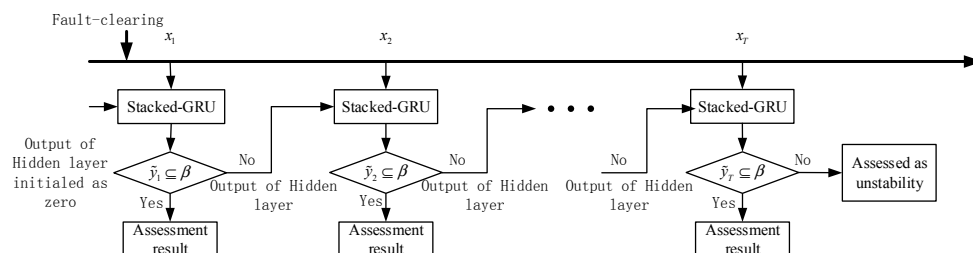


Figure 3. Online application of transient stability intelligent assessment.

## 4. Case Studies

### 4.1. Data Generation

All simulations are implemented on a laptop with Intel® Core(TM) i5-7300HQ CPU@2.50GHz, 8 GB memory and 4 GB NVIDIA GEFORCE GTX 1050 ti GPU, and the intelligent assessment model is built on the deep learning framework Tensorflow1.4 [26].

As shown in Figure 4, a New England power system [27] including 10 generators and 39 buses with frequency of 60 Hz is built in the simulation. The time domain simulation is carried out using the software of PSS/E (Power System Simulator/Engineering, Siemens PTI, Berlin, Germany). The GENCLS models and ZIP models are used to describe the generators and the loads respectively. The simulation step size was set to 0.0083 s. In order to construct a relatively complete sample space, the operating conditions of the power system includes 70%, 75%, 80%, 85%, 90% ..., 115%, and 120% of the benchmark operation. The generation level is changed according to the load level. Power flow calculation is performed in these 11 modes of operation. If the power flow converges, a three-phase ground short circuit fault is set at all bus bars and at 20%, 40%, 60%, and 80% of the line end. The fault clearing time is 0.1 s and 0.3 s, and the simulation duration is 20 s. At the end of the simulation, if rotor angle deviation between the any two generators exceeds  $360^\circ$ , it is determined to be unstable, transient stable and unstable cases are shown in Figure 5a,b, respectively. Finally, a total of 3848 valid samples were obtained, of which 2693 samples were used as training sets and the remaining 1155 samples were used as test sets. The first 20 cycles (40 sampling points) of the training set were input into the model for training and the model was validated with the test set.

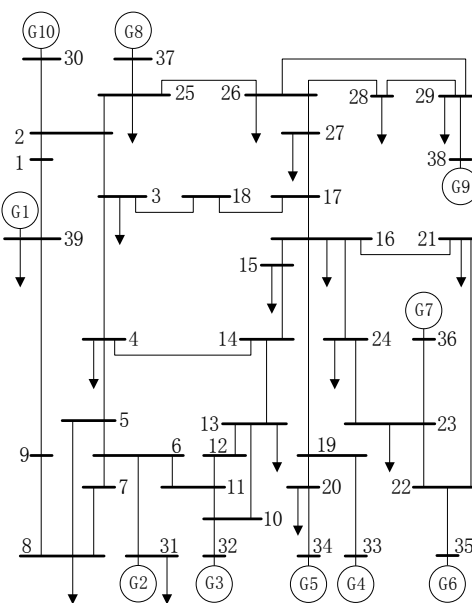


Figure 4. The New England power system.

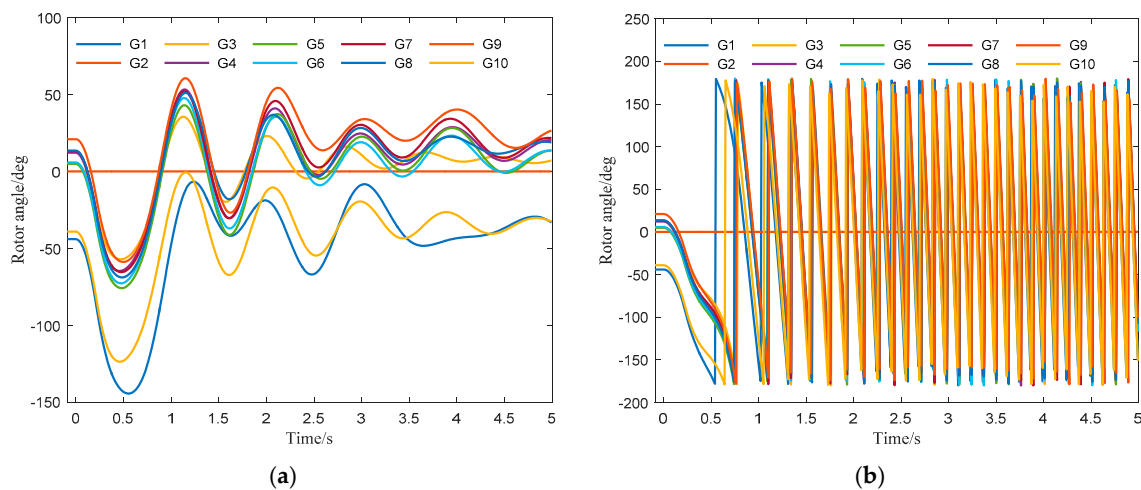
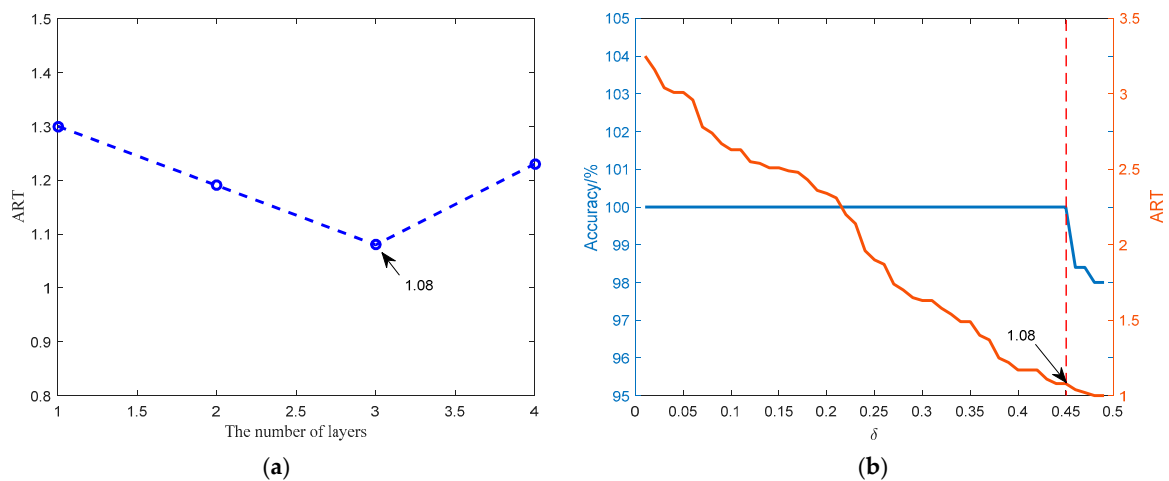


Figure 5. (a) Transient stable case; (b) transient unstable case.

## 4.2. Discussion

### 4.2.1. Different Layers of Stacked-GRU Performance Assessment

According to the deep learning theory [28], the more layers of the model, the stronger the ability to approximate arbitrary functions, but it also causes over-fitting problems at the same time considering that the data in a 10-machine and 39-buses power system. The amount of large data that can be processed compared to deep learning is not huge; in addition, the more model layers, the larger the computational resources required. Therefore, according to the amount of data utilized and the hardware platform, it can be determined that the number of layers required by the model is not very large. In order to determine the number of layers of the model stacked-GRU, the grid search method is used to adjust the parameters. The number of hidden layer neurons in each layer of GRU is set to 100, and the number of model training epochs is 5000, a dropout layer after each layer of GRU [29] is added to avoid over-fitting problems. Generally speaking, the critical stable area of the power system is not large, so the value of  $\alpha$  is also small. But the value of  $\alpha$  is too small to cause misclassification, and the number of uncertain samples will be too large, so a grid search should be implemented to find the best  $\alpha$ . The test results are shown in Figure 6.



**Figure 6.** (a) The relationship between ART and number of layers; (b)  $\delta$  sensitivity analysis.

It can be seen from the Figure 6a that ART obtains the minimum value of 1.08 when the number of layers of stacked-GRU is three, which indicates that the more layers, the faster the assessment of the model. However, too many layers will lead to a decrease in accuracy due to the limitation of the amount of data. Therefore, the number of layers of the model should be selected according to the needs of the system in practical applications. In this paper, the number of stacked-GRU layers is set to three.

$\delta$  sensitivity analysis is shown in Figure 6b when  $\delta$  gets larger, ART become smaller, however, the accuracy gets lower. This is because a larger  $\delta$  means less unknown samples, which leads to higher misjudgment rate. Accuracy can be as high as 100% when  $\delta = 0.45$ , therefore, to implement reliable and quick assessment, the value of  $\delta$  is set to 0.45.

### 4.2.2. Performance Comparison of Different Models

Compared with long-short term memory (LSTM) [20], GRU replaces LSTM's forget gate and input gate with update gate which is equivalent to combining cell state and hidden layer state. Fewer parameters and simpler structure make it easier to train GRU. Experiments show that the effect of GRU is slightly stronger than or equivalent to LSTM on most tasks. The transient stability adaptive assessment method constructed by using 20 LSTM classifiers in reference [18] requires more training

parameters than 10 times the amount of model parameters built by parameter sharing in 3-layers stacked-GRU, so the computational resources of LSTM required are very large. Therefore, the model proposed in this paper adopts stacked-GRU.

The intelligent assessment process of single-layer GRU, 3-layer stacked-GRU, and LSTM is compared in order to further verify that the proposed model has the advantage of high accuracy and lower model complexity. The test conditions of the three models are set to be equal to Section 4.2.1, the assessment results are shown in Figure 7.

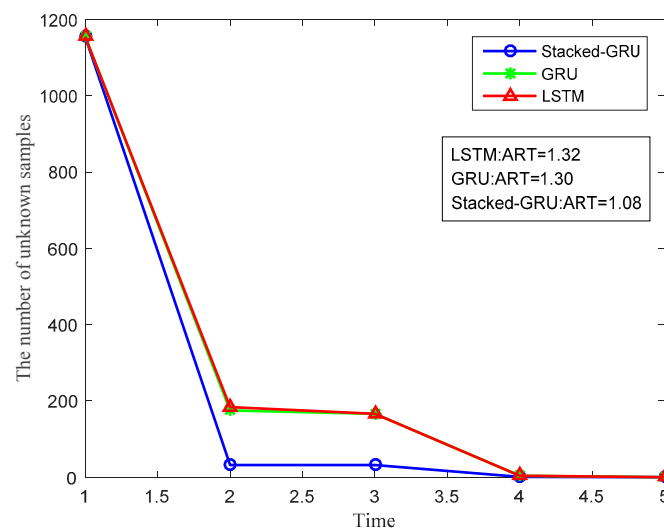


Figure 7. Comparison of different model performance.

As can be seen from the Figure 7 and Table 1, the GRU and LSTM have the same assessment performance. The three models can correctly identify all the test sets in the fourth round of assessment, and the stacked-GRU has reached nearly 100% assessment in the first round of assessment and its assessment result is ART = 1.08, its assessment speed is greatly improved compared to that in a past paper [18], ART = 1.448. Meanwhile, the simulation also shows that stacked-GRU has stronger feature extraction capability than GRU and LSTM which results in achieving shorter assessment time.

Table 1. Assessment process of stacked-GRU.

Time	Number of Unknown Samples	Number of Known Samples	Accuracy
1	1155	1123	100%
2	32	1	100%
3	31	30	100%
4	1	0	100%

## 5. Conclusions

In this paper, a stacked-GRU-based transient stability intelligent assessment method was proposed. By using the parameter sharing method in time, the stacked-GRU classification model was trained by using the transient time series data after the power system failure. An intelligent assessment method was constructed and each sample was evaluated in chronological order using a confidence framework. The simulation results of the New England power system show that some conclusions can be drawn as follows:

(1) With the stacking of several GRU units, stacked-GRU has strong ability to extract higher feature from data to get state-of-art effect.



(2) The proposed model consists of less parameters compared with traditional adaptive assessment methods, it does not need to construct so many classifiers. As a result, the model is easy to train and use.

(3) The proposed model is a good way to do reliable and accurate assessment of transient stability, and it realizes shorter response time to leave enough time for emergency control.

However, the real power system has the problem that the length of the time series data is not completely equal and the instability is unbalanced, which is the focus of our future work.

**Author Contributions:** Conceptualization, F.P. and J.L.; Methodology, B.T.; Validation, C.Z., X.J., and L.L.; Formal Analysis, J.Y.

**Funding:** This research was funded by Hunan Provincial Power Company grant number 5216K0160001.

**Acknowledgments:** This work is supported by the Hunan Provincial Power Company.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

- Zhou, Y.Z.; Wu, J.Z.; Yu, Z.H.; Ji, L.Y.; Hao, L.L. Support vector machine ensemble classifier and its confidence evaluation for transient stability prediction of power systems. *Power Syst. Technol.* **2017**. [[CrossRef](#)]
- Mei, S.W. *Half-Volume Product Method in Power System Transient Analysis*; Tsinghua University Press: Beijing, China, 2010; pp. 142–144.
- Wang, X.F. *Modern Power System Analysis*; Science Press: Beijing, China, 2003; pp. 354–362.
- Xue, Y.; Van Cutsem, T.; Ribbens-Pavella, M. A simple direct method for fast transient stability assessment of large power systems. *IEEE Trans. Power Syst.* **1988**, *3*, 400–412. [[CrossRef](#)]
- Ye, S. Study on Power Systems Transient Stability Assessment Based on Machine Learning Method. Ph.D. Thesis, Southwest Jiaotong University, Chengdu, China, June 2010.
- Liu, L. Power System Transient Stability Assessment Based on Machine Learning. Master's Thesis, North China Electric Power University, Beijing, China, March 2017.
- Yu, Y.X. Review of study on methodology of security regions of power system. *J. Tianjin Univ.* **2008**, *6*, 635–646. [[CrossRef](#)]
- Xi, X.F.; Zhou, G.D. A survey on deep learning for natural language processing. *Acta Autom. Sin.* **2008**, *41*, 635–646.
- Zhang, R.; Xu, Y.; Dong, Z.Y.; Wong, K.P. Post-disturbance transient stability assessment of power systems by a self-adaptive intelligent system. *Gener. Trans. Distrib. IET* **2015**, *9*, 296–305. [[CrossRef](#)]
- Li, Y.; Yang, Z. Application of EOS-ELM with Binary Jaya- Based Feature Selection to Real-Time Transient Stability Assessment Using PMU Data. *IEEE Access* **2017**, *5*, 23092–23101. [[CrossRef](#)]
- James, J.Q.; Lam, A.Y.; Hill, D.J.; Li, V.O. Delay aware intelligent transient stability assessment system. *IEEE Access* **2017**, *5*, 17230–17239. [[CrossRef](#)]
- Yu, J.L.; Guo, Z.Z. Neural-network based transient security assessment for electrical power systems. *Autom. Electr. Power Syst.* **1991**, *15*, 44–50.
- Gomez, F.R.; Rajapakse, A.D.; Annakkage, U.D.; Fernando, I.T. Support vector machine-based algorithm for post-fault transient stability prediction using synchronized measurements. *IEEE Trans. Power Syst.* **2011**, *26*, 1474–1483. [[CrossRef](#)]
- Ai, Y.D.; Chen, L.; Zhang, W.L. Power system transient stability assessment based on multi-support vector machines. *Proc. CSEE* **2016**, *36*, 1173–1180.
- Guo, T.; Milanovic, J.V. Probabilistic framework for assessing the accuracy of data mining tool for online prediction of transient stability. *IEEE Trans. Power Syst.* **2013**, *29*, 377–385. [[CrossRef](#)]
- Zhang, C.; Li, Y.; Yu, Z.; Tian, F. A weighted random forest approach to improve predictive performance for power system transient stability assessment. In Proceedings of the 2016 IEEE PES Asia-Pacific Power & Energy Engineering Conference (APPEEC), Xi'an, China, 25–30 October 2016; pp. 1259–1263. [[CrossRef](#)]
- Liao, R.J.; Wang, Q.; Luo, S.; Liao, Y.X.; Sun, C.X. Condition assessment model for power transformer in service based on fuzzy synthetic evaluation. *Autom. Electr. Power Syst.* **2008**, *32*, 18–21.
- James, J.Q.; Hill, D.J.; Lam, A.Y.; Gu, J.; Li, V.O. Intelligent Time-Adaptive Transient Stability Assessment System. *IEEE Trans. Power Syst.* **2018**, *33*, 1049–1058.

19. Goodfellow, I.; Bengio, Y.; Courville, A.; Bengio, Y. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016; pp. 331–372.
20. Olah, C. Understanding-LSTMs. Available online: <http://colah.github.io/posts/2015-08-Understanding-LSTMs> (accessed on 11 July 2018).
21. Jozefowicz, R.; Zaremba, W.; Sutskever, I. An empirical exploration of recurrent network architectures. In Proceedings of the 32nd International Conference on Machine Learning (ICML 2015), Lille, France, 6–11 July 2015; pp. 2342–2350.
22. Bianchini, M.; Maggini, M. *Handbook on Neural Information Processing*; Jain, L.C., Ed.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 29–65. [[CrossRef](#)]
23. Cheng, Y. The Identification of Fault Type in Transmission Lines Based on Neural Network. Master's Thesis, Dalian University of Technology, Dalian, China, January 2011.
24. Tian, Z. Research on Sentiment Analysis Based on Deep Feature Representation. Master's Thesis, Shandong University, Shandong, China, May 2017.
25. Bersini, H.; Gorrini, V. A simplification of the backpropagation-through-time algorithm for optimal neurocontrol. *IEEE Trans. Neural Netw.* **1997**, *8*, 437–441. [[CrossRef](#)] [[PubMed](#)]
26. Abadi, M.; Barham, P.; Chen, J.; Chen, Z.; Davis, A.; Dean, J.; Devil, M.; Ghemawat, S.; Irving, G.; Isard, M.; et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. Available online: <https://arxiv.org/abs/1606.04467> (accessed on 11 July 2018).
27. Pai, M.A. *Energy Function Analysis for Power System Stability*; Springer Science & Business Media: New York, NY, USA, 2012; pp. 223–227.
28. Grubb, A.; Bagnell, J.A. Stacked training for overfitting avoidance in deep networks. In Proceedings of the ICML 2013 Workshop on Representation Learning, Atlanta, GA, USA, 21 June 2013.
29. Srivastava, N.; Hinton, G.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R. Dropout: A simple way to prevent neural networks from overfitting. *J. Mach. Learn. Res.* **2014**, *15*, 1929–1958.



© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).