Data-Driven Outage Modelling and Restoration Time Prediction in Distribution Grids

Zhaoyu Wang
Northrop Grumman Associate Professor
Iowa State University

Motivation of Data-Driven Outage Time Prediction

- Power outages have significant impacts on production, transportation, communication, and health supply service, resulting in significant economic losses.

- In recent years, customers experienced longer outages. In 2018, each customer lost power for around 5.8 hours.

- In February 2021, the state of Texas suffered a major power crisis, more than 4.5 million homes and businesses were affected.

- From the customer's perspective, the most important and concerned information is timely and accurate outage recovery time prediction, which will greatly help them plan for subsequent arrangements in advance.

Source: https://www.eia.gov/ https://poweroutage.us

Existing Work and Challenges

Reference	Data Source	Approach	Case Study	Cons
[1]	Severe weather records	Accelerated failure time model	Estimate duration of historical outages	Data distribution assumption, uses only weather data as variables, limited data source
[2]	Radar observations data	Bayesian prediction algorithm	Provide an estimation of outage duration	
[3]	Historical outage data with severe weather records	Deep neural network	Predict repair and restoration time with respect to severe weather events	Single global model, each outage recovery is treated as an isolated process

- Challenges:
  - Outages occurring together in a time period can impact restoration time; previous studies ignored the correlation among overlapped outages.
  - Outages may have different scales (i.e., a couple of minutes to several hours) and unbalanced distributions (i.e., some scales are rare); previous studies trained a single model for the entire dataset, which may cause an overfitting problem.

Real-World Outage Dataset - Overview

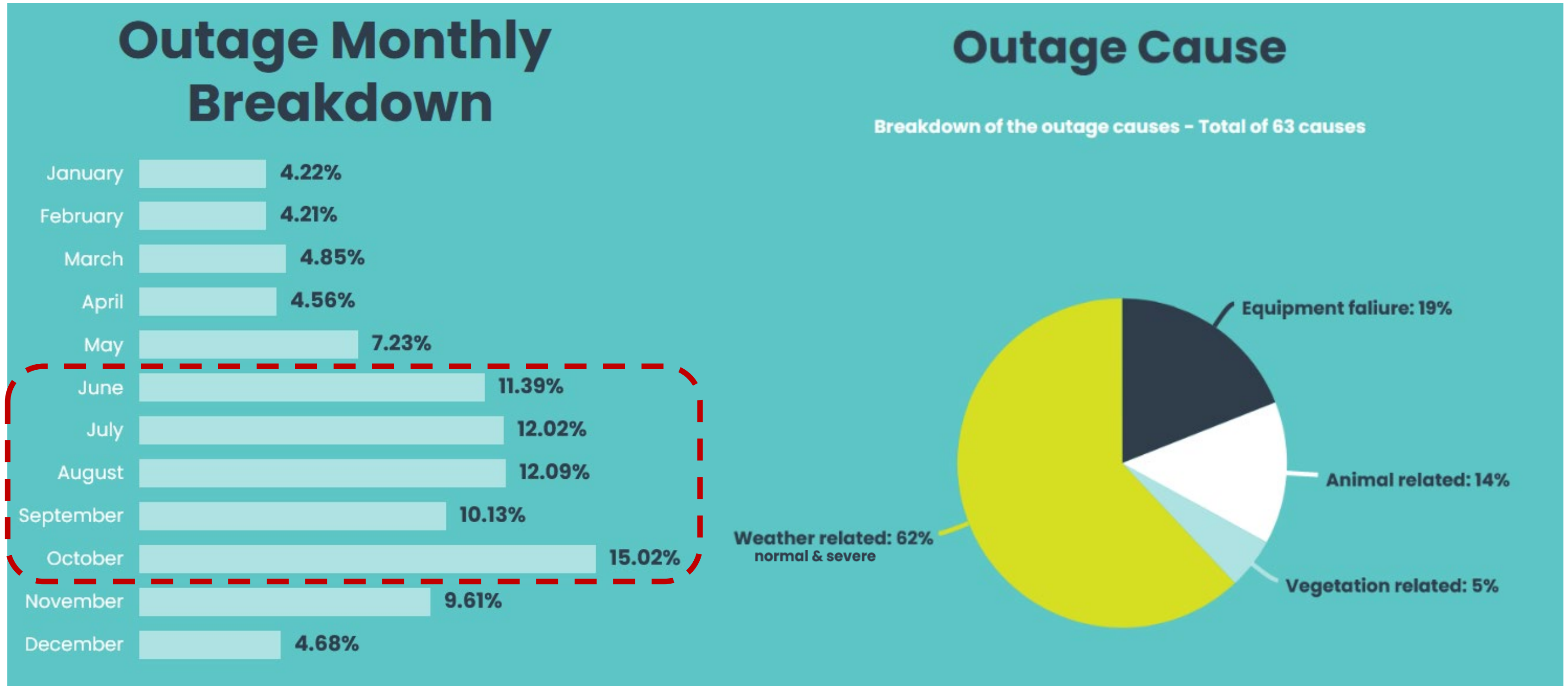
- The available outage reports are recorded by a utility provider located in the U.S., including over 16,000 outage records over a six-year period (2011 ~ 2016).
- The initial outage data features include:
  - Start and end time (accurate to seconds resolution)
  - Outage locations (latitude, longitude)
  - Numbers of customers interrupted
  - Repair and Restoration time (accurate to seconds resolution)
  - Causes (i.e., animal, tree, connector failure)

Start Time	End Time	Customers Interrupted	Latitude	Longitude	Duration(mins)	Pole Number	Cause Key	Sub Cause Key
Dec 1, 2012 3:21:00 AM	Dec 1, 2012 11:31:00 PM	126	xxx	xxx	2,010	B 0001	Connector	Tree/Limb In Clearance Zone
Dec 1, 2012 3:39:00 AM	Dec 1, 2012 6:56:00 AM	55	xxx	xxx	317	N 0006	System Failure	Animal, Squirrel

Dataset visualization with sample entries

Real-World Outage Dataset - Analysis

- Outages mostly occur during summer and fall.
- There are 63 causes, categorized as vegetation-related, animal-related, and equipment-related.
- The collected ***severe*** weather report has transferred to 8 discrete codes.



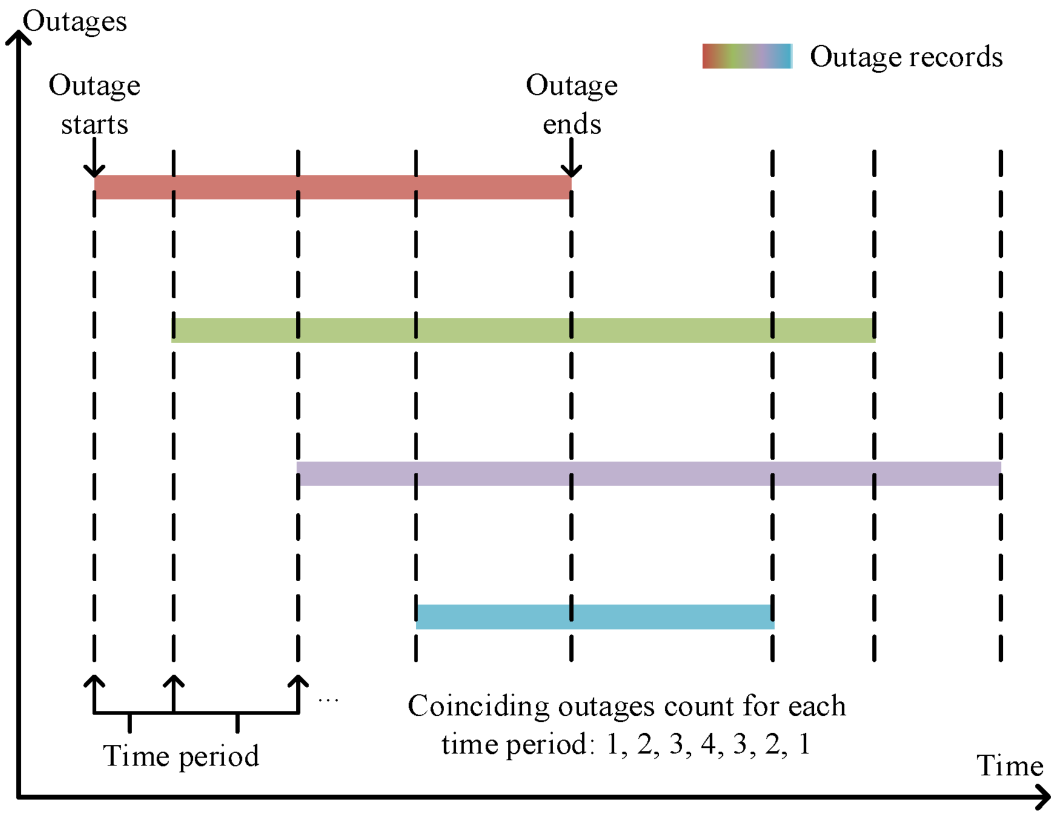
Real-World Outage Dataset - Analysis

- Outages can have overlapped time periods. We considered a new training feature: cumulative number of coinciding outages (i.e., the quantity of outages presented at a certain time period that has not yet been resolved).

$$c_{outages}^{t_i} = c_o^{t_i} - c_r^{t_i} \quad \textbf{(1)}$$

where is the $c_o^{t_i}$ cumulative total outages at time t_i, $c_r^{t_i}$ is the cumulative total restorations at time t_i, and $c_{outages}^{t_i}$ is the number of coinciding outages at time t_i.

- The cumulative number of customers interrupted is also considered as a new feature.

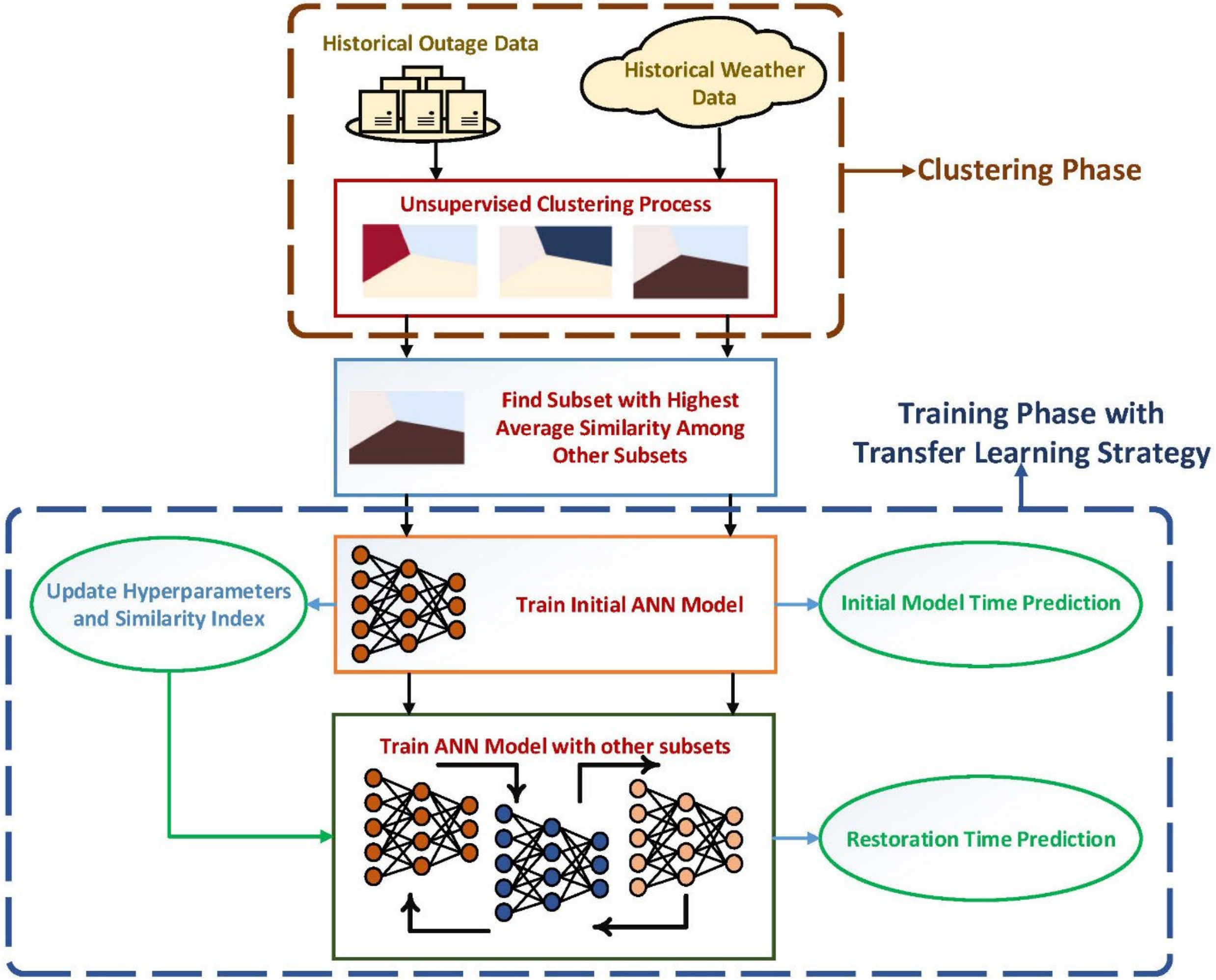


Example of the cumulative number of coinciding outages

Proposed Outage Restoration Time Prediction Methodology

We propose a multi-stage framework to estimate the restoration time in a cluster-wise manner:

- Dataset clustering to deal with the data imbalanced distribution problem.
- Find the optimal subset with the highest average similarity with other subsets to train an initial ANN model.
- Update hyperparameters and similarity index based on the trained model. Use the trained model as a source for the next training session.
- Each of the other outage subsets is assigned with an ANN to predict restoration time.



Outage Pattern Discovery Using Cluster Ensembles

- The sparse dictionary-based ensemble spectral clustering (SDESC) is leveraged to cluster the dataset.
- Unlike k-means, spectral clustering can better handle high-dimensional data and is robust against data noise.
- In SDESC, the sparse coding technique greatly decreases the complexity and cost of practical implementation.
- **Algorithm summary**:

Step I: The high-dimensional dataset is factorized into a low-dimensional dictionary matrix and a representation matrix.

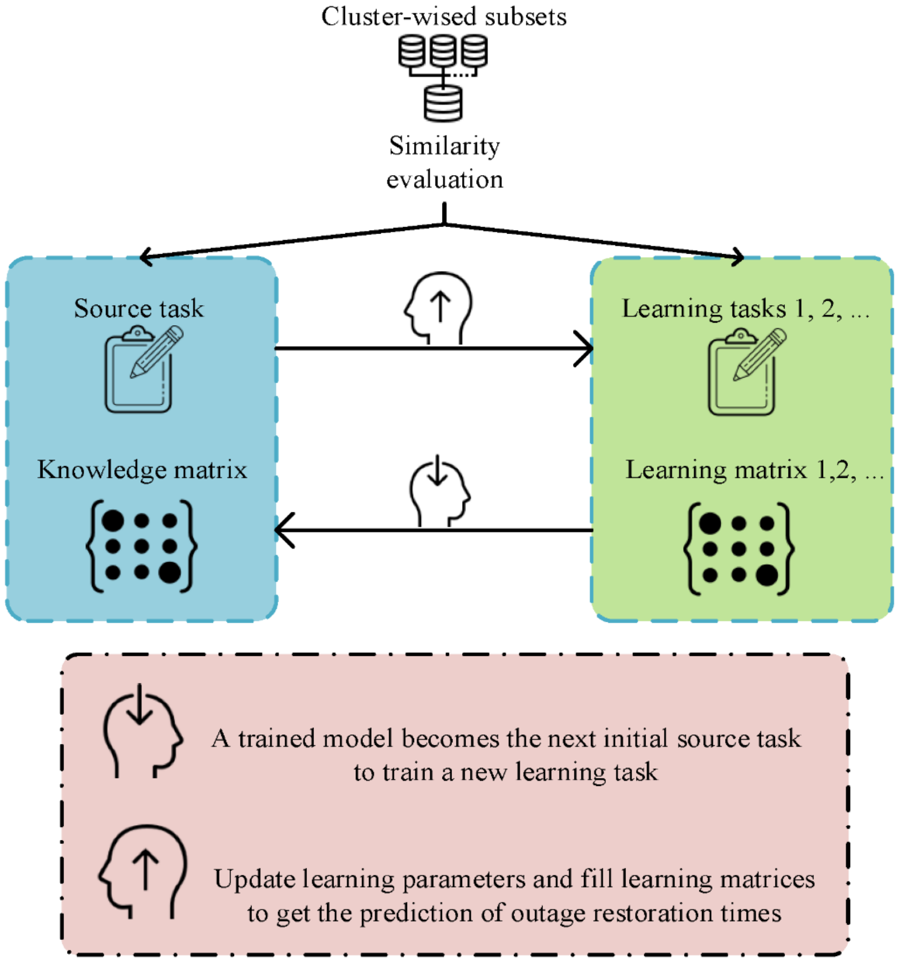
Step II: The similarity between different data points is used to distinguish data partitions in a graphical manner.

Step III: Solve the graph partition problem according to the optimal value of clusters.

Step IV: The optimal value of clusters can be determined by various clustering evaluation metrics such Davies-Bouldin validation index (DBI).

Outage Restoration Time Prediction Framework

- A transfer learning strategy discovers outage features and structures under different but related subsets.
- The transfer learning firstly gathers features and the output (i.e., actual restoration time) in the pre-trained model, and stores them as a source task.
- In this study, learning tasks are the training assignments of each outage subset, by exploiting the similarity between the learning task and the source task, the learning parameters can be updated for training a new prediction model.
- The learned model can be utilized in a recursive manner when dealing with a new learning task.



Numerical Results – Clustering Summary

- Using feature selection and clustering metrics evaluation, the dataset consists of 10 features (i.e., customer interrupted, cumulative outages, cause, and weather information) is clustered into 4 subsets using the SDESC algorithm.
- Collected high-precision weather-related data from the National Oceanic and Atmospheric Administration (NOAA) :
  - Hourly temperature
  - Hourly wind speed
  - Hourly precipitation
  - Severe weather reports
- Weather data aligned with each outage data record based on the start time.

Cluster	Samples	Avg. CI	Avg. RT (min)
$\mathbf{C_1}$	2379	170	740.5
$\mathbf{C_2}$	5302	21	288.4
$\mathbf{C_3}$	2884	16	144.5
$\mathbf{C_4}$	5872	22	82.2

- $\mathbf{C_1}$ refers to severe outages with higher Avg. RT and Avg. CI, but relatively infrequent.
- $\mathbf{C_2}$ and $\mathbf{C_4}$ represent intermediate and least serious outages, which are twice as frequent as severe outages.
- $\mathbf{C_4}$ represents a subset of minor outages, which occur frequently but can typically be resolved in a timely manner.

CI: Customer interrupted RT: restoration time

Source: https://www.ncei.noaa.gov/products/severe-weather https://www.ncdc.noaa.gov/cdo-web

Numerical Results – Similarity Matrix

- The similarity index (ranged 0-1) among different data subsets is calculated by a cross-validation principle with an unsupervised process.

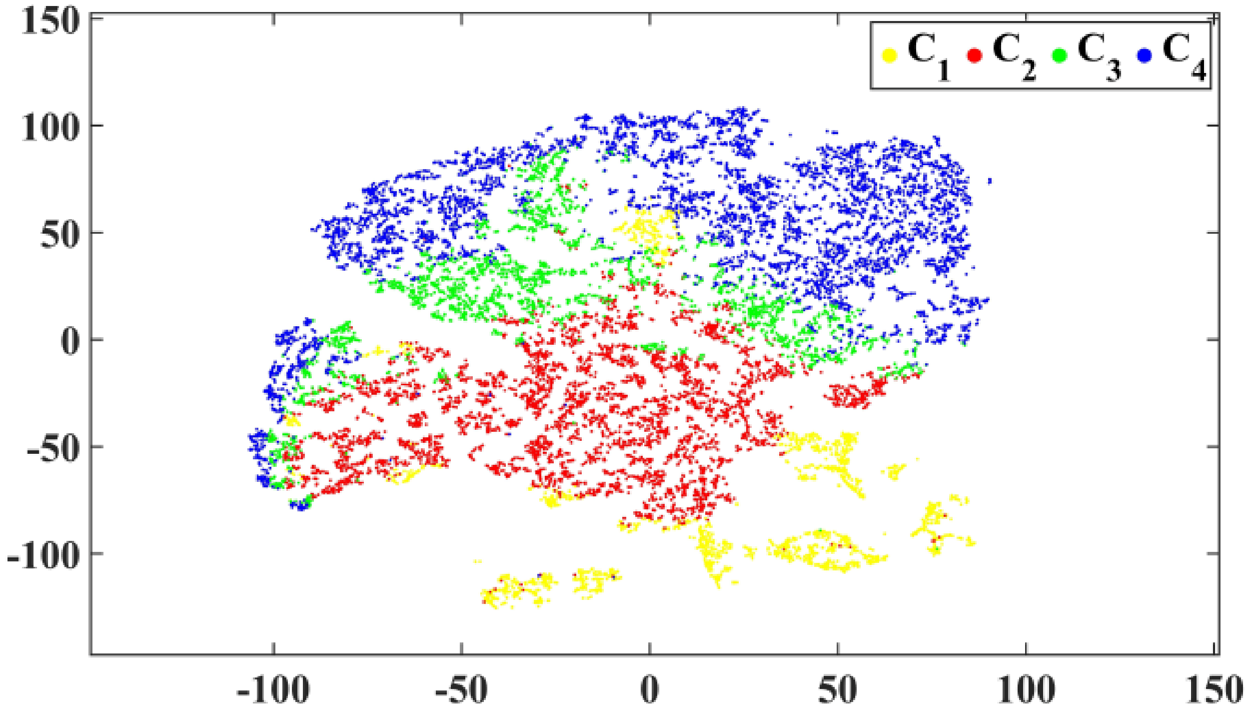
Similarity Index (%)	Cluster 1	Cluster 2	Cluster 3	Cluster 4
Cluster 1 with…		17.7	18.6	26.2
Cluster 2 with…	17.7		49.8	58.7
Cluster 3 with…	18.6	49.8		52.3
Cluster 4 with…	26.2	58.7	52.3	

- Average similarity with other subsets:
  - Cluster 1: 20.83%
  - Cluster 2: 42.06%
  - Cluster 3: 40.23%
  - Cluster 4: 45.73%

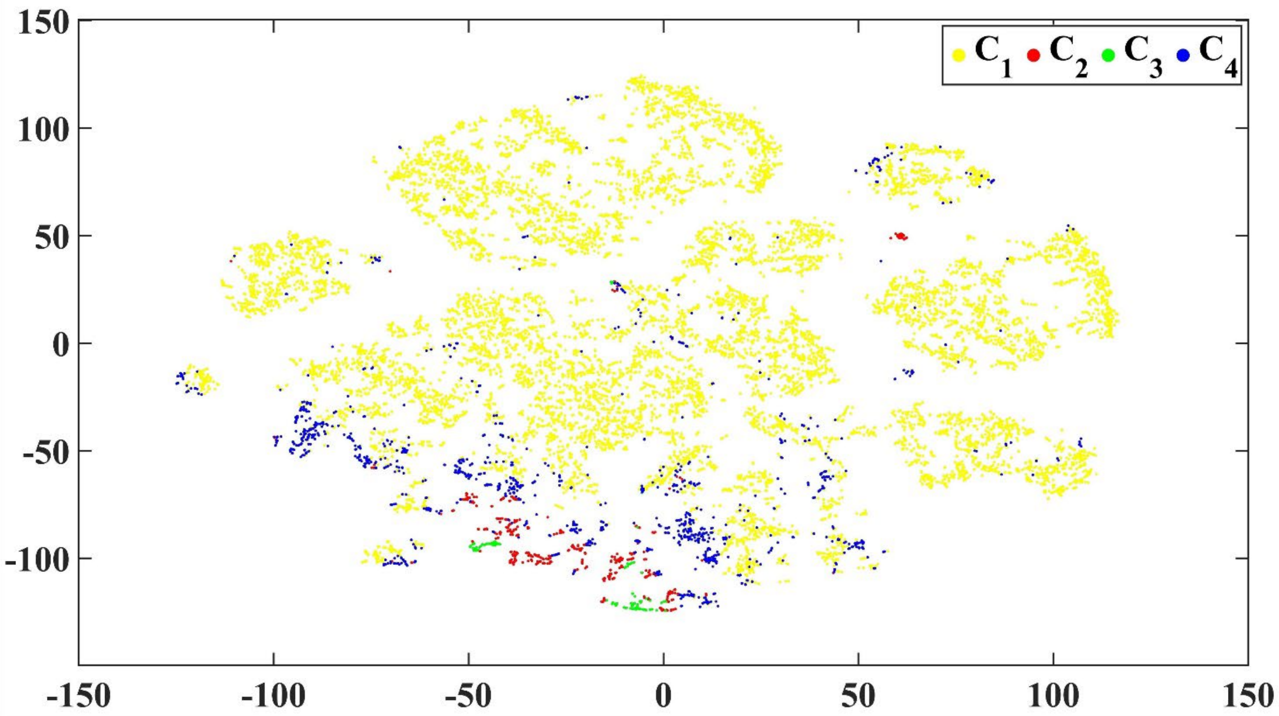
Similarity Index: 0 means no similarity between two samples, and 1 means two samples are the same.

Numerical Results – Clustering Visualization

- t-SNE [4] for high-dimensional data visualization and enhancing the overall interpretability of the framework.

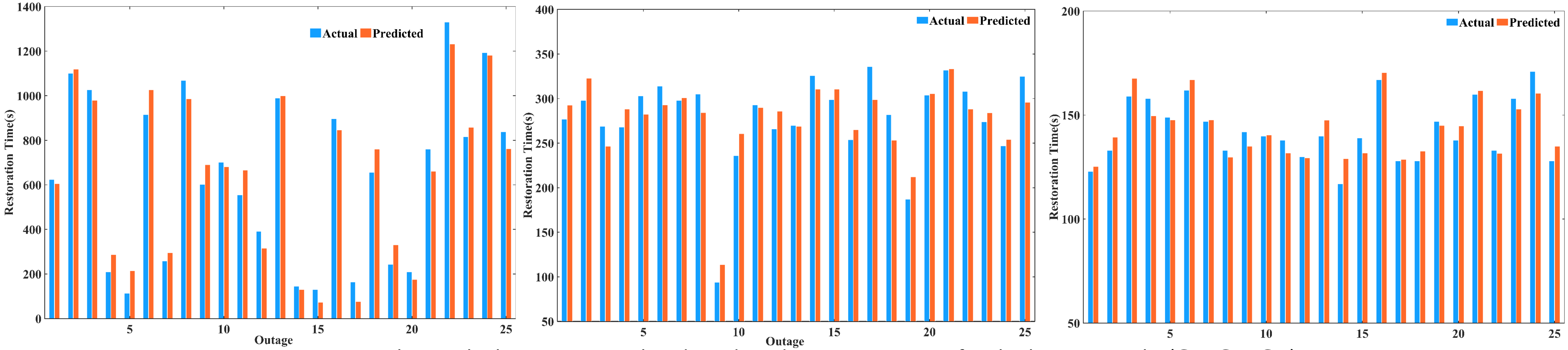


t-SNE plot of clustered data using the proposed SDESC method

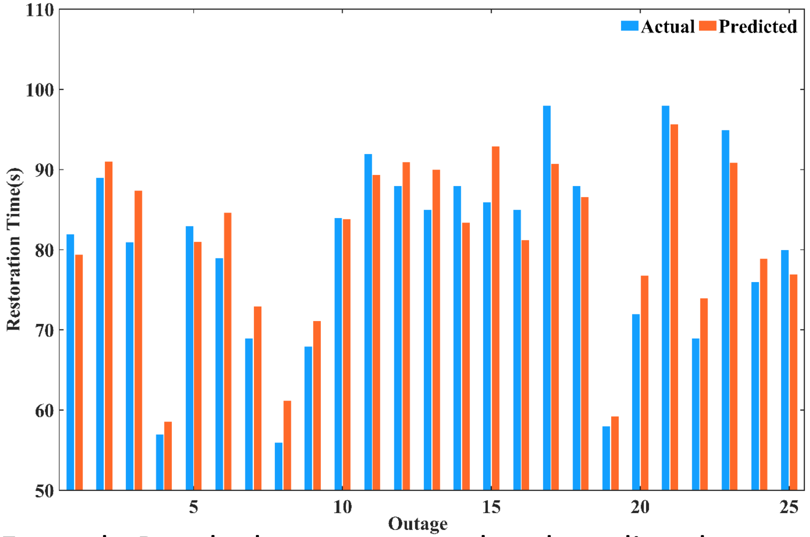


t-SNE plot of clustered data using the advanced k-means method

Numerical Results – Time Estimation



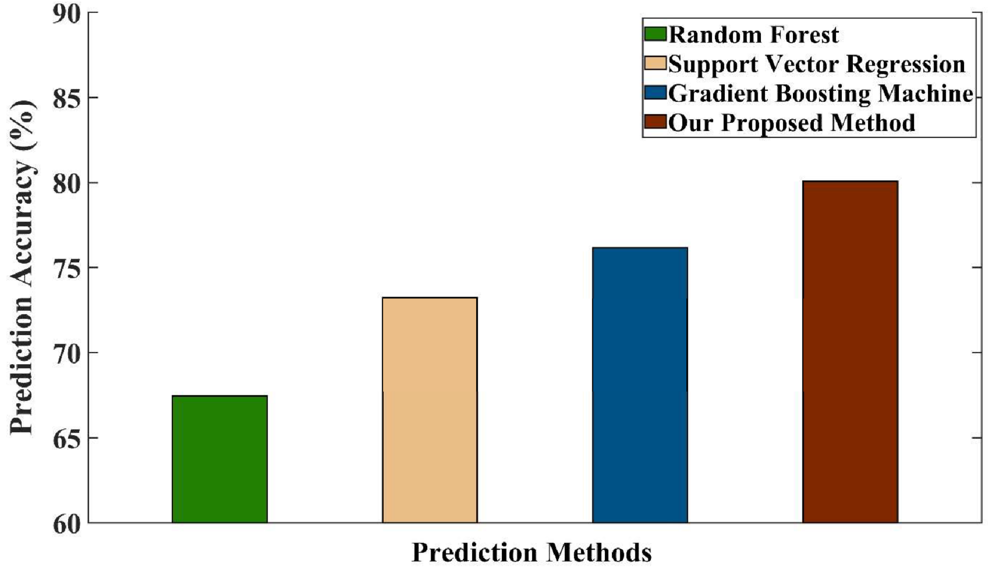
Example Results between actual and predicted restoration time for the learning tasks ($\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{C}_3$)



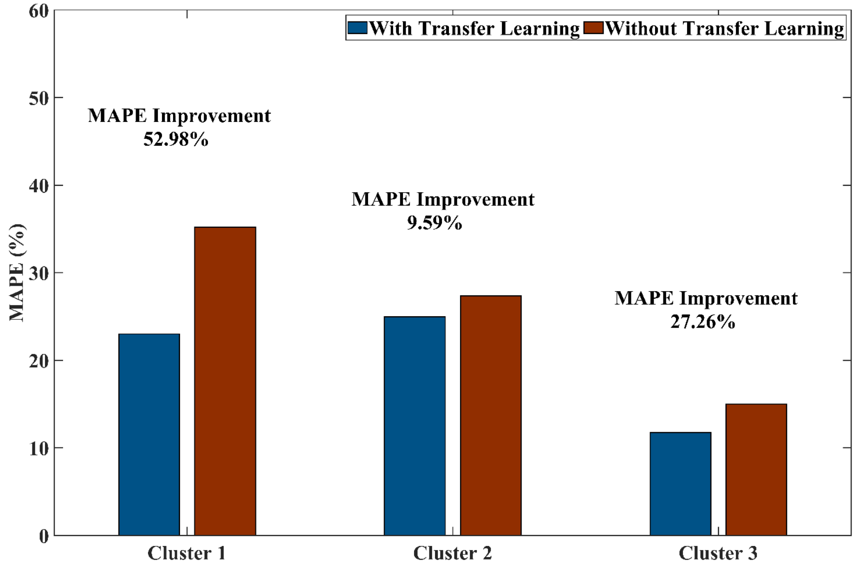
Example Results between actual and predicted restoration time for the source task ($\mathbf{C}_4$)

- ✓ $\mathbf{C}_4$ is chosen to be the source task based on the similarity evaluation.
- ✓ Other training tasks are utilizing the pre-trained model $\mathbf{C}_4$.
- ✓ Only 3% of the total predicted time is more than 60 minutes of the actual restoration time.

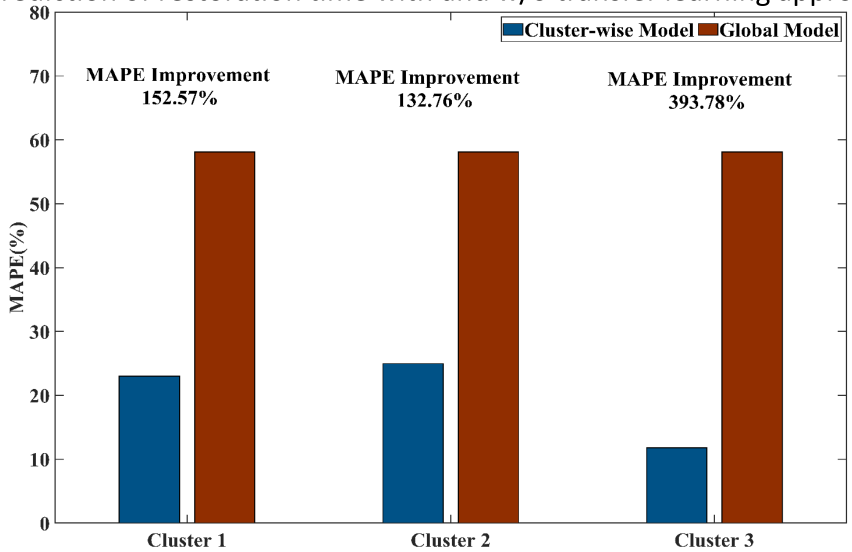
Numerical Results – Comparison



Comparison of prediction results with three existing methods



Prediction of restoration time with and w/o transfer learning approach



Restoration time comparison between cluster-wised model and global model

- ✓ We have conducted numerical comparisons with three existing works [5-7] (left figure) and global models without transfer learning and cluster-wise strategies (right figures).
- ✓ The proposed method can outperform the previous works. Also, the combination of transfer learning and cluster-wise strategies have proven to be valuable. The largest MAPE improvements are 393.78% and 52.98% respectively for implementing clustering and transfer learning strategies.

Conclusions

- Accurate outage restoration time predictions will greatly help customers and utilities plan for subsequent arrangements in advance.

- The proposed method estimates the restoration time in a cluster-wise manner to deal with the uncertainty caused by the heterogeneity of outage events.

- The transfer learning embedded framework solves the data imbalance problem caused by the data scarcity of the specific outage patterns.

- The results show that the proposed method has improved performance compared to existing methods and overcome large-scale data challenges.

References

- [1] H. Liu, R. A. Davidson, and T. V. Apanasovich, “Statistical forecasting of electric power restoration times in hurricanes and ice storms,” *IEEE Trans. Power Syst*., vol. 22, no. 4, pp. 2270–2279, 2007.
- [2] M. Yue, T. Toto, M. P. Jensen, S. E. Giangrande, and R. Lofaro, “A bayesian approach-based outage prediction in electric utility systems using radar measurement data,” *IEEE Trans. Smart Grid*, vol. 9, no. 6, pp. 6149–6159, November, 2018.
- [3] A. Arif and Z. Wang, “Distribution network outage data analysis and repair time prediction using deep learning,” *2018 IEEE International Conference on Probabilistic Methods Applied to Power Systems (PMAPS*), pp. 1–6, June 2018.
- [4] A. C. Belkina, C. O. Ciccolella, R. Anno, R. Halpert, J. Spidlen, and J. E. Snyder-Cappione, “Automated optimized parameters for t-distributed stochastic neighbor embedding improve visualization and analysis of large datasets,” *Nature Communications*, vol. 10, no. 1, pp. 1–12, 2019.
- [5] A. J. Smola and B. Scholkopf, “A tutorial on support vector regression,” *Statistics and computing*, vol. 14, no. 3, pp. 199–222, 2004
- [6] D. Cerrai, D. W. Wanik, M. A. E. Bhuiyan, X. Zhang, J. Yang, M. E. B. Frediani, and E. N. Anagnostou, “Predicting storm outages through new representations of weather and vegetation,” *IEEE Access*, vol. 7, pp. 29 639–29 654, 2019.
- [7] T. Lawanson, V. Sharma, V. Cecchi, and T. Hong, “Analysis of outage frequency and duration in distribution systems using machine learning,” *2020 52nd North American Power Symposium (NAPS)*, pp. 1–6, 2021.

Thank you!

Q&A