Advancing Distribution System Restoration via an Innovative Physics-Informed Decision Transformer

Hong Zhao[1], **Jin Wei-Kocsis**[1], **Adel Heidari Akhijahani**[2], **Karen L Butler-Purry**[2],

[1]Purdue University, West Lafayette
[2]Texas A&M University, College Station
{zhao1211,kocsis0}@purdue.edu, adelheidari@exchange.tamu.edu, klbutler@tamu.edu

Abstract

Driven by advancements in sensing and computing, deep reinforcement learning (DRL)-based methods have demonstrated significant potential in effectively tackling distribution system restoration (DSR) challenges under uncertain operational scenarios. However, DRL's data-intensive nature poses obstacles to achieving satisfactory DSR solutions for large-scale and complex distribution systems. Inspired by the promising potential of emerging causal transformers, which widely work as foundation models for large language models (LLMs) such as GPT-x, this paper explores an innovative approach harnessing causal transformers' powerful computing capabilities to address scalability challenges inherent in conventional DRL methods for solving DSR. To our knowledge, this study represents the first exploration of foundation models in revolutionizing conventional DRL applications in power system operations. Our contributions are twofold: 1) introducing a novel Physics-Informed Decision Transformer (PIDT) framework that exploits a GPT-based causal transformer to transform conventional DRL methods for DSR operations, and 2) conducting comparative studies to assess the performance of the proposed PIDT framework at its initial development stage for solving DSR problems. While our primary focus in this paper is on DSR operations, the proposed PIDT framework can be generalized to optimize sequential decision-making across various power system operations.

Introduction

In recent years, reliance on a stable and continuous power supply has reached unprecedented levels across all industry and daily life sectors. Consequently, a reliable and resilient power supply has become paramount for critical infrastructures such as hospitals, transportation systems, communication networks, and manufacturing facilities. At the same time, the increasing frequency and severity of extreme events, such as extreme weather, natural disasters, and cyber-attacks, pose critical challenges for maintaining the resilience and reliability of modern power systems. For instance, in July 2024, Hurricane Beryl caused over 2.7 million households and businesses in Houston, Texas, to suffer from prolonged power outages amid high heat and humidity, resulting in at least 42 deaths and over $6 billion in property damage. These challenges highlight the urgent need to enhance the resilience of power grids to ensure they can withstand and recover from severe disruptions. Given that the distribution system serves as a crucial link in the power delivery process from the transmission grid to end-users, research into resiliency in distribution systems is pivotal for enhancing grid resilience and mitigating the impacts of power outages.

One key indicator of resiliency in distribution systems is the ability to restore service to critical loads after disruptions on the main grid (Chen, Wang, and Ton 2017). Following power outages, DSR aims to rapidly restore affected loads by leveraging advanced emerging controls once the outages are isolated. Forming microgrids (MGs) with dynamic boundaries as a service restoration strategy is a promising solution for enabling effective DSR (Liang et al. 2022; Igder, Liang, and Mitolo 2022; Zhao and Wang 2022). By incorporating various energy sources, distributed generators (DGs), along with remotely controlled switches, a distribution system can be partitioned into multiple self-sufficient MGs. This approach enhances the system's restoration capability and maintains power supply continuity to critical loads, thereby significantly improving overall grid resilience. In this paper, we focus on advancing the DSR solution through the sequential formation of MGs with dynamic boundaries, for enhancing the resiliency of distribution systems.

The existing methods for solving DSR problems can be generally grouped into four categories: mathematical programming methods, heuristic methods, expert systems, and machine learning methods. Mathematical programming methods, such as mixed-integer programming (MIP)-based methods, formulate the DSR problems as mixed-integer linear or non-linear programming problems and solve them using off-the-shelf solvers (Wang and Wang 2015; Patsakis et al. 2019; Yang et al. 2019; Bassey and Butler-Purry 2020; Chen et al. 2019). While these methods ensure the optimality of DSR solutions, they typically require accurate physical models of the distribution systems, which are not always available in dynamic and uncertain operation scenarios. Additionally, the computational complexity can increase dramatically with the number of controllable components, which limits the scalability of the solutions. Heuristic methods leverage algorithms, such as genetic algorithms, Tabu search, and greedy algorithms, to search for satisfactory solutions (Arefifar, Mohamed, and El-Fouly 2012; Sedzro et al. 2019; Wang et al. 2023). Compared to mathematical programming, heuristic methods are capable of handling dynamic and uncertain operation environments. However, these methods also face scalability issues for large-scale distribution systems. Additionally, the quality of the solutions can be sensitive to the hyperparameters of heuristic algorithms. Expert systems use knowledge-based techniques, such as rule-based systems and fuzzy logic systems, to achieve DSR solutions (Chen, Tsai, and Kuo 2005; Pao-La-Or et al. 2009; Singh, Mehfuz, and Kumar 2016). These methods are effective in providing consists and quick decision makings based on encoded knowledge. However, capturing, encoding, and updating expert knowledge base can be challenging. Machine learning methods, especially emerging DRL-based methods, have recently gained attention for enabling more efficient, adaptive, and robust DSR solutions in uncertain operation scenarios. These methods formulate decision making for DSR under uncertainties as a Markov decision process (MDP) or a partially observable Markov decision process (POMDP) and solved them iteratively using data-driven DRL techniques such as deep Q-learning, advantage actor critic (A2C) algorithms, and proximal policy optimization (PPO) algorithms (Wu et al. 2019; Yao et al. 2020; Gao et al. 2020; Du and Wu 2022). While DRL-based methods have shown great potential in efficiently addressing DSR in uncertain operation scenarios, their data-intensive nature poses challenges in achieving satisfactory DSR solutions for large-scale complex distribution systems. Various research efforts have been conducted to tackle the scalability issues, with one main trend being formulating the distribution systems as multi-agent system and then developing multi-agent DRL methods for DSR (Zhao and Wang 2022; Yao et al. 2023; Al-Hinai and Alhelou 2021). While these methods have demonstrated efficiency in addressing DSR in large-scale distribution systems, the coordination between the agents can introduce high communication overhead and operation complexity. Additionally, as indicated in Canese et al. and Du and Ding, achieving stable learning and convergence can be challenging due to the non-stationary nature of the multi-agent system.

Recently, we have observed the great potential of causal transformers that widely work as foundation models of emerging LLMs, revolutionizing various application domains including virtual assistants, healthcare, and education (Wei et al. 2024; Thirunavukarasu et al. 2023; Kasneci et al. 2023). In these applications, causal transformers have demonstrated their advanced capabilities, such as context awareness on long-term dependencies, generative sequence modeling, and large-scale high-dimensional data processing. Inspired by the immense potential of these models, we pose a research question (RQ): Can we leverage the powerful computing capabilities of causal transformers to address the previously discussed scalability issue of DRLs for solving DSR? The work presented in Chen et al. 2021 implies potential direction on addressing this RQ. In Chen et al. 2021, a causal transformer-powered concept, Decision Transformer, was proposed to first transform conventional DRL by modeling it as a conditional sequence modeling problem, and then leverage a causally masked transformer originally developed for LLMs, such as GPT-x model, to generate optimal actions for the DRL by conditioning on desired returns, past states, and actions. Given the powerful computing capabilities of causal transformers in context awareness for long-term dependencies, generative sequence modeling, and large-scale high-dimensional data processing, it is reasonable to investigate the concept of a Decision Transformer for addressing our RQ and explore a transformative computing solution for solving the DSR problem. However, to the best of our knowledge, no existing work has explored the capability of the Decision Transformer in any power system operations, including DSR. This gap may be due to the complex inherent physical constraints present in power systems.

To address this gap, in this paper, we aim to develop a novel Physics-Informed Decision Transformer (PIDT) for solving the DSR problem. As far as we know, this is the first paper that explores the capability of foundation models in revolutionizing conventional DRLs for power system operations. While our focus is on the DSR problem, the proposed PIDT framework can be generalized to optimize sequential decision-making for other power system operations. The main contributions of our proposed work are twofold:

- A novel PIDT framework is proposed as the first-ever effort to explore the powerful computing capabilities of causal transformers, which work as foundation models for LLMs, in transforming conventional DRL for DSR operations.
- Comparative studies are conducted to analyze the performance of the proposed causal transformer-powered PIDT framework in its initial development stage for solving the DSR problem.

The next section illustrates the problem settings for our work. After this, we describe our proposed causal transformer-powered PIDT framework for the DSR problem. The following section shows the case studies and performance evaluations of the proposed PIDT framework. Conclusions are presented in the last section.

Problem Settings

In our work, we formulate the DSR problem as an MDP within DRL framework. Table 1 shows the definitions of the parameters and variables that will be used in this paper.

DSR Problem Modeling

In the initial stage of this research, the proposed DSR method is modeled as a sequential decision-making process involving sequences of control actions on switches to restore loads to their normal operational states. These sequences of switching actions, referred to as energization paths, are each associated with a single active distributed generator (DG). Specifically, each energization path begins at the switch connected to the associated DG and aims to maximize load restoration by forming multiple microgrids, while ensuring compliance with all operational and physical constraints. Additionally, to reduce the space of control actions in the modeling, we adopt the concept of node cell as introduced in Chen et al. 2019. A node cell is defined as a set of nodes that

$s_t, \mathcal{S}$	State at time t, the state space
$a_t, \mathcal{A}$	Action at time t, the action space
R	Reward
$\mathcal{P}$	Transition probability
r_t	Reward at time t
γ	Discount factor
T	Time Horizon
L	Set of all the loads
N	Set of all nodes
C	Set of all node cells
$x_{i,t}^L, x_{i,t}^N, x_{i,t}^C$	Energization status of Load i at time t, energization status of Node i at time t, energization status of node cell i at time t
$P_{l,t}^L, P_{l,t}^C, P_{l,t}$	Active power of Load l at time t when restored, accumulative active power of all loads in node cell l at time t when restored, nominal active power of Load l at time t.
$H_{i,t}^L$	Squared voltage magnitude of Load i at time t
$H_i^{\min}, H_i^{\max}$	Minimum and maximum squared nodal voltage of Load i
$V_{l,t}^{L,p}$	Voltage penalty function of Load l at time t
π	Policy function
θ	Learnable parameter for the policy function (or the proposed PIDT)

Table 1: Definition of variables and parameters

are interconnected directly by non-switchable lines. Consequently, all the lines and loads within a node cell will be energized simultaneously. Furthermore, in our current work, the constraints include: 1) the voltage limits, 2) power flow constraints, 3) DGs' generation capacities, and 4) topological constraints including prevention of loop formation and ensuring no node cell is visited more than once.

DSR Problem Formulation within a DRL Framework

We further formulate the sequential decision-making process for energization paths in the DSR problem model as an MDP within the framework of DRL. The formulated MDP can be described as *MDP* $= \{\mathcal{S}, \mathcal{A}, P, R, \gamma, \mathcal{T}\}$, where:

- State $s_t \in \mathcal{S}$ is defined as the available observation vector of the overall distribution system at time t. The vector consists of total loads that are currently restored, the voltages at the individual load, the status of the operable switches, and the status of energization trajectories.
- Action $a_t \in \mathcal{A}$ is defined as the control actions applied to the operable switches, specifically selecting which switch to activate, in order to maximize the load restoration during DSR operations while ensuring operational and physical constraints.
- Transition probability $\mathcal{P} : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$: Given state s_t and action a_t at time t, the distribution system transits to state s_{t+1} at time step $t + 1$ according to the transition probability $\mathcal{P}(s_{t+1}|s_t, a_t)$.
- Time horizon T defines the sequence of decision-making steps.
- Discount factor $\gamma \in [0, 1]$ is to formulate the importance of future reward.
- Reward $R : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ evaluates the effectiveness of action a_t taken in state s_t in achieving the objective of our DSR problem, which is to maximize the load restoration in the time horizon T while adhering to operational and physical constraints. Therefore, the reward function will be formulated as:

$$r_t = R_t(s_t, a_t) = R_t^A(s_t, a_t) + w_p \times R_t^V(s_t, a_t) \quad (1)$$

where $R_t^A(s_t, a_t)$ is the reward related to the total active power restoration at t, which is defined as:

$$R_t^A(s_t, a_t) = \left(\sum_{l \in L} x_{l,t}^L P_{l,t}^L\right) \times \Delta t \quad (2)$$

By incorporating the concept of node cell, Eq. (2) can be rewritten as:

$$R_t^A(s_t, a_t) = \left(\sum_{l \in C} x_{l,t}^C P_{l,t}^C\right) \times \Delta t \quad (3)$$

Additionally, $R_t^V(s_t, a_t)$ is a penalty term to penalize actions that violate the voltage constraints. It is defined as:

$$R_t^V(s_t, a_t) = -\left(\sum_{l \in L} x_{l,t}^L V_{l,t}^{L,p}\right) \times \Delta t \quad (4)$$

where

$$V_{l,t}^{L,p} = \max\left(0, H_{l,t}^L - H_l^{\max}\right) + \max\left(0, H_l^{\min} - H_{l,t}^L\right) \quad (5)$$

Where $V_{l,t}^{L,p}$ represents the penalty terms associated with individual loads to ensure their voltage magnitude do not violate the constraints. These constraints are formulated using squared nodal voltage range $[H_i^{\min}, H_i^{\max}]$. The weight term w_p is set to ensure that the penalty is comparable to the active restored power term. Additionally, in our current initial stage of development, the topological constraints are hard-coded with a check function to automatically filter out violated actions.

Within the DRL framework, the objective of the DSR problem is to adaptively learn an optimal policy π^* that maximizes the expected sum of rewards $\mathbb{E}[\sum_{t=1}^{T} r_t]$ over the trajectory $\tau = (s_1, a_1, r_1, s_2, a_2, r_2, \ldots, s_T, a_T, r_T)$.

Proposed PIDT Framework for DSR Decision Makings

We continue to introduce our proposed innovative PIDT framework that transforms our formulated DRL framework by exploring the powerful computing capabilities of causal transformers. Figure. 1 illustrates the overview structure our proposed PIDT framework. As shown in Fig. 1, our

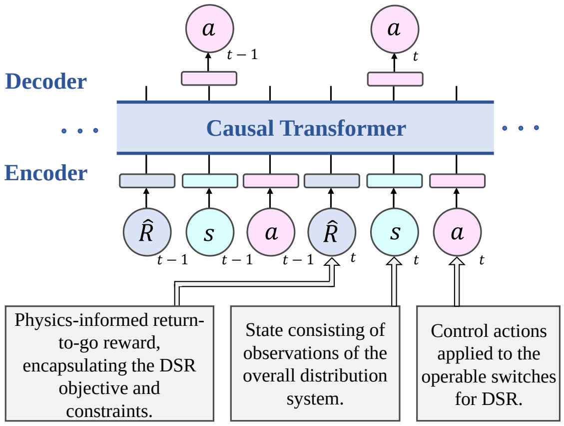


Figure 1: Overview of the architecture of our proposed PIDT framework.

proposed PIDT framework mainly consists of: 1) an encoder that consists of linear layers followed by an embedding layer; 2) a GPT-based causal transformer with a causal self-attention mask; and 3) a decoder. The encoder is developed to take a trajectory of rewards, states, and actions as input, and process the trajectory to generate token embeddings that will be fed into the GPT-based causal transformer model for further processing. To achieve this, the trajectory input $\hat{\tau}_{-k:t}$ of length T of the encoder is formulated $\hat{\tau}_{-k:t} = \{\hat{R}_k, s_k, a_k, \hat{R}_{k+1}, s_{k+1}, a_{k+1}, \ldots, \hat{R}_{t-1}, s_{t-1}, a_{t-1}, \hat{R}_t, s_t\}$ as illustrated in Fig. 1, which is different from the trajectory formulated in the original DRL framework that is described in previous section.

To appropriately formulate the trajectory input, it is essential to define a physics-informed return-to-go reward $\hat{R}_t$ such that 1) the trajectory $\hat{\tau}_{-k:t}$ can be effectively processed by the GPT-based causal transformer, and 2) the objective and physical/operational constraints for DSR are accurately characterized. To achieve this, we formulate $\hat{R}_t$ by transforming the reward r_t for the original DRL framework, which is defined in Eq. (1), by defining the desired target return as $\hat{R}^* = \sum_{t=1}^{T} \left(\sum_{l \in L} P_{l,t}\right)$, where $P_{l,t}$ is nominal active power of Load l at time t. Based on $\hat{R}^*$ and r_t defined in Eq. (1), we are able to represent $\hat{R}^*$ in training and inference procedures as follows:

1) Training Procedure:

$$\hat{R}_t = \sum_{i=t}^{T} r_i \tag{6}$$

2) Inference Procedure:

$$\begin{cases} \hat{R}_0 = \hat{R}_1 = \hat{R}^* \\ \hat{R}_t = \hat{R}^* - \sum_{i=1}^{t-1} r_i, \quad t = 2 \ldots T \end{cases} \tag{7}$$

After projecting the trajectory $\hat{\tau}_{-k:t}$ to the embedding dimension, the encoder forwards the token embeddings for $\hat{\tau}_{-k:t}$ to GPT-based casual transformer that is designed to generate a deterministic action at time t such that $a_t = \hat{\pi}(\hat{\tau}_{-k:t})$, as shown in Fig. 1. The policy $\hat{\pi}$ within our PIDT framework is parameterized by the GPT-based casual transformer where the action sequences are generated via autoregressive modeling. The policy $\hat{\pi}$ is trained by minimizing the cross-entropy loss between the predicted actions and the ground-truth actions in a sampled batch of trajectory data. The output of the GPT-based casual transformer is fed to the decoder that projects the token embeddings of the predicted trajectory back to the original action space, resulting in the final predicted control actions on the operational switches for DSR.

Model Training. The overall training procedure is described as follows:

1) We prepare a dataset D consisting of "offline" trajectories for DSR operations on the targeted distribution system. These trajectories can be collected from experts in power system domain or can be collected from simple off-line random walks to generate the sequences of control actions applied to the operable switches. These trajectories do not need to be optimal.
2) The minibatches of sequence length K from the dataset D will be fed to the PIDT framework to make decision on switching actions, and parameters of the PIDT framework will be updated according to the cross-entropy loss.
3) Repeat M episodes for Step 2.

The procedure is shown in Algorithm 1.

Algorithm 1: The PIDT Model Training

Initialize: Dataset D, PIDT model with learnable parameter θ, sample size L, maximum number of episodes M, and minibatch size b.

```
1: for episode m = 1 to M do
2:     Sample a random minibatch B of b sequences of
       length K from D.
3:     Obtain predictions using PIDT from the batch B.
4:     Calculate the cross-entropy loss between predicted
       and ground-truth values of action in sequence.
5:     Update θ.
6: end for
```

Model Inference. The overall inference procedure is shown in Algorithm 2. It is worth noting that we always keep only the last K time steps in the trajectory.

Performance Evaluations

In this section, we evaluate the performance of our proposed PIDT framework in two case studies with the modified IEEE 13-node test feeder (IEEEStd 2014) and the modified IEEE 123-node test feeder (IEEEStd 2014), respectively. To evaluate the performance of the PIDT framework, we compare it with two benchmark DRLs, the PPO algorithm (Schulman et al. 2017), and the A2C algorithm (Mnih et al. 2016), in the DSR operation.

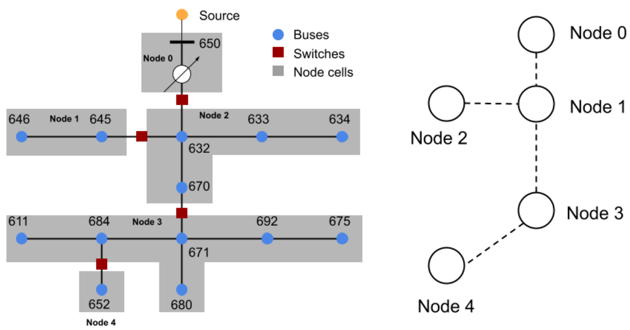


Figure 2: The physical topology and the corresponding graph representation of the modified IEEE 13-node test feeder.

Modified IEEE 13-Node Test Feeder

In this case study, we use a modified IEEE 13-node test feeder that is available in the Open Distribution Simulator Software (OpenDSS). The system topology and its corresponding node-cell graph are shown in Fig. 2. As illustrated in Fig. 2, in this system, there is a substation located in a node named "source" that is connected to Node 650, with an additional node (namely Node 670) that connects between Node 632 and 670. After reformed into node cells, there will be a total of 5 node cells to be energized.

Additionally, based on objective and physical/operational constraints for the DSR operation, the ground-truth energization path in this case is determined as (Node 0) $\rightarrow$ (Node 1) $\rightarrow$ (Node 3) $\rightarrow$ (Node 4). And the corresponding final restored power should be 3006.509 kW, which is considered as the objective of the DSR operation and is also the desired target return in our PIDT method. The learning curves for the average return of the first 1500 gradient updates by using our PIDT method as well as the other two benchmark DRL methods are shown in Fig. 3. We would like to mention that, as described in the section of our proposed PIDT framework, return is defined as the accumulated reward. The evaluation results of the three methods across 50 independent trails in the inference stage are stated in Table 2. As shown in Table 2, we compare the performance of these three methods from four perspectives, including average return, standard deviation of return values, the number of optimal solutions in the 50 trials, and the number of suboptimal solutions. Additionally, a bar char in Fig. 4 shows more insights of the simulation results, which presents the distribution of power restoration levels in the 50 independent trials using the three methods, respectively.

From Figs. 3 and 4 and Table 2, we can observe that, during the learning process, the A2C method demonstrates the quickest convergence and achieves the highest training return. However, it typically settles for a suboptimal solution, maintaining power restoration within the range of $[1000, 3006]$ kW for the majority of trials. In contrast, our method, while not showing the fastest convergence or the highest training return, achieves optimal solutions in 42 out of 50 independent trials. This is reasonable since the return-

---

Algorithm 2: Inference of the Trained PIDT Model

Initialize: PIDT model with trained parameter θ, return-to-go $\hat{R}_0$, and initial state of the distribution system s_0.

1: Set $\hat{R}_1 \leftarrow \hat{R}_0, s_1 \leftarrow s_0, \tau \leftarrow (s_1, \hat{R}_1)$
2: **for** time step $t = 1$ to $T - 1$ **do**
3: Obtain $a_t \leftarrow PIDT(\tau)$ and $r_t \leftarrow R_t(s_t, a_t)$
4: Observe the next state s_{t+1} after taking switching action a_t, and obtain $\hat{R}_{t+1} \leftarrow \hat{R}_t - r_t$.
5: Append $(a_t, s_{t+1}, \hat{R}_{t+1})$ to τ, and keep the last K time steps of τ (i.e., $\tau \leftarrow \tau_{-K:t+1}$).
6: **end for**

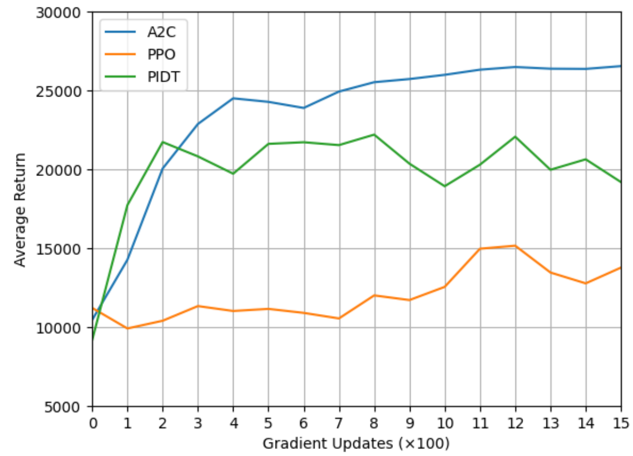


Figure 3: Learning curves for the average returns of the first 1500 gradient updates using our PIDT method and the two other benchmark DRL methods, PPO and A2C, for the DSR operations in the modified 13-node test feeder. The curve is updated per hundred gradient updates.

Evaluation Results in the Inference Stage	**A2C**	**PPO**	**Our method**
Average return	26659.982	14810.724	20361.480
Standard deviation of returns	643.940	8622.008	6130.914
Number of optimal solutions	17	18	42
Number of suboptimal solutions	33	32	8

Table 2: Further performance comparison between our PIDT method and other two benchmark DRL methods for the DSR operation in the modified IEEE 13-node test feeder

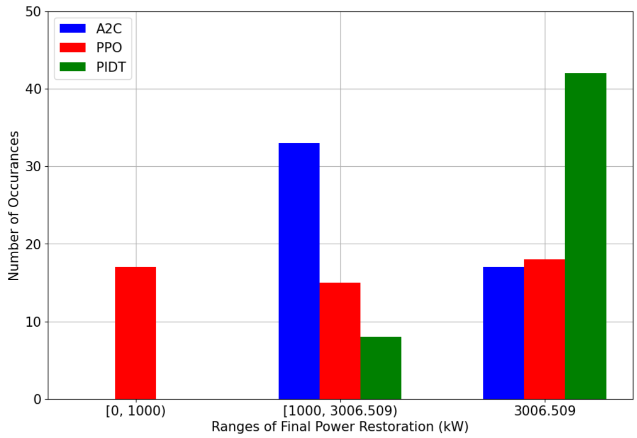


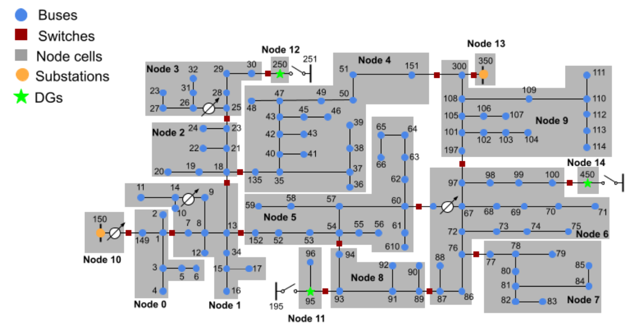
Figure 4: Distribution of power restoration levels in the 50 independent trials using the three methods, respectively, in the modified 13-node test feeder.

to-go reward defined in our proposed PIDT method is different from the rewards in conventional DRL. Additionally, from the simulation results, we can also see that the performance of the PPO method falls between that of A2C and our PIDT methods.

Modified IEEE 123-Node Test Feeder

In this case study, we evaluate the performance of our PIDT method by comparing with the PPO- and A2C-based benchmark DRL methods in DSR operations for a modified 123-node test feeder. The system topology and its corresponding node cell graph are shown in Fig. 5. As shown in Fig. 5, in the modified 123-node test feeder, there are five energy sources in the system, two of which are substations located in Nodes 150 and 350, and the three other sources are distributed generators (DGs) located in Nodes 95, 250 and 450. Additionally, the system has a total of 15 node cells to be energized. We also incorporate the modification on the energization ability to the DG in Node 250, such that this DG can only fully energize loads in Node Cell 3 and Node Cell 2. The DG will be automatically shut down if it tries to energize more loads other than those in Node Cell 3 and Node Cell 2.

The actual total power of the system, if all the loads are energized properly, would be 3350.757 kW, which is considered as the objective of the DSR operation and is also the desired target return in our PIDT method. An expected optimal solution would power all the loads properly. However, due to the large-scale and complex topology of the modified IEEE 123-node test feeder, a simple optimal solution to power all the loads is very challenging. The learning curve of the first 6000 gradient updates using our PIDT method and the two benchmark DRL methods are shown in Fig. 6. As shown in Fig. 6, for the modified IEEE 123-node test feeder that has large-scale and complex topology, our method outperforms the other two methods in convergence rate and convergence reward in general. The evaluation results of these three methods across 50 independent trials in the inference



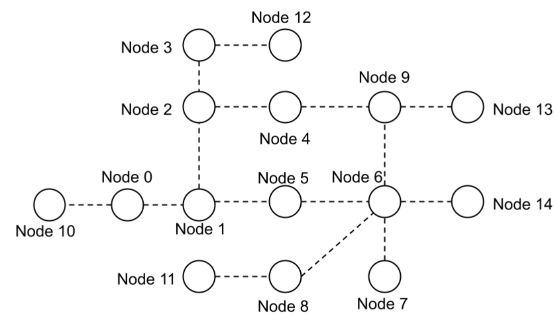


Figure 5: The physical topology and the corresponding graph representation of the modified IEEE 123-node test feeder.

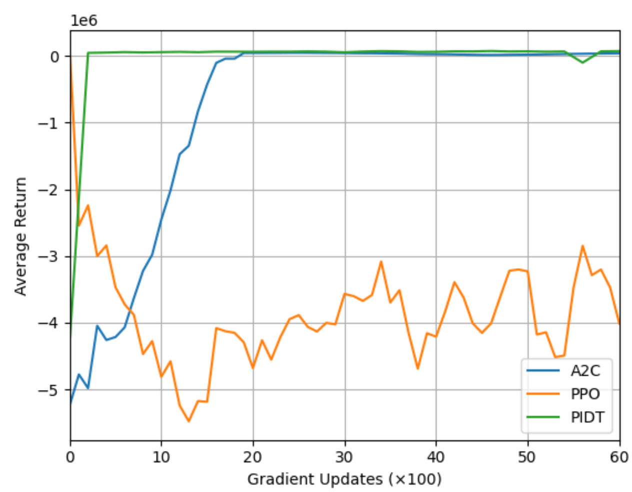


Figure 6: Learning curve for the average return of the first 6000 gradient updates using our PIDT method and the two other benchmark DRL methods, PPO and A2C, for the DSR operations in the modified 123-node test feeder. The curve is updated per hundred gradient updates.

stage are presented in Table 3. As shown in Table 3, our proposed PIDT method can achieve near-optimal solutions in 30 of the 50 independent trials, which outperforms both the PPO and A2C methods (in this case study, the near-optimal

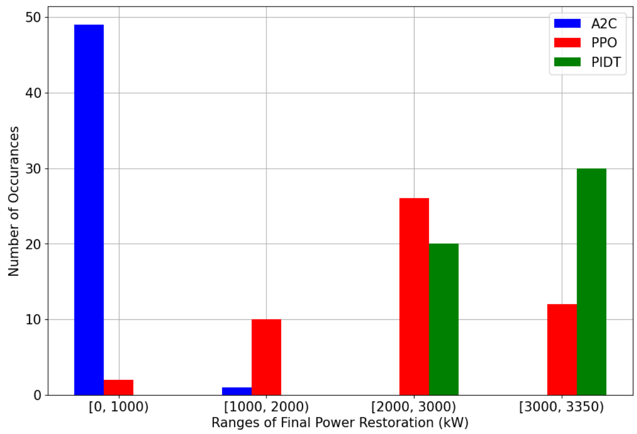


Figure 7: Distribution of power restoration levels in the 50 independent trials using the three methods, respectively, in the modified 123-node test feeder.

solutions are defined as the solutions whose final restored power is more than 3000 kW).

Furthermore, Fig. 7 provides deeper insights into the simulation results shown in Table 3 by illustrating the distribution of power restoration levels across 50 independent trials using the three methods. It shows that our method achieves near-optimal power restoration in 30 trials and between 2000 kW to 3000 kW in the remaining 20 trials. In contrast, the solutions generated by the A2C-based method predominantly fall within the range of 0 kW to 1000 kW. The PPO-based method shows a more varied distribution across all three ranges, with 12 trials achieving near-optimal results. These findings underscore the superior performance of our method over the other two conventional DRL methods.

Evaluation Results in the Inference Stage	**A2C**	**PPO**	**Our method**
Average return	47414.036	-4659896.17	100915.721
Standard deviation of returns	1675.582	5724583.67	19100.489
Average power restoration (kW)	988.044	2452.702	2910.923
Standard deviation of power restorations (kW)	29.400	662.671	369.444
Number of near-optimal solutions	0	12	30

Table 3: Further performance comparison between our PIDT method and other two benchmark DRL methods for the DSR operation in the modified IEEE 123-node test feeder

Conclusions

While different DRL-based approaches have been developed for effectively addressing DSR challenges amidst uncertain operational conditions, the data-intensive nature of DRL presents barriers to achieving robust solutions for large-scale, complex distribution systems. This paper explores an innovative strategy inspired by the transformative impact of emerging causal transformers that widely work as foundation models for LLMs. Specifically, we present a first-ever effort to explore the powerful computing capabilities of a GPT-based causal transformer to tackle scalability challenges inherent in traditional DRL methods for DSR operations. This study marks the pioneering application of emerging causal transformers to revolutionize conventional DRL practices in power system operations. Our contributions are twofold: 1) introducing a novel PIDT framework that explores a GPT-based causal transformer to transform conventional DRL methods for DSR operations, and 2) conducting comparative studies to analyze the performance of the proposed PIDT framework in its initial development stage for solving DSR problem. Simulation results underscore the effectiveness of our proposed PIDT method in resolving DSR problems within large-scale distribution systems. In our ongoing work, we are leveraging the insights gained from our current initial-stage development to further enhance the scalability and resilience of our proposed PIDT method. While our primary focus in this paper lies in DSR, the proposed PIDT framework demonstrates potential applicability in optimizing sequential decision-making across a spectrum of power system operations.

References

Al-Hinai, A.; and Alhelou, H. H. 2021. A multi-agent system for distribution network restoration in future smart grids. *Energy Reports*.

Arefifar, S. A.; Mohamed, Y. A.-R. I.; and El-Fouly, T. H. M. 2012. Supply-Adequacy-Based Optimal Construction of Microgrids in Smart Distribution Systems. *IEEE Transactions on Smart Grid*, 3(3): 1491–1502.

Bassey, O.; and Butler-Purry, K. L. 2020. Black Start Restoration of Islanded Droop-Controlled Microgrids. *Energies*, 13(22).

Canese, L.; Cardarilli, G. C.; Di Nunzio, L.; Fazzolari, R.; Giardino, D.; Re, M.; and Spanò, S. 2021. Multi-Agent Reinforcement Learning: A Review of Challenges and Applications. *Applied Sciences*, 11(11).

Chen, B.; Ye, Z.; Chen, C.; and Wang, J. 2019. Toward a MILP Modeling Framework for Distribution System Restoration. *IEEE Transactions on Power Systems*, 34(3): 1749–1760.

Chen, C.; Wang, J.; and Ton, D. T. 2017. Modernizing Distribution System Restoration to Achieve Grid Resiliency

Against Extreme Weather Events: An Integrated Solution. *Proceedings of the IEEE*, 105: 1267–1288.

Chen, L.; Lu, K.; Rajeswaran, A.; Lee, K.; Grover, A.; Laskin, M.; Abbeel, P.; Srinivas, A.; and Mordatch, I. 2021. Decision Transformer: Reinforcement Learning via Sequence Modeling. In *Neural Information Processing Systems*.

Chen, W.-H.; Tsai, M.-S.; and Kuo, H.-L. 2005. Distribution system restoration using the hybrid fuzzy-grey method. *IEEE Transactions on Power Systems*, 20(1): 199–205.

Du, W.; and Ding, S. 2020. A survey on multi-agent deep reinforcement learning: from the perspective of challenges and applications. *Artificial Intelligence Review*, 54: 3215 – 3238.

Du, Y.; and Wu, D. 2022. Deep Reinforcement Learning From Demonstrations to Assist Service Restoration in Islanded Microgrids. *IEEE Transactions on Sustainable Energy*, 13(2): 1062–1072.

Gao, Y.; Wang, W.; Shi, J.; and Yu, N. 2020. Batch-Constrained Reinforcement Learning for Dynamic Distribution Network Reconfiguration. *IEEE Transactions on Smart Grid*, 11(6): 5357–5369.

IEEEStd. 2014. IEEE Recommended Practice for Electric Power Distribution System Analysis. *IEEE Std 1729-2014*, 1–20.

Igder, M. A.; Liang, X.; and Mitolo, M. 2022. Service Restoration Through Microgrid Formation in Distribution Networks: A Review. *IEEE Access*, 10: 46618–46632.

Kasneci, E.; Sessler, K.; Küchemann, S.; Bannert, M.; Dementieva, D.; Fischer, F.; Gasser, U.; Groh, G.; Günnemann, S.; Hüllermeier, E.; Krusche, S.; Kutyniok, G.; Michaeli, T.; Nerdel, C.; Pfeffer, J.; Poquet, O.; Sailer, M.; Schmidt, A.; Seidel, T.; Stadler, M.; Weller, J.; Kuhn, J.; and Kasneci, G. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103: 102274.

Liang, X.; Saaklayen, M. A.; Igder, M. A.; Shawon, S. M. R. H.; Faried, S. O.; and Janbakhsh, M. 2022. Planning and Service Restoration Through Microgrid Formation and Soft Open Points for Distribution Network Modernization: A Review. *IEEE Transactions on Industry Applications*, 58(2): 1843–1857.

Mnih, V.; Badia, A. P.; Mirza, M.; Graves, A.; Lillicrap, T. P.; Harley, T.; Silver, D.; and Kavukcuoglu, K. 2016. Asynchronous Methods for Deep Reinforcement Learning. In *International Conference on Machine Learning*.

Pao-La-Or, C.; Bhasaputra, P.; Kulworawanichpong, T.; and Numsomran, A. 2009. Multiagent-based service restoration system with rule-based approach for distribution systems. *Journal of Electrical Engineering and Technology*, (4): 454–462.

Patsakis, G.; Rajan, D.; Aravena, I.; and Oren, S. 2019. Strong Mixed-Integer Formulations for Power System Islanding and Restoration. *IEEE Transactions on Power Systems*, 34(6): 4880–4888.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal Policy Optimization Algorithms. *ArXiv*, abs/1707.06347.

Sedzro, K. S. A.; Shi, X.; Lamadrid, A. J.; and Zuluaga, L. F. 2019. A Heuristic Approach to the Post-Disturbance and Stochastic Pre-Disturbance Microgrid Formation Problem. *IEEE Transactions on Smart Grid*, 10(5): 5574–5586.

Singh, R.; Mehfuz, S.; and Kumar, P. 2016. Intelligent decision support algorithm for distribution system restoration. *Springerplus*, 5(1): 1175.

Thirunavukarasu, A. J.; Ting, D. S. J.; Elangovan, K.; Gutierrez, L.; Tan, T. F.; and Ting, D. S. W. 2023. Large language models in medicine. *Nat. Med.*, 29(8): 1930–1940.

Wang, F.; Cetinay, H.; He, Z.; Liu, L.; Van Mieghem, P.; and Kooij, R. E. 2023. Recovering Power Grids Using Strategies Based on Network Metrics and Greedy Algorithms. *Entropy*, 25(10).

Wang, Z.; and Wang, J. 2015. Self-Healing Resilient Distribution Systems Based on Sectionalization Into Microgrids. *IEEE Transactions on Power Systems*, 30(6): 3139–3149.

Wei, J.; Kim, S.; Jung, H.; and Kim, Y.-H. 2024. Leveraging Large Language Models to Power Chatbots for Collecting User Self-Reported Data. *Proc. ACM Hum.-Comput. Interact.*, 8(CSCW1).

Wu, J.; Fang, B.; Fang, J.; Chen, X.; and Tse, C. K. 2019. Sequential topology recovery of complex power systems based on reinforcement learning. *Physica A: Statistical Mechanics and its Applications*, 535: 122487.

Yang, L.-J.; Zhao, Y.; Wang, C.; Gao, P.; and Hao, J.-H. 2019. Resilience-Oriented Hierarchical Service Restoration in Distribution System Considering Microgrids. *IEEE Access*, 7: 152729–152743.

Yao, S.; Gu, J.; Zhang, H.; Wang, P.; Liu, X.; and Zhao, T. 2020. Resilient Load Restoration in Microgrids Considering Mobile Energy Storage Fleets: A Deep Reinforcement Learning Approach. In *2020 IEEE Power & Energy Society General Meeting (PESGM)*, 1–5.

Yao, Y.; Zhang, X.; Wang, J.; and Ding, F. 2023. Multi-Agent Reinforcement Learning for Distribution System Critical Load Restoration. In *2023 IEEE Power & Energy Society General Meeting (PESGM)*, 1–5.

Zhao, T.; and Wang, J. 2022. Learning Sequential Distribution System Restoration via Graph-Reinforcement Learning. *IEEE Transactions on Power Systems*, 37(2): 1601–1611.