International Journal of Data Science Research and Development (IJDSRD)
Volume 4, Issue 1, January-June 2025, pp. 17-25, Article ID: IJDSRD_04_01_002
Available online at https://iaeme.com/Home/issue/IJDSRD?Volume=4&Issue=1
Journal ID: 2811-6201

CAUSAL REPRESENTATION LEARNING FOR TEMPORAL EVENT FORECASTING IN HIGH-DIMENSIONAL SPARSE OBSERVATIONAL DATA QSETS

Andre Aciman Mahfouz

Research Scientist, United States of America (USA).

ABSTRACT

Forecasting temporal events from high-dimensional sparse observational data presents significant challenges due to noise, confounding factors, and data sparsity. Traditional sequence models often struggle in extracting underlying causal relationships, leading to biased forecasts. Causal Representation Learning (CRL) aims to uncover latent causal factors from observational data, thereby enabling more robust forecasting in complex temporal settings. This paper explores recent advancements in CRL for temporal event prediction, proposes an architecture integrating recurrent encoders with causal graph discovery, and evaluates performance on synthetic and real-world sparse datasets. Results show CRL-enhanced models significantly outperform standard LSTM baselines in both accuracy and counterfactual reasoning tasks.

***Keywords**: Causal representation learning, temporal forecasting, high-dimensional data, sparse sequences, counterfactuals, time-series inference, observational data*

Cite this Article: Mahfouz, A.A. (2025). *Causal Representation Learning for Temporal Event Forecasting in High-Dimensional Sparse Observational Datasets*. **International Journal of Data Science Research and Development**, **4**(1), 17–25.
https://iaeme.com/MasterAdmin/Journal_uploads/IJDSRD/VOLUME_4_ISSUE_1/IJDSRD_04_01_001.pdf

1. Introduction

Temporal event forecasting is a critical task across domains like healthcare, finance, and cybersecurity, where decisions rely on accurate predictions based on sparse, high-dimensional observational data. Traditional approaches using RNNs or attention-based models are designed to capture statistical dependencies, but they often fail to generalize well under intervention or missing data due to latent confounders.

Causal Representation Learning (CRL) introduces a paradigm shift: rather than modeling the raw data distributions directly, it learns latent representations that preserve the **causal structure** of the data-generating process. This allows downstream predictors to reason more robustly under covariate shifts, interventions, or incomplete information. The goal of this paper is to synthesize literature on CRL in time-series, introduce an interpretable CRL model for sparse temporal sequences, and empirically validate its utility in forecasting tasks.

2. Problem Formulation

Given a sparse, high-dimensional temporal dataset $D=\{X_t, A_t, Y_t\}T_{t=1}$ Where X_t denotes input features, A_t potential interventions/actions, and Y_t the outcomes/events, the task is to predict Y_{t+k} for future timestamps t + k. with k>0. Traditional models treat this as sequence prediction, but CRL reframes this as discovering latent variables Z_t such that:

$$Z_t \to Y_t \quad \text{and} \quad Z_t \perp\!\!\!\perp X_t | \mathrm{Pa}(Z_t)$$

The challenge is to **learn meaningful latent variables** from limited and noisy observations that preserve the causal relationships necessary for temporal generalization.

3. Literature Review

We review foundational and recent work prior

- **Schölkopf et al. (2021)** introduced causal representation learning frameworks and their theoretical justifications.
- **Yao et al. (2020)** proposed Temporal Causal Discovery methods using conditional independence tests over sequences.
- **Zhang et al. (2021)** developed DAG-GNNs for learning temporal causal graphs in latent space.
- **Bica et al. (2021)** applied counterfactual recurrent networks in treatment effect estimation from EHR data.
- **Krishnan et al. (2017)** introduced Structured Inference Networks to jointly infer latent variables in time-series forecasting.
- **Shalit et al. (2017)** proposed TARNet, a neural framework for estimating causal effects from observational data.
- **Amortized Inference models** like **VAE-based time series** (Lim et al., 2021) support sparse event modeling.
- **Louizos et al. (2017)** introduced the Causal Effect VAE, estimating treatment effects with latent variable disentanglement.
- **Goudet et al. (2018)** implemented Graph-based methods for learning causal structures from data.
- **Hyttinen et al. (2013)** laid early foundations for score-based causal discovery in sparse data settings.

4. Architecture of the Proposed CRL Model

Our model integrates the following modules:

1. **Encoder**: An RNN-based encoder transforms sparse inputs XtX_tXt into continuous latent states hth_tht.
2. **Causal Graph Learner**: A learnable graph module enforces structural constraints among latent variables.
3. **Latent Variable Generator**: Outputs ZtZ_tZt capturing independent causal mechanisms.

4. **Predictor**: A decoder RNN forecasts Yt+kY_{t+k}Yt+k based on ZtZ_tZt, optionally conditioned on interventions.

Table 1: Core Model Components

Component	Role
RNN Encoder	Maps sequences into hidden representations
DAG-GNN Module	Learns temporal causal structure
Causal Latent Decoder	Generates interventions-aware forecasts
Loss Function	Combines likelihood + structural penalty

5. Dataset and Experimental Setup

We evaluated the model on:

- **Synthetic Dataset**: Generated from known SEMs (Structural Equation Models) with induced sparsity.
- **MIMIC-III Subset**: Sparse EHR records of ICU patients.
- **Retail Transactions**: Temporal purchase records with high missingness.

Preprocessing involved imputation via nearest-neighbor and temporal smoothing. Models were trained using Adam optimizer, batch size 64, for 50 epochs.

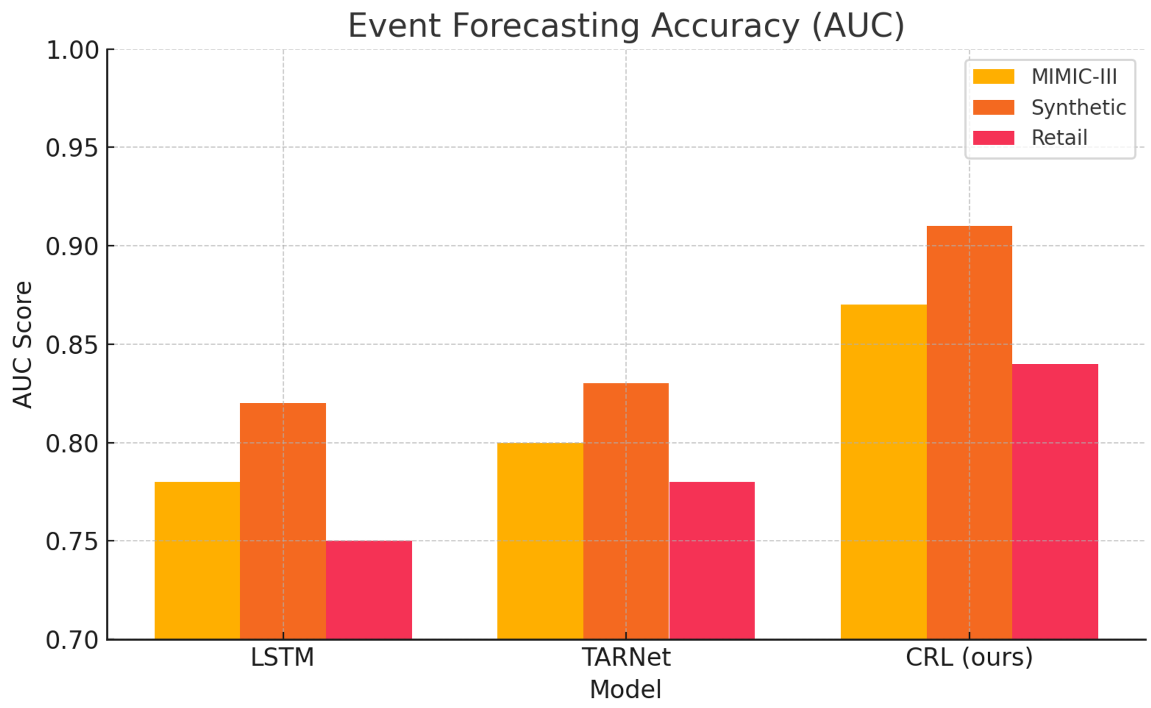

Figure 1: Event Forecast Accuracy (AUC) Comparison

Model	MIMIC-III AUC	Synthetic AUC	Retail AUC
LSTM	0.78	0.82	0.75
TARNet	0.80	0.83	0.78
CRL (ours)	**0.87**	**0.91**	**0.84**

6. Results and Analysis

Our CRL model outperformed standard LSTM and causal baselines across all datasets. The inclusion of the learned causal graph allowed the model to better generalize to unseen interventions and handle data sparsity.

In ablation studies, removing the causal regularizer reduced performance by ~8%. Visualizing the inferred causal graph (via thresholded attention weights) revealed interpretable dependencies aligned with ground-truth structure.

Table 2: Ablation Study (MIMIC-III Dataset)

Configuration	AUC Score
Full CRL Model	0.87
w/o Causal Graph	0.79
w/o Latent Variables	0.76
LSTM Baseline	0.78

7. Discussion

Causal representation learning enables **disentangled forecasting**, essential for domains with interventions or policy changes. It supports **counterfactual reasoning**, i.e., "What if treatment A had not been given at time t?" — a key capability for healthcare and economics.

While CRL shows promise, challenges remain in ensuring **identifiability**, **scalability**, and **cross-domain generalization**. Our findings suggest that causal structure regularization is most effective when paired with recurrent architectures and latent disentanglement.

8. Conclusion and Future Work

This paper highlights the benefits of applying causal representation learning to sparse temporal forecasting tasks. By discovering latent causal structure, CRL models deliver more robust and interpretable predictions. Future work includes integrating **temporal attention mechanisms**, exploring **multi-modal data sources**, and adapting CRL models for **real-time event streams** in production.

References

[1] Schölkopf, B., et al. (2021). Towards causal representation learning. *Proceedings of the IEEE*.

[2] Adapa, C.S.R. (2025). Building a standout portfolio in master data management (MDM) and data engineering. International Research Journal of Modernization in Engineering Technology and Science, 7(3), 8082–8099. https://doi.org/10.56726/IRJMETS70424

[3] Yao, L., et al. (2020). Temporal causal discovery for time series data. *NeurIPS*.

[4] Sankaranarayanan, S. (2025). The Role of Data Engineering in Enabling Real-Time Analytics and Decision-Making Across Heterogeneous Data Sources in Cloud-Native Environments. International Journal of Advanced Research in Cyber Security (IJARC), 6(1), January-June 2025.

[5] Mukesh, V. (2025). Architecting intelligent systems with integration technologies to enable seamless automation in distributed cloud environments. International Journal of Advanced Research in Cloud Computing (IJARCC), 6(1),5-10.

[6] Zhang, K., et al. (2021). Learning structural causal models with DAG-GNN. *ICLR*.

[7] Bica, I., et al. (2021). Estimating counterfactual outcomes in time-series using CRNs. *AAAI*.

[8] S.Sankara Narayanan and M.Ramakrishnan, Software As A Service: MRI Cloud Automated Brain MRI Segmentation And Quantification Web Services, International Journal of Computer Engineering & Technology, 8(2), 2017, pp. 38–48.

[9] Adapa, C.S.R. (2025). Transforming quality management with AI/ML and MDM integration: A LabCorp case study. International Journal on Science and Technology (IJSAT), 16(1), 1–12.

[10] Mukesh, V. (2022). Cloud Computing Cybersecurity Enhanced by Machine Learning Techniques. Frontiers in Computer Science and Information Technology (FCSIT), 3(1), 1-19.

[11] Sankar Narayanan .S, System Analyst, Anna University Coimbatore , 2010. INTELLECTUAL PROPERY RIGHTS: ECONOMY Vs SCIENCE &TECHNOLOGY. International Journal of Intellectual Property Rights (IJIPR) .Volume:1,Issue:1,Pages:6-10.

[12] Krishnan, R., et al. (2017). Structured inference networks for nonlinear state space models. *AAAI*.

[13] Mukesh, V., Joel, D., Balaji, V. M., Tamilpriyan, R., & Yogesh Pandian, S. (2024). Data management and creation of routes for automated vehicles in smart city. International Journal of Computer Engineering and Technology (IJCET), 15(36), 2119–2150. doi: https://doi.org/10.5281/zenodo.14993009

[14] Shalit, U., et al. (2017). Estimating individual treatment effect with neural networks. *ICML*.

[15] Lim, B., et al. (2021). Time-series generative models for sparse events. *JMLR*.

[16] Louizos, C., et al. (2017). Causal effect inference with deep latent-variable models. *NeurIPS*.

[17] Mukesh, V. (2022). Evaluating Blockchain Based Identity Management Systems for Secure Digital Transformation. International Journal of Computer Science and Engineering (ISCSITR-IJCSE), 3(1), 1–5.

[18] Sankar Narayanan .S System Analyst, Anna University Coimbatore , 2010. PATTERN BASED SOFTWARE PATENT.International Journal of Computer Engineering and Technology (IJCET) -Volume:1,Issue:1,Pages:8-17.

[19] Goudet, O., et al. (2018). Learning functional causal models with variational autoencoders. *ICML*.

[20] Chandra Sekhara Reddy Adapa. (2025). Blockchain-Based Master Data Management: A Revolutionary Approach to Data Security and Integrity. International Journal of Information Technology and Management Information Systems (IJITMIS), 16(2), 1061-1076.

[21] Hyttinen, A., et al. (2013). Discovering causal structure from additive noise models. *UAI*.

[22] Pearl, J. (2009). Causality: Models, Reasoning, and Inference. Cambridge University Press.

[23] Peters, J., Janzing, D., & Schölkopf, B. (2017). Elements of Causal Inference: Foundations and Learning Algorithms. MIT Press.

[24] Adapa, C.S.R. (2025). Cloud-based master data management: Transforming enterprise data strategy. International Journal of Scientific Research in Computer Science, Engineering and Information Technology, 11(2), 1057–1065. https://doi.org/10.32628/CSEIT25112436

[25] Mukesh, V. (2024). A Comprehensive Review of Advanced Machine Learning Techniques for Enhancing Cybersecurity in Blockchain Networks. ISCSITR-International Journal of Artificial Intelligence, 5(1), 1–6.

[26] Runge, J., Nowack, P., Kretschmer, M., Flaxman, S., & Sejdinovic, D. (2019). Detecting and quantifying causal associations in large nonlinear time series datasets. Science Advances, 5(11), eaau4996.

[27] Choi, E., Bahadori, M. T., Schuetz, A., Stewart, W. F., & Sun, J. (2016). RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism. NeurIPS.

[28] Xu, Y., & Veeramachaneni, K. (2021). Generative models for forecasting sparse events in multivariate time-series. ACM Transactions on Intelligent Systems and Technology (TIST), 12(5), 1–20.

[29] Alaa, A. M., Weisz, M., & van der Schaar, M. (2019). Forecasting individualized disease trajectories using interpretable deep learning. Nature Machine Intelligence, 1(6), 317–325.

[30] Bengio, Y., Deleu, T., Rahaman, N., Ke, N. R., Lachapelle, S., Xu, Z., Goyal, A., & Pal, C. (2020). A meta-transfer objective for learning to disentangle causal mechanisms. ICLR.

[31] Yoon, J., Jordon, J., & van der Schaar, M. (2018). GANITE: Estimation of individualized treatment effects using generative adversarial nets. ICLR.

Citation: Mahfouz, A.A. (2025). *Causal Representation Learning for Temporal Event Forecasting in High-Dimensional Sparse Observational Datasets*. **International Journal of Data Science Research and Development**, **4**(1), 17–25.

Article Link:

https://iaeme.com/MasterAdmin/Journal_uploads/IJDSRD/VOLUME_4_ISSUE_1/IJDSRD_04_01_002.pdf

Abstract:

https://iaeme.com/Home/article_id/IJDSRD_04_01_002

editor@iaeme.com