World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Research Article)

Check for updates

Disaster impact prediction in the power grid using artificial intelligence based on Texas synthetic grid data replication

Hossein Rahimighazvini *, Milad Hadizadeh Masali, Sahand Saeidi and Reza Barzegaran

Department of Electrical Engineering, Lamar University, Beaumont, TX, USA.

World Journal of Advanced Research and Reviews, 2024, 21(03), 1631–1641

Publication history: Received on 09 February 2024; revised on 16 March 2024; accepted on 19 March 2024

Article DOI: https://doi.org/10.30574/wjarr.2024.21.3.0881

Abstract

Power grids are endangered yearly by disasters that could well affect our livelihood. This paper investigates an AI (Artificial Intelligence) solution to understand better and ultimately predict disasters using the data from previously observed disasters and artificially generated data. The data is fed to the AI in the form of resiliency calculations, which are handled by another sub-program and then go through stages of linear regression models, normal curving models, and decision tree models. Afterward, the main body assigns a value percentile to each of the predictions and reaches a final prediction that is reliable, fast, and accurate. The AI then reviews its decision and checks it with the actual data to see if the forecast is accurate and refines itself according to the newly gathered data. This Algorithm has massive potential implications for decision-making while facing a disaster. It can also help test out different approaches to the problems in a safe environment before the actual procedure initiates, saving time and reducing costs. This paper is an extension to this lab's effort to simulate a real-life disaster and grid via disaster data and multi-layer analysis to predict the steps of disaster and the grid's responses. The raw data (generator's generations, load's amount, lines current, etc.) is refined through the resiliency equations introduced by this lab to be evaluated via the algorithm.

Keywords: Artificial Intelligence; Machine Learning; Natural Disasters; Power Grid; Resilience Analysis; Restoration.

1. Introduction

Numerous critical decisions are made daily based on a multitude of datasets. The precision, speed, and overall effectiveness of the responses given to these numerical problems determine the quality of our life in many aspects. As an example, power grid modernization and advancement over the years are no different. Undeniably, the dependence on electrical power consumption in daily life is growing, resulting in society's incapability to properly function in its absence [1]. Thus, increasing the need to control the factors that might contribute to a partial or total loss of power. One of which can be a catastrophic disaster, especially in the United States, which faces the most significant number of power outages in the modern world [2]. According to the data, there have been 679 separate outage events between the years 2003 to 2012 only [3]. Numerous research has been conducted to evaluate all the visible further and hidden implications such events might have on our systems, communities, and overall livelihood [4]. There is no doubt in the unfathomable profit of solidification against such scenarios. However, to react to a threat, one must first understand its behavior. The closest we can get to understanding the behavior of a disaster is to predict its consequences via analyzing the earlier incidents and the data acquired from it [5]. Obtaining such data will result in better risk assessment, response, and recovery [6]. This study's main objective is to discuss the strategies applicable to predict and consequentially react to an outage caused by weather-related disturbances. The study targets the Texas power grid via the usage of publicly available data in the Texas Synthetic Power Grid. Alongside direct threats such as a power outage caused by major events like hurricane Harvey in 2017, which was approximated to more than 100 billion dollars in total damage [7],

* Corresponding author: Hossein Rahimighazvini

there are indirect damages as seen in cases like hurricane Rita in 2004 and Ike in 2008 resulting in market declination of fuel suppliance, and ultimately noticeable inflation in gas and goods price [8].

The need for a more resilient grid has attracted governmental interest as well as researchers in the past few years due to the reasons mentioned above and more. As evident in the presidential 13653 executive order, attending to subjects regarding the US preparation for extreme weather and climate change impacts [9].

While there is a variety of definitions for what resilience in a power grid is, a void can still be felt when it comes to an accurate depiction of how this translation of resilience behaves and if there are commonalities in the routes that the system partakes in a while facing an outage and trying to recover from it [10]. Resilience can be explained as the system's ability to return to its operational status after facing a disturbance. Resiliency can also be referred to as robustness against challenging circumstances. When the system withstands a challenge, it has previously failed to address, it can also be considered resiliency [11], which is the part this paper mainly focuses on.

The most definitive definition of resilience is given by the Presidential Policy Directive (PPD) 21 [12], stating that resiliency can be considered as "the ability to prepare for and adapt to changing conditions and withstand and recover rapidly from disruptions. Resilience includes the ability to withstand and recover from deliberate attacks, accidents, or naturally occurring threats or incidents."

When discussing a system's ability to recover from a disruptive event while keeping its flexible and timely characteristics, it is crucial to find the optimal recovery strategy grounded in resilience-based methods [13]. In this paper, we are maneuvering over strategies such as normal curve finding and decision trees when it is also discussed in other papers to use strategies like following the Markov chain metrics, which is a replication method more focused on the ease of calculation and resource allocation. Even though the latter reduces the redundancy by 18-45%, it lacks the accuracy offered by the former [14].

Other strategies such as monte Carlo are also discussed in order to determine a resilience factor for the whole network, focusing on reliability engineering that operates based on measuring and ordering components based on their importance [15]. Time complexity is another concern regarding calculating the reliability of a network in an accurate manner. Various solutions have been offered to combat this obstacle, such as fixed-node unconnected subgraphs (FUSA) [16]. While being successful in reducing the needed time to execute the operations, FUSA is more of an algorithmic response to divide and conquer the same objective rather than introducing new methods that require less time to be calculated.

Resilience also constitutes concepts like hardening and self-healing. Hardening is acquiring the ability to withstand harsh impacts that are uncalled for and take place in a brief time, whereas self-healing takes effect once the damage is done and helps the system stay on while under pressure from losses [17].

As stated, the academic society has yet not reached a verdict as to which of the options best describes resilience. It is also worth noting that there is another element in the system called reliability which is closed to resiliency but not entirely. Reliability determines the system's status while it is in a functioning state, whereas resiliency is mostly calculated after the disturbance strikes the system. They are both critical to the power grid and dictate the grid's performance [18].

The restoration phase of the power system in an outage should be fast, accurate, and reliable, ensuring that each tweak is to its maximum [19]. In order to evaluate the options available while restoring the system, it is critical to have an accurate and reliable virtual model of the system capable of predicting its reaction precisely [20].

Giving credit to all pieces of literature regarding the subject in hand that we just had a quick review on, it is apparent that while the nature of resilience has been discussed many times, an accurate depiction of the system's behavior in the event of a disturbance that can reflect the system's reaction to each decision made is needed. The contributions of this study are as follows:

Implementing a data generating code that receives actual data from the Texas Synthetic Power Grid and creates new sets of data that are similar in nature to the actual data but are different in momentarily value for the purpose of training the AI

Creating a prediction AI that guesses the next step value of resilience in a given system based on the previous data on the same scenario and other scenarios via different techniques. Namely, the Regression model, the normal curve model, and the decision tree model.

The AI reaches a final prediction by weighting different calculations in different situations suiting the scenario.

The final prediction is then compared with the actual number from the next step to see if the prediction was close enough or not. If not, the AI trains itself accordingly to prevent the reoccurrence of this mistake over time

This algorithm is the main body of the next project in this lab, which consists of a fully functioning simulation of the disaster and the grid and their most likely responses to each measure introduced through the user.

1.1. Problem Statement

As is the case with many studies regarding AI development, the available sets of data on disasters, while improved vastly during recent years, are not quite sufficient to train a multi-level decision-making algorithm [21]. The subject that this algorithm is trying to understand is a phenomenon relatively uncommon, even though resulting in a magnitude of consequences. A realistic and reliable secondary source of data is needed that has similar behavioral properties to an actual disaster but has multiplication abilities alongside [22]. Having such code in hand allows us to train the algorithm better, putting it through many stages of disaster without the need to actually experience these disasters, which would further help out cause to proceed in having a better analysis of what we are to expect at the time of a disaster.

To determine the current stability and reliability of a system, it is critical to establish a unit based on which it is possible to evaluate the power grid and sort out the constructive steps necessary to be taken. This, in return, gives us the ability to respond effectively to the events that are disturbing the power grid.

The equation that was chosen for this unit, named “resilience”, is shown below:

$$R_r{}^X(t) = \begin{cases} k_1 \dfrac{Q_{1,r}^X(t)-Q_{1,r}^X(t_1)}{Q_{1,r}^X(t_0)-Q_{1,r}^X(t_1)} + k_2 \dfrac{Q_{2,r}^X(t)-Q_{2,r}^X(t_1)}{Q_{2,r}^X(t_0)-Q_{2,r}^X(t_1)} \end{cases} \quad (1)$$

$$t \le t_1, \qquad t_1 < t \le t_2, \qquad t > t_2$$

This equation demonstrates a definitive value based on the difference between the nominal state $Q_1^X(t_0)$, post-disaster state $Q_1^X(t_1)$, and current state $Q_1^X(t)$, as well as the numbers of parts in total $Q_2^X(t_0)$, the post-disaster available parts $Q_2^X(t_1)$, and the currently available parts $Q_2^X(t)$. The options are weighted to account for the sudden fluctuations that might occur in the values that are experienced (K1 and k2 represent weights dictated by the distribution of the power, determining whether the grid is in need of new elements introduction or mostly suffering from a lack of coherence in distribution).

The equation can be used for all the elements available on the grid, where the Q would be filled by:

$$Q_{1,r}^G(t) = \sum_{i=1}^{N_B} P_{G,i,r}(t), \quad Q_{2,r}^G(t) = N_{G,r}^{UP} \qquad (2)$$

$$Q_{1,r}^D(t) = \sum_{i=1}^{N_B} P_{D,i,r}(t), \quad Q_{2,r}^D(t) = N_{D,r}^{UP} \qquad (3)$$

$$Q_{1,r}^L(t) = \sum_{j=1}^{N_L} P_{L,j,r}(t), \quad Q_{2,r}^L(t) = N_{L,r}^{UP} \qquad (4)$$

$$Q_{1,r}^S(t) = \sum_{k=1}^{N_S} P_{S,k,r}(t), \quad Q_{2,r}^S(t) = N_{S,r}^{UP} \qquad (5)$$

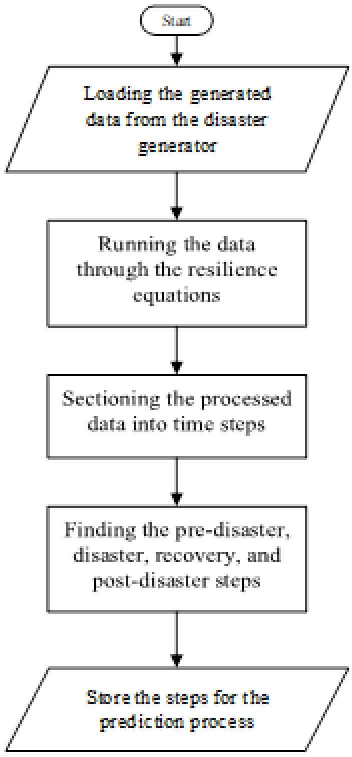

Figure 1 The raw data coming from the disaster generator gets processed into meaningful data by applying the resistance equations. The data then gets separated into four different time periods of pre-disaster (nominal state), disaster, recovery, and post-disaster. The steps are code translations of the same process for the purpose of being used in a binary environment.

Now that a good definition of resiliency is available, it is possible to address that there are inaccuracies in trailing the behavior of the disaster, especially in the event of an extreme change in values, also known as a leap. The path of the disaster might not always be linear, and the excessive number of variables makes it hard to illustrate the resiliency in one equation. The algorithm's objective is to predict the value of resiliency the grid is experiencing in the next step point based on the current and previous data points with minimal inaccuracies. Knowing the benefit of the decision made before it is executed could significantly help us determine our next step.

2. Disaster Generation

Due to the lack of reliable sets of data sources, a disaster generator program was coded. The objective in mind while coding the program was to create a data simulator that changes the overall environment with respect to the original peaks, mean, and overall value of the disaster [23] so that it is a different set of data that still represents similar nature of the original disaster. Generating simulated data is one of the most trending ways used to train an AI algorithm working on a problem with insufficient data in hand. This program reads the available data from disaster cases on the Texas synthetic grid and creates semi-similar cases with nearly infinite possibilities. The process consisted of analyzing the disaster data, randomizing the steps between the pre-disaster and post-recovery steps with randomly generated seeds, and then fitting the newly generated steps with the normal path of the data we had in hand. This would ensure that the data is different but reacts, in the same manner, as the actual disaster data set would. Considering X_i as a random variable, $f(X_1)$ would be the marginal distribution corresponding column in D. The first synthetic data column generated would be:

$$(x_1^1, \ldots, x_1^D)^T. \quad (6)$$

The condition following the lines afterward would follow as:

$$f(X_2|X_1) \quad (7)$$

$$f\,(X_j|X_1, \ldots, X_{j-1}) \quad (8)$$

To determine whether the resulted data is distinguishable from the original data set, the propensity is introduced, whereas:

$$pMSE = N^{-1} \sum (p_i - 0.5)^2 \quad (9)$$

The propensity score varies between 0 and 0.25, with 0 indicating no distinguishability between the two datasets. This can happen if the generator overfits the original dataset and creates a synthetic that is indistinguishable from the original [24].

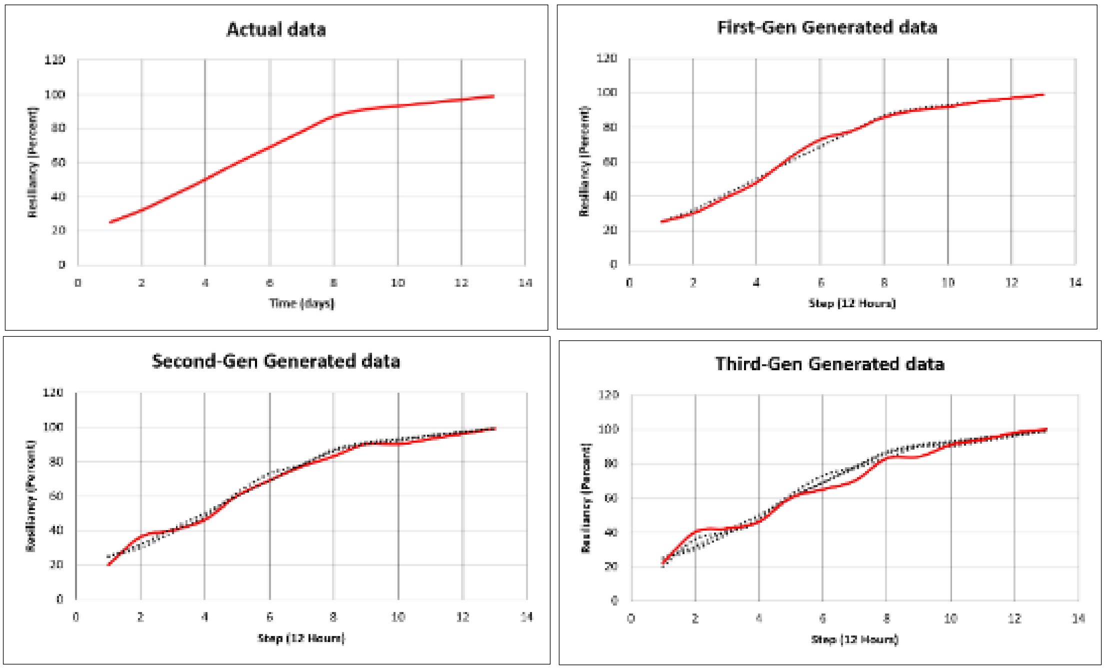

Figure 2 The different generations of generated data and how they start diverting further away from the original data while keeping key Features like the general direction of the data as the data grows more sophisticated, challenging the algorithm's ability to predict sudden Movement. Each read line indicates the current set of data and its change flow where the grad dots are the previous versions for comparison.

The figures above are samples of generated data collected from the code discussed above. As the algorithm responsible for predicting the outcome grows, the data-generating code takes the liberty of increasing the temperature of the data [25]. Temperature is a term in coding that is used to describe the element of randomness and overall radical changes in a point or set of data while creating an output; hence higher temperatures result in more considerable shifts in the path taken by the data, which in terms makes it harder for the algorithm to predict the next step.

3. Multi-Level refined AI resiliency calculator

As mentioned before, while giving a good background on where the point is expected to land, a linear approach is incapable of fully predicting a sudden shift in data. Furthermore, when a linear module sees a significant change, the indicator misses the next mark post the sudden event since it expects the explosion to follow a linear path, ultimately leaving us with at least two inaccuracies during the prediction phase.

Counter-measuring this occurrence, the prediction algorithm consists of the original resiliency equation developed by this lab and introduced above as a measuring base while adding multiple steps to reach less than 1% inaccuracy with regard to keeping the short reaction time the original equation already had.

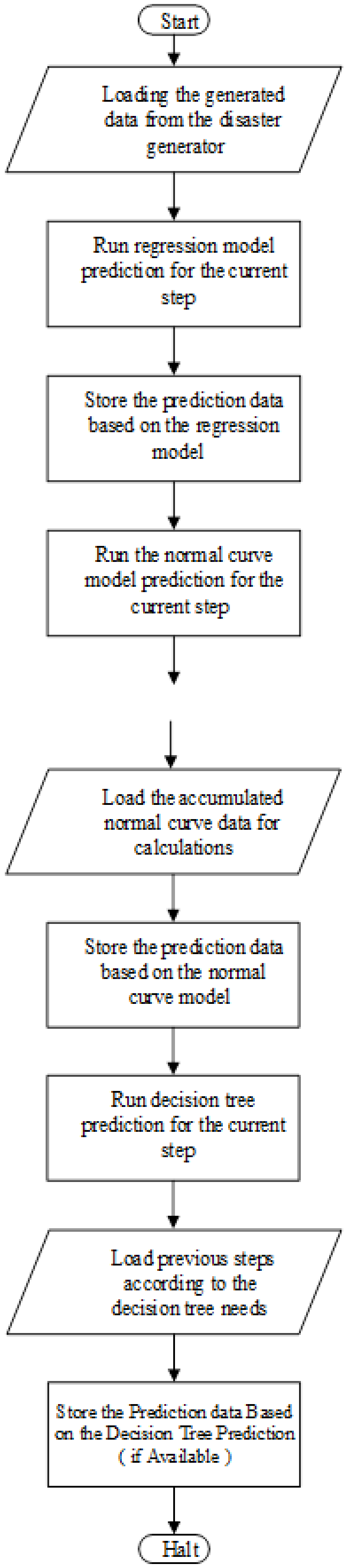

Figure 3 Each unit casts a prediction for the next step (J + 1) based on the current step (J) and the previously accumulated data (the models). The scenario value will later evaluate these predictions to reach a final prediction.

3.1. Normal Curve

A normal curve is the most likely path the disaster is to take. By knowing the normal curve, we have the opportunity to learn the average result that is expected in each scenario. This value is needed to calculate the tolerance experienced from it to reach the final prediction [26]. The process used to get this value consisted of brute calculating all the points in the generated scenarios and averaging them to a single case. While this method is incapable of predicting any sudden motion, it is a reliable reference to determine the general direction of the disaster [27].

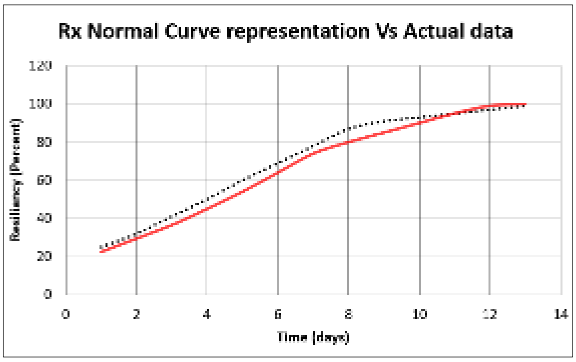

Figure 4 The actual data alongside the normal curve representation of it where the red line is the prediction and the gray line is the actual data.

This is the algorithm's normal curve result after ten thousand practices. While the amount of training could be tremendously increased using modernized computers, it is evident from the results that, as expected, the normal curve appears to be insensitive to the sharper edges of the data, only outlaying the closest path that could be taken by it.

3.2. Decision Tree

Decision trees are a case to case and step-to-step profiles that try to explain each new event based on the previous occurrence. To make this prediction accurate, we used the second and third-step decision trees, meaning the result is predicted based on the two or three last available steps. Needless to say, it is all possible when the requirements in the number of steps are met. These profiling trees need an immense amount of data set to function properly and accurately, fed by the data generator.

The decision is of binary nature, whereas each $i = k \in [1, K]$ constitutes the leaf nodes of the tree. Each $n_i = w_k$ is a row vector from the fully connected layer's weights

$$W \in R^{D\times K}. \qquad (10)$$

The average weight of leaves in node I, when chosen from the prior sample portion, can be calculated by the term:

$$n_i = \sum_{k\in L(i)} w_k/|L(i)| \qquad (11)$$

From there, to evaluate the probability of an event j reoccurring in node i of sample x would be:

$$p(j|i) = SOFTMAX(<n_i , x_i>)[j] \qquad (12)$$

Where:

$$n_i = (<n_j , x>)_{j\in C(i)} \qquad (13)$$

using the path probability to choose a leaf, The probability of each node $i \in P_k$ traversing the next node in the path $C_k(i) \in P_k \cap C(i)$ is denoted $p(C_k(i)|i)$. Then, the probability of leaf and its class k is [28]:

$$p(k) = \Pi_{i\in P_k}\, p(C_k(i)|i) \qquad (14)$$

Same as the normal curve, the prediction was made on the actual case after a ten thousand test run on generated data. As evident by the pictures, while still relatively young, the algorithm behind the decision trees is able to find and navigate the path of the actual data. The overshoots and undershoots are expected to decrease after the program ages and gain a deeper insight into generated data sets over time [29-30].

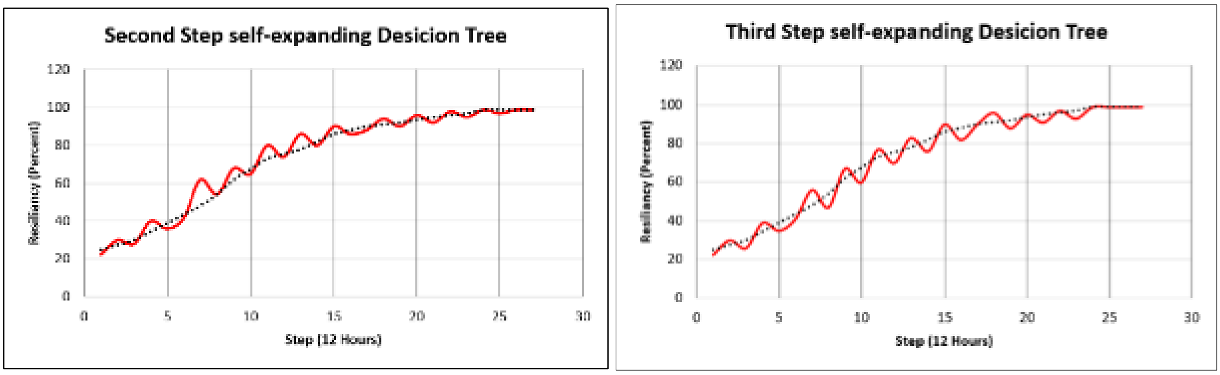

Figure 5 The second and third-step decision trees decisions per step where the red line is the prediction and the gray line is the actual data.

3.3. Scenario Value

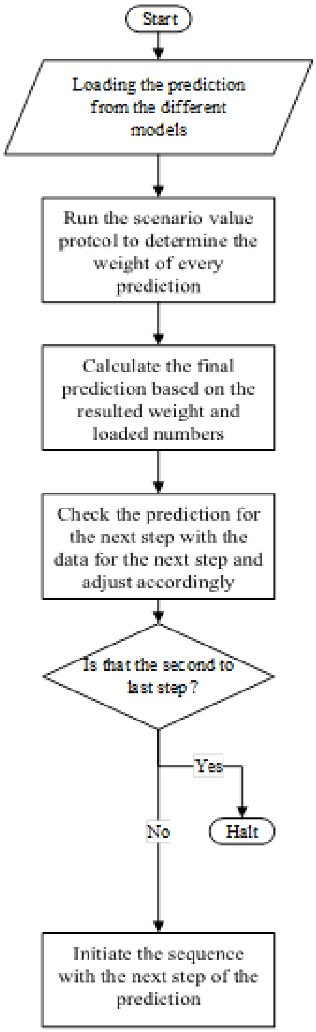

Figure 6 The scenario evaluator makes a prediction based on the outcome of all the procedures. The prediction would then be compared with the actual numbers from the next step that is stored in the data to determine and improve the prediction accuracy.

Having four different values from four different equations in hand, namely the gaussian regression, the normal curve, the second step decision tree, and the third step decision tree, we can accurately predict the path of the disaster while anticipating any sudden changes with a substantially more acceptable rate of failure. Since the result from each set of calculations are not always equally valid, there is a need for an evaluating mechanism to make a case-to-case point for each scenario and credit each response as needed to get the best output.

$SV(n)=k_1GR(n)+k_2NC(n)+k_3SD(n)+kTD(n)$ (15)

The code behind scenario value is a complex AI algorithm that decides the weight of all answers from four explained equations based on the relativeness of each equation to the scenario at hand. The relativeness is calculated via the previous experiences that the code had with similar scenarios and multi-level profiling.

Nomenclature

- $N_G^{UP}(t)$: Number of online generators at time *t.*
- $N_D^{UP}(t)$: Number of connected loads at time *t.*
- $N_L^{UP}(t)$: Number of available transmission/distribution lines at time *t.*
- $N_S^{UP}(t)$: Number of working substations at time *t.*
- N_B : The total number of buses.
- N_L : The total number of transmission/distribution lines.
- N_S : The total number of substations.
- $P_{G,i}(t)$: Power generation of bus *i* at time *t*.
- $P_{D,i}(t)$: Power demanded of bus *i* at time *t*.
- $P_{L,j}(t)$: Transferred power of line *j* at time *t*.
- $P_{S,k}(t)$: Output power of substation k at time *t*.
- $Q_{1X}(t_0)$: Nominal state.
- $Q_{1X}(t_1)$: Post-disaster state.
- $Q_{1X}(t)$: Current state.
- $Q_{2X}(t_0)$: The number of parts in total.
- $Q_{2X}(t_1)$: Post-disaster available parts.
- $Q_{2X}(t)$: Currently available parts.
- *SV(n)*: Scenario value at stage n.
- *GR(n)*: Gaussian regression model value at stage n.
- *NC(n)*: Normal curve model value at stage n.
- *SD(n)*: Second-step decision tree value at stage n.
- *TD(n)*: Third-step decision tree value at stage n.

4. Conclusion

The calculation methods discussed in this paper provide an accurate digital representation of the disaster, its behavior, and its consequences which could be well used to investigate further the best options available to react to the disaster in a timely manner since each option could be weighted and fully considered before putting into action. The discussed algorithm successfully predicted the next step in the unfolding of the sample case disasters up to a small fault via using different approaches, namely gaussian regression, normal curves, and decision trees that has not been achieved prior to this paper. This technology can shorten the downtime of the power systems and cut the losses drastically if engaged with the right strategies for recovery.

Disaster generation code provides a feed of real-like disaster data that can be used to test multiple scenarios and their effects. At the same time, the prediction algorithm ensures a mixture of speed, with the normal curve, accuracy with the decision tree, and the scenario value method to balance it all to the direction we desire.

Using such advancement, a responsive modeling system will be far more feasible, helping at improving the short- and long-term strategies regarding responses to disasters that in most cases need fast and decisive decision making since they're in close contact with the life of thousands of people.

Compliance with ethical standards

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Susan L. Cutter, “The landscape of disaster resilience indicators in the USA”, *Natural Hazards 80*, pp. 741-758, 2015.

[2] Tara Dodrill, “Off the Grid News”, August 1, 2014.

[3] Adam B. Smith, “US billion-dollar weather and climate disasters: data sources, trends, accuracy and biases”, Natural Hazards, vol. 67, pp. 387-410, 2013.

[4] Behzad Rouhanizadeh, “Exploratory analysis of barriers to effective post-disaster recovery”, International Journal of Disaster Risk Reduction, vol. 50, 2019.

[5] Abdelsalam, Hany A., Abdelrahman Eldosouky, and Anurag K. Srivastava. "Enhancing distribution system resiliency with microgrids formation using weighted average consensus." *International Journal of Electrical Power & Energy Systems*, vol. 141, 2022.

[6] M. Sakurai “Information technologies and disaster management - Benefits and issues -“, Progress in Disaster Science, vol. 2, 2019.

[7] Eric S. Blake, “Hurricane Harvey,” Tropical Cyclone Report National Hurricane Center, January 2018.

[8] Margaret Saizan, “A Visual Story of Hurricane Rita,” New York Review, September 2005.

[9] Presidential Executive Order, “Preparing the United States for the Impacts of Climate Change,” 2013

[10] Shahbazi, A., Aghaei, J., Pirouzi, S., Niknam, T., Vahidinasab, V., Shafie-khah, M., & Catalão, J. P. ,“ Holistic approach to resilient electrical energy distribution network planning”, International Journal of Electrical Power & Energy Systems, vol. 132, 2021

[11] David D Woods. "Four concepts for resilience and the implications for the future of resilience engineering." Reliability Engineering & System Safety, vol. 141, pp.5-9, 2015.

[12] The White House, “Presidential Policy Directive Critical Infrastructure Security and Resilience”, February 2013.

[13] Tao Liu, “Modeling and evaluation method for resilience analysis of multi-state networks”, Reliability Engineering & System Safety, vol. 226, October 2022.

[14] Maryam NooraeiAbadeh, “Resiliency-aware analysis of complex IoT process chains”, Computer Communications, vol. 192, pp. 245-255, August 2022.

[15] Kash Barker, “Resilience-based network component importance measures,” Reliability Engineering & System Safety, Vol. 117, pp. 89-97, 2013.

[16] Hongjun Cui, “A novel fixed-node unconnected subgraph method for calculating the reliability of binary-state networks”, Reliability Engineering & System Safety, vol. 226, October 2022

[17] S. S. M. Venkata, "Grid Resilience: Elasticity Is Needed When Facing Catastrophes [Guest Editorial]", IEEE Power and Energy Magazine, vol. 13, no. 3, pp. 16-23, June 2015.

[18] Z. Li, "Networked Microgrids for Enhancing the Power System Resilience", Proceedings of the IEEE, vol. 105, no. 7, pp. 1289-1310, July 2017.

[19] National Academies of Sciences, Engineering, and Medicine. 2017. Enhancing the Resilience of the Nation's Electricity System. Washington, DC: The National

[20] Lin, C., Chen, C., Liu, F., Li, G., & Bie, Z. “ Dynamic MGs-based load restoration for resilient urban power distribution systems considering intermittent RESs and droop control”, International Journal of Electrical Power & Energy Systems, vol. 140, 2022.

[21] Y. Pan, "Heading toward artificial intelligence 2.0", *Engineering*, vol. 2, no. 4, pp. 409-413, 2016.

[22] Manzhu Yu, “Big Data in Natural Disaster Management: A Review”, geosciences 8(5), pp. 165-181, 2018.

[23] Anand Narasimhamurthy, "A FRAMEWORK FOR GENERATING DATA TO SIMULATE CHANGING ENVIRONMENTS ", Artificial intelligence and applications, pp.415-420, 2007.

[24] Wenjuan Sun, "Applications of artificial intelligence for disaster management", Natural Hazards volume 103, 2020.

[25] Fida K. Dankar, "Fake It Till You Make It: Guidelines for Effective Synthetic Data Generation", Applied Science 11(5), 2021.

[26] Aaron Jaech, "Real-Time Prediction of the Duration of Distribution System Outages", in IEEE Transactions on Power Systems, vol. 34, no. 1, pp. 773-781, 2019.

[27] Vladimir Sedlacek, "A Formula for Disaster: A Unified Approach to Elliptic Curve Special-Point-Based Attacks", Advances in Cryptology – ASIACRYPT, pp 130-159, 2021.

[28] Alvin Wan, "NEURAL-BACKED DECISION TREE", arXiv, 2020.

[29] S. B. Kotsiantis, "Decision trees: a recent overview", Artificial Intelligence Review, Vol. 39, pp. 261-283, 2013.

[30] Reza Khorramabadi, "Post-buckling and vibration analysis of double-curved sandwich panels with SMA embedded faces", Composites Part C: Open Access, Vol. 12, pp. 100419, 2023.