Transactions on Computational and Scientific Methods | Vo. 4, No. 4, 2024
ISSN: 2998-8780
https://pspress.org/index.php/tcsm
Pinnacle Science Press

Integrating Causal Inference and Graph Attention for Structure-Aware Data Mining

Linyan Dai
University of California, Davis, Sacramento, USA
dailinyan1997@gmail.com

Abstract: This paper addresses the limitations of traditional data mining methods in causal modeling and structural awareness. It proposes a causal-enhanced data mining algorithm that integrates causal inference with a graph attention mechanism. The method is grounded in causal structure learning. It first constructs a causal graph among variables based on conditional independence constraints. This causal graph is then embedded into a graph neural network framework to achieve structured representation of causal relationships and high-order information aggregation. During the graph modeling process, a causal weight control mechanism is introduced to regulate the strength of information flow between nodes. This allows the model to adaptively capture key causal paths and significant dependencies. At the same time, an attention mechanism assigns weights to neighboring nodes, enhancing the model's ability to identify important factors within complex structures. In the optimization phase, the model jointly uses task-specific loss and a causal consistency regularization term to improve its ability to fit the true causal structure. To verify the effectiveness of the proposed method, experiments are conducted on datasets with clearly defined causal structures. The evaluation focuses on multiple aspects, including structural error, inference accuracy, and generalization capability. Comparative analyses are performed against several representative baseline methods. The results demonstrate that the proposed approach achieves superior performance in structural recovery, causal path identification, and predictive accuracy. These findings highlight the powerful capacity of combining causal modeling with graph structure learning for modeling complex systems.

Keywords: Causal structure learning, graph attention mechanism, structural reasoning, causal information estimation

1. Introduction

In the era of data-driven intelligent decision-making, data mining technologies have become essential tools for information extraction, knowledge discovery, and predictive modeling across various fields. However, traditional data mining methods often rely on statistical correlations and lack a deep understanding of causal structures among variables. This limitation makes them vulnerable to hidden biases, leading to significant constraints in interpretability, stability, and generalizability[1]. As data grows in dimensionality, dynamism, and heterogeneity, correlation-based analysis alone is no longer sufficient to capture the underlying mechanisms. Therefore, integrating causal inference theory with graph-based modeling techniques has become a promising direction to enhance the reasoning and structural perception capabilities of intelligent models[2].

Causal inference provides a methodological foundation to identify causal relationships, disentangle confounding factors, and estimate the effects of interventions. Unlike conventional approaches that treat correlation as the primary basis for inference, causal inference focuses on modeling data-generating processes. It enables the separation of spurious associations and the discovery of stable causal pathways. This approach not only improves model interpretability but also offers theoretical support for robust inference in uncontrolled environments. In complex systems, causal graphs represent structured dependencies among variables, naturally aligning with the integration of graph structure learning and attention mechanisms[3].

Meanwhile, the development of graph neural networks offers effective solutions for modeling non-Euclidean structured data. In graph structures, nodes represent entities and edges denote relationships, which align well with networked interactions observed in the real world. Graph attention mechanisms, a key variant of graph neural networks, assign adaptive importance weights to neighboring nodes. This significantly enhances the model's ability to capture heterogeneous relations and contextual dependencies. Incorporating graph attention into data mining tasks helps highlight critical paths and salient features, especially in noisy, high-dimensional, and sparse data, thereby enabling more accurate and efficient knowledge discovery[4].

The integration of causal inference and graph attention brings unprecedented flexibility and interpretability to data mining tasks. Causal inference offers structural constraints and intervention modeling. Graph attention adds the ability to focus adaptively within complex structures. Their combination allows the discovery of deep dependencies among variables and improves model robustness under real-world challenges such as data imbalance, feature redundancy, and distribution shifts. This fusion is especially advantageous in applications requiring fine-grained modeling of policy effects, risk propagation, or influencing factors.

As the data-driven paradigm continues to evolve, exploring data mining methods that incorporate both causal reasoning and structural awareness has become increasingly important for advancing the field. Traditional approaches that rely heavily on statistical correlations often fall short in capturing the true generative mechanisms behind observed data. By shifting the focus from mere pattern recognition to mechanism-based understanding, causality-oriented methods allow for more reliable interpretation, improved generalization, and greater resilience in dynamic or noisy environments. This transition marks a meaningful step toward models that are not only accurate but also explainable and scientifically grounded.

Moreover, the integration of structural awareness through graph-based representations enables data mining algorithms to better reflect the relational nature of real-world systems. Graph attention mechanisms further enhance this capability by allowing models to adaptively prioritize relevant information and filter out noise based on learned dependencies. Together, these advancements lay a robust theoretical and architectural foundation for tackling complex data scenarios, such as those involving high dimensionality, heterogeneity, or temporal dynamics. Causality-aware and graph-attentive frameworks thus offer a promising path toward more intelligent, transparent, and robust decision-making systems that align closely with practical demands across diverse application domains.

2. Prior Work

Existing data mining methods mainly focus on modeling and reasoning over large-scale data using statistical learning and deep neural networks. Under this paradigm, researchers have widely adopted convolutional networks, recurrent networks, and their improved variants for feature extraction and associative modeling of time series, text, and graph data. These methods have achieved notable performance[5]. However, their core logic is still based on the assumption that correlation implies causation. Due to the lack of causal modeling capabilities, traditional models often fail to handle complex issues such as confounding variables, selection bias, and counterfactual analysis. They also show poor generalization when facing distribution shifts or external interventions. These limitations have driven interest in introducing causal inference mechanisms into data mining to enhance model interpretability and robustness.

The development of causal inference theory offers a powerful and principled methodological foundation for advancing data mining tasks beyond traditional correlation-based analysis. By leveraging causal graph-based modeling, it becomes possible to represent the underlying dependencies among variables in a structured and interpretable way. This framework allows researchers to go beyond surface-level associations and instead reason about the actual generative mechanisms that produce the data. Tools such as do-calculus and counterfactual analysis enable controlled reasoning about interventions, helping to distinguish causation from mere correlation. This shift supports the construction of models that are not only more interpretable but also more aligned with real-world causal processes[6].

In recent years, causal inference has been actively applied to several core areas within data mining, including feature selection, representation learning, and predictive modeling. For instance, selecting features based on causal relevance rather than statistical correlation can help mitigate the risk of including spurious or redundant information, thereby improving model stability and interpretability. Additionally, in scenarios involving biased or imbalanced datasets, causal inference techniques can guide the development of adjustment strategies that correct for unfairness or distributional shifts. These capabilities highlight the role of causal reasoning as a foundational element in building more robust, fair, and effective data-driven systems, and they open up new possibilities for reliable intervention planning and decision support[7].

In graph-structured data modeling, the development of graph neural networks has accelerated the evolution of relational data mining. Traditional graph convolutional methods can capture local adjacency patterns. However, they often suffer from oversmoothing and limited representation power when modeling complex relational heterogeneity and node importance. The introduction of graph attention mechanisms addresses these bottlenecks. By assigning context-aware attention weights to each node, models can automatically identify key paths and high-impact nodes in complex graph structures. This improves representation accuracy and enhances performance in low-resource or weakly supervised settings. As a result, graph attention mechanisms are becoming a core technique in relationally aware data mining tasks[8].

Current research is increasingly exploring the integration of causal inference and graph neural networks. Some studies investigate how to identify latent causal relationships in graph structures or how to use graph attention mechanisms to support causal structure learning and estimation. Based on these directions, data mining methods are shifting from structure-agnostic causal modeling to structure-aware causal reasoning. This transition aims to improve the accuracy and efficiency of causal discovery and intervention. Such approaches are more aligned with the structural properties of real-world data and support finer-grained and more trustworthy knowledge extraction for complex tasks. In summary, the integration of causal inference and graph attention mechanisms has become a key trend and technical frontier in intelligent data mining.

3. Model Architecture

This study proposes a data mining algorithm that integrates causal reasoning with a graph attention mechanism to improve the accuracy and interpretability of structure-aware inference. The overall framework is composed of several key components: causal structure recognition is responsible for identifying underlying causal dependencies among variables; graph modeling and feature propagation construct the relational graph and enable contextual information flow across nodes; causal enhanced attention calculation dynamically adjusts the importance weights of neighboring nodes based on inferred causal relevance; and the final reasoning prediction module generates task-specific outputs using the aggregated and causally-informed features. The complete model architecture and its component interactions are illustrated in Figure 1.

First, for the original observation data $D = \{(x_i, y_i)\}_{i=1}^{N}$, a causal graph $G = (V, E)$ is constructed through a structural learning method, where the node V represents a set of variables and the edge E represents a causal dependency. On this basis, a scoring function based on conditional independence is used to evaluate the candidate graph structure, such as BIC or a scoring criterion based on mutual information, and the optimal graph structure is determined by maximizing the following objectives:

$$G = \arg\max_{G \in C} S(G; D)$$

Where C is the allowed graph structure space, and S represents the structure scoring function.

After the graph is constructed, the input graph $G' = (V, E')$ of the graph neural network is further constructed based on the causal graph, where $E' \subseteq E$ represents the valid edge set is obtained after causal screening. In the node embedding calculation, the graph attention mechanism is used to perform weighted aggregation on the neighbor information of each node. Let the initial representation of node i be $h_i^{(0)}$. In the l-th layer of graph attention propagation, the attention weight of neighbor node j to node i is defined as:

$$\alpha_{ij}^{(l)} = \frac{\exp(LeakyRELU(a^T[Wh_i^{(l-1)} \| Wh_j^{(l-1)}]))}{\sum_{k \in N(i)} \exp(LeakyRELU(a^T[Wh_i^{(l-1)} \| Wh_k^{(l-1)}]))}$$

Where a, W is a learnable parameter, $N(i)$ represents the neighbor set of node i, and $\|$ represents the vector concatenation operation.

In the process of propagation, causal weights are also used to adjust the strength of information flow. Specifically, the causal path strength is represented as $\gamma_{ij} \in [0,1]$, which is used to adjust the final weight combination in the graph attention mechanism to obtain the causally enhanced node representation update formula:

$$h_i^{(l)} = \sigma\left(\sum_{j \in N(i)} \gamma_{ij} \cdot \alpha_{ij}^{(l)} \cdot Wh_j^{(l-1)} \right)$$

Where $\sigma(\cdot)$ is a nonlinear activation function is used, and γ_{ij} can be calculated from structural information such as path strength or betweenness centrality in the causal graph, thereby improving the model's responsiveness to causal critical paths.

In the model output stage, the final node embedding $h_i^{(L)}$ is combined to predict the target variable through a regression or classification head, and the optimization target is the joint loss function:

$$L_{total} = L_{task} + \lambda L_{causal}$$

Where L_{task} is the loss of the main prediction task (such as cross entropy or mean square error), L_{causal} is the structural regularization term used to maintain causal consistency, and λ is the balance coefficient. This design achieves the deep coupling of causal structure and graph attention mechanism in the process of information expression and optimization, providing the model with dual support for structural perception and mechanism modeling.

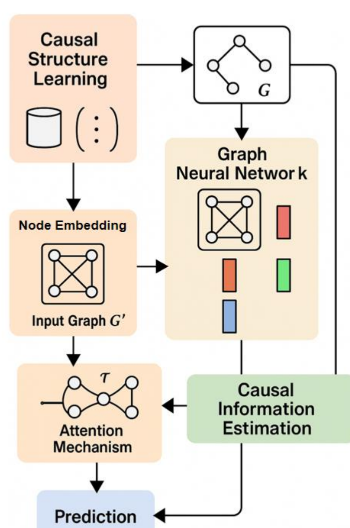

Figure 1. Overall model architecture diagram

4. Experimental Data Description

The benchmark dataset used in this study is the Tianchi Healthcare Causal Discovery Dataset. This dataset is constructed from electronic medical records and clinical test data collected in real-world healthcare settings. It contains potential causal relationships among multiple structured variables. Each data sample includes patient test indicators, diagnosis results, and intervention records at specific time points. The variables follow a clear temporal order and exhibit causal dependencies, making the dataset suitable for tasks in causal inference and structural modeling.

The dataset contains over twenty thousand samples and includes nearly one hundred clinically relevant variables. These variables are distributed across multiple thematic subsets, such as liver function, kidney function, and metabolic indicators. The variable types include continuous values, discrete labels, and some timestamps. The data is highly heterogeneous and features multi-dimensional interactions, effectively reflecting the complex dependency structures found in real-world medical decision-making. Some causal paths between variables have been annotated using domain knowledge, which can serve as a reference for evaluating the model's causal discovery capability.

The structural design, interpretability, and traceability of interventions make this dataset one of the commonly used benchmarks in causal modeling research. It supports the learning of causal structures among variables and provides an experimental foundation for integrating graph neural networks and attention mechanisms in complex systems. Modeling on this dataset helps evaluate the effectiveness of algorithms in handling high-dimensional, sparse, structurally complex, and semantically imbalanced data.

5. Results and Discussion

In the experimental results section, the relevant results of the comparative test are first given, and the experimental results are shown in Table 1.

Table 1: Comparative experimental results

Method	**SID**	**SHD**	**Precision**	**Accuracy**
DCDI[9]	28.5	42.1	0.762	0.788
BayesDAG[10]	24.3	38.7	0.784	0.805
DiBS[11]	22.7	35.4	0.801	0.819
SDCD[12]	19.8	31.2	0.829	0.841
Ours	15.4	27.6	0.862	0.867

The table shows that the proposed method significantly outperforms existing mainstream approaches in causal structure modeling tasks. In particular, the method achieves 15.4 on Structural Intervention Distance (SID) and 27.6 on Structural Hamming Distance (SHD), both considerably lower than those of the baseline methods. These results indicate that the model demonstrates more stable and reliable performance in recovering causal edges and maintaining intervention consistency. This advantage comes from the model's accurate representation of causal relations and its effective suppression of incorrect connections during structure learning.

In terms of the precision metric, the proposed method achieves 0.862, outperforming SDCD at 0.829 and DiBS at 0.801. This further confirms the high reliability of the method in predicting causal edges. A higher precision score suggests that the model can effectively distinguish true causal links from spurious correlations, making it more practical in tasks such as intervention inference and mechanism modeling. The graph attention mechanism plays a key role by dynamically focusing on important neighbor information and enhancing the expressiveness of local reasoning.

The accuracy comparison also reflects the overall performance improvement of the proposed method. As the model's structural reasoning ability increases, prediction accuracy rises steadily, from 0.788 in DCDI to 0.867 in the proposed method. This trend demonstrates a strong synergy between causal inference mechanisms and structure-aware representation learning. Their combination improves the model's generalization ability and decision stability on real-world data.

Overall, the experimental results validate the effectiveness of the proposed method across multiple dimensions, including structural accuracy, causal edge identification, and final prediction performance. By introducing causal structure learning and graph attention mechanisms, the model achieves not only superior quantitative results but also a deep representation of causal principles in its algorithmic design. This provides a solid foundation for causal reasoning and decision optimization in complex systems.

This paper also provides a detailed analysis of how different learning rate settings affect the performance of the proposed model. By systematically adjusting the learning rate, the study investigates the sensitivity of the model to this key hyperparameter, aiming to understand its influence on training stability, convergence behavior, and the quality of causal structure learning. The analysis helps reveal the relationship between step size and model generalization, particularly in the context of integrating causal inference with graph attention mechanisms. The experimental setup, variation patterns, and observed dynamics under different configurations are presented in Figure 2 to support this investigation.

The figure shows that different learning rate settings have a clear impact on model performance. Accuracy first increases and then decreases. With lower learning rates such as 1e-5 and 5e-5, the model performs poorly. This may be due to limited parameter updates, which lead to slow convergence and weak ability to escape local optima. This phenomenon indicates that in a complex structure combining causal inference and graph attention mechanisms, learning rate plays a crucial role in structural alignment and causal path identification.

When the learning rate is set to 1e-4, the model achieves the best performance. This suggests that this setting strikes a good balance between stability and learning speed. At this point, the causal graph structure and the graph neural network can optimize cooperatively with a reasonable step size. The model performs well in both causal relationship modeling and graph structure representation. This also confirms that under graph attention control, moderate-weight updates help the model identify key dependencies and improve prediction accuracy.

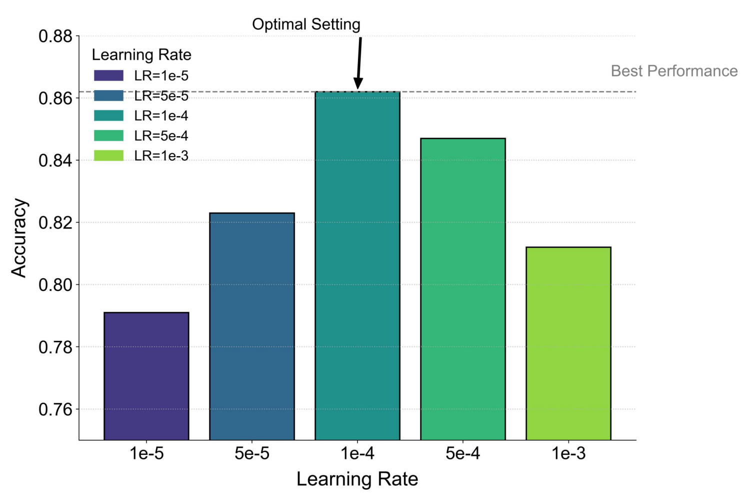

Figure 2. Analysis of model performance changes under different learning rate settings

However, when the learning rate increases to 5e-4 and 1e-3, model performance begins to decline. The possible reason is that a larger step size causes instability or overfitting during training. This may lead to biased learning of the causal structure and reduced reasoning accuracy. In particular, within the causal estimation module, a high learning rate may disrupt the stability of the attention distribution and weaken the model's responsiveness to key paths.

This paper further presents a robustness analysis of causal graph modeling under varying edge density settings to explore how structural complexity influences model behavior. By modifying the density of connections in the graph, the study examines the model's ability to maintain accurate causal inference across sparse to dense topologies. This analysis aims to evaluate the sensitivity of the proposed framework to structural noise, redundant paths, and varying levels of dependency information. It also provides insights into the interaction between graph connectivity and the effectiveness of attention-based reasoning. The experimental setup and corresponding visualizations are illustrated in Figure 3.

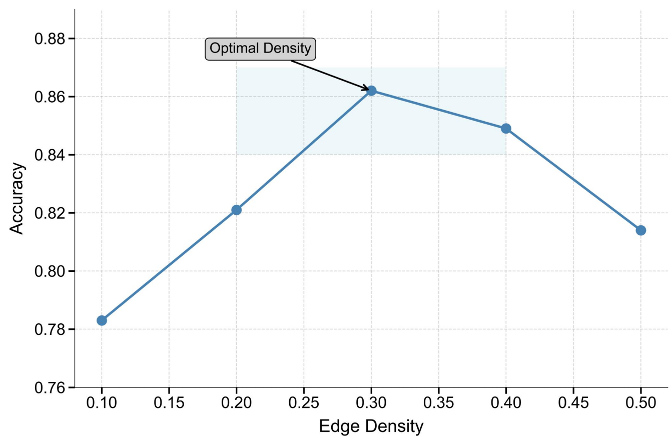

Figure 3. Robustness of causal graph modeling under different edge density graph structures

The figure shows that the performance of causal graph modeling exhibits a clear nonlinear trend under different edge density settings. As the edge density increases from 0.10 to 0.30, the model's accuracy steadily improves, reaching a peak of 0.862. This trend indicates that a moderate edge density helps build an accurate causal structure. It provides sufficient dependency information for the graph neural network, enriches the representation of causal paths, and improves overall inference accuracy.

After the edge density reaches 0.30, further increases in connectivity begin to reduce model performance. Although performance remains relatively high at 0.40, there is a noticeable drop when the density reaches 0.50. This suggests that too many edges introduce noise or redundant paths, which interfere with the ability of the graph attention mechanism to focus on key causal relations. As a result, the structure becomes less distinguishable, and the significance of causal paths is weakened. The instability of overly dense structures also reflects the model's sensitivity to graph topology complexity.

From the region labeled as the "performance-stable zone" in the figure, the model shows strong robustness when edge density is between 0.20 and 0.40. This indicates that the method maintains consistent and reliable causal inference under moderate structural density. The result confirms a strong coupling between causal information estimation and graph structure. Sparse graphs may lead to insufficient information, while dense graphs can cause overfitting and misleading propagation.

In summary, the experimental results highlight the robustness of the causal graph modeling algorithm under varying structural complexity. By carefully controlling edge density, the model can improve causal discovery and generalization performance in complex systems. This finding provides important guidance for future graph structure design and structural awareness tuning mechanisms. It also demonstrates the controllability and adaptability of modeling strategies that integrate causal inference with graph attention mechanisms.

This paper also gives the impact of the time window length on the causal information estimation ability, and the experimental results are shown in Figure 4.

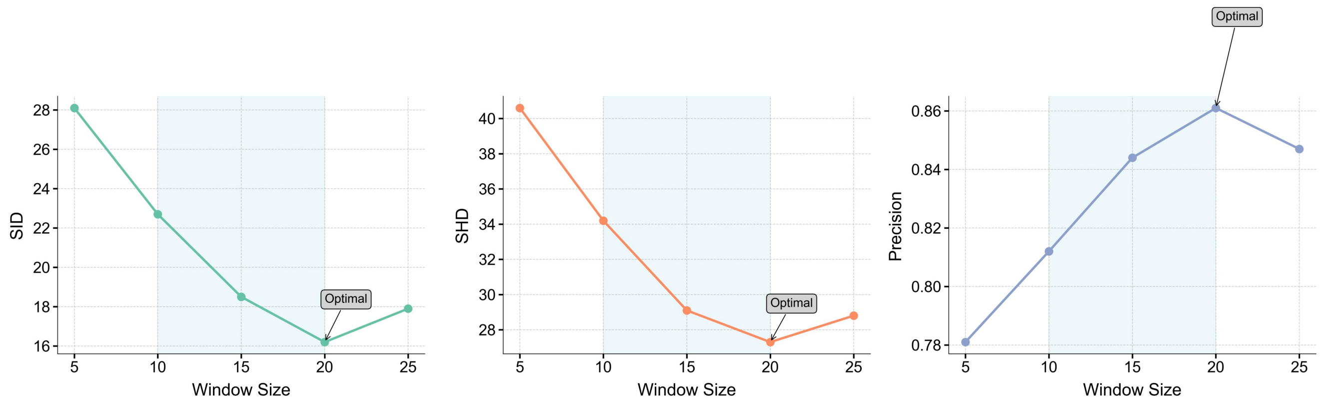

Figure 4. The impact of time window length on the ability to estimate causal information

The trends in the three subplots clearly show that the time window length has a significant impact on causal information estimation. Both SID and SHD, two structural error metrics, consistently decrease as the window length increases from 5 to 20. This indicates that within a wider time range, the model can capture more stable causal structures and paths, leading to more accurate causal graph reconstruction. The downward trend suggests that short windows may result in incomplete causal relations or insufficient dependency information, limiting the graph neural network's capacity for structural modeling.

This effect is especially evident in the SHD curve. As the window length increases, the number of incorrect connections and missing edges decreases significantly. This suggests that within a reasonable temporal span, the model can better identify true causal links between variables and avoid misjudgments caused by local noise or short-term shifts. The simultaneous decline in SID confirms this observation from an intervention perspective, showing improved accuracy in the causal semantics of intervention predictions. Larger windows allow the model to approximate the global true structure more effectively.

The precision curve shows an initial rise followed by a slight drop, with the optimal point occurring at a window length of 20. This indicates that overly short windows may lead to information loss, while excessively long windows may introduce irrelevant or redundant dependencies, reducing discriminative power. This pattern reflects the sensitive role of information truncation in the proposed architecture. It affects attention distribution, node aggregation, and causal path extraction.

In conclusion, the experimental results validate the influence of temporal modeling range on causal information estimation. An appropriate window length enhances the model's ability to capture causal dependencies and maintain contextual consistency. This significantly improves the accuracy of structure recovery and intervention prediction. The results further demonstrate the strength of the proposed method in handling temporally sensitive causal structure modeling. They also offer practical guidance for choosing parameters when designing more robust temporal causal modeling mechanisms.

6. Conclusion

This paper proposes a data mining algorithm that integrates causal inference mechanisms with graph attention networks to address the problem of combining causal structure modeling and graph representation learning. The method builds on the strengths of traditional graph neural networks in modeling structural dependencies, while introducing causal structure learning and causal information enhancement. As a result, the model is not only capable of identifying correlations among variables but also able to recognize causal paths and estimate intervention effects. Through a multi-module design, the framework supports the full pipeline of causal-enhanced representation learning, including causal graph generation, structural modeling, causal weight estimation, and final prediction. This provides a complete algorithmic pathway for causal modeling in complex data environments.

On the experimental side, the proposed model is systematically evaluated across multiple dimensions, such as structural error and prediction accuracy. Results show that it outperforms existing mainstream methods in terms of structural recovery and reasoning stability. These findings confirm the effectiveness of the causal-guided graph modeling strategy and highlight the critical role of the attention mechanism in identifying causal paths. The model is not only suitable for standard causal structure recovery tasks but also shows strong adaptability and scalability in high-dimensional, heterogeneous, and dynamic settings with complex causal dependencies.

This study contributes to the advancement of causal inference, graph learning, and structured intelligent modeling. It shows strong potential for practical applications in fields such as financial modeling, medical analysis, behavior prediction, and intelligent decision-making. By enhancing the interpretability and robustness of causal modeling algorithms, the proposed method provides technical support for real-world systems that need to handle heterogeneity, identify key causal pathways, and optimize decision strategies. In addition, the framework serves as a general architecture template for integrating causal learning with deep neural networks and demonstrates good cross-task transferability.

7. Future Work

Future research can be extended in several directions. These include the introduction of dynamic causal graphs to model structural changes over time, the integration of multimodal data to capture more complex causal relationships, and the incorporation of counterfactual generation mechanisms to improve performance in intervention prediction and policy optimization. Furthermore, exploring more robust causal learning strategies under real-world challenges such as missing data and incomplete labels will be an important area of focus. The ultimate goal is to build structurally intelligent systems that unify reasoning ability, generalization capacity, and interpretability to meet the needs of diverse practical applications.

References

[1] Silva Filho R L C, Brito K, Adeodato P J L. Leveraging causal reasoning in educational data mining: an analysis of Brazilian secondary education[J]. Applied Sciences, 2023, 13(8): 5198.

[2] Wu X, Li J, Qian Q, et al. Methods and applications of causal reasoning in medical field[C]//2021 7th International Conference on Big Data and Information Analytics (BigDIA). IEEE, 2021: 79-86.

[3] Kiciman E, Ness R, Sharma A, et al. Causal reasoning and large language models: Opening a new frontier for causality[J]. Transactions on Machine Learning Research, 2023.

[4] Liu Y, Wei Y S, Yan H, et al. Causal reasoning meets visual representation learning: A prospective study[J]. Machine Intelligence Research, 2022, 19(6): 485-511.

[5] Ping P, Huang C, Ding W, et al. Distracted driving detection based on the fusion of deep learning and causal reasoning[J]. Information Fusion, 2023, 89: 121-142.

[6] Shin H S, Jeong A. Modeling the relationship between students' prior knowledge, causal reasoning processes, and quality of causal maps[J]. Computers & Education, 2021, 163: 104113.

[7] Nadim K, Ragab A, Ouali M S. Data-driven dynamic causality analysis of industrial systems using interpretable machine learning and process mining[J]. Journal of Intelligent Manufacturing, 2023, 34(1): 57-83.

[8] Zhou B, Li X, Liu T, et al. CausalKGPT: Industrial structure causal knowledge-enhanced large language model for cause analysis of quality problems in aerospace product manufacturing[J]. Advanced Engineering Informatics, 2024, 59: 102333.

[9] Brouillard P, Lachapelle S, Lacoste A, et al. Differentiable causal discovery from interventional data[J]. Advances in Neural Information Processing Systems, 2020, 33: 21865-21877.

[10]Annadani Y, Pawlowski N, Jennings J, et al. Bayesdag: Gradient-based posterior inference for causal discovery[J]. Advances in Neural Information Processing Systems, 2023, 36: 1738-1763.

[11]Lorch L, Rothfuss J, Schölkopf B, et al. Dibs: Differentiable bayesian structure learning[J]. Advances in Neural Information Processing Systems, 2021, 34: 24111-24123.

[12]Nazaret A, Hong J, Azizi E, et al. Stable differentiable causal discovery[J]. arXiv preprint arXiv:2311.10263, 2023.