

The Message Passing Mechanism of Graph Neural Networks and Its Potential for Processing Structured Data

Xinzheng Wu

Monash Business School, Monash University, Melbourne, Australia

Abstract: Graph neural networks (GNNs) achieve deep representation learning of graph-structured data through a message passing mechanism, becoming a core technical paradigm for processing structured data. This paper systematically analyzes the architectural principles of the message passing mechanism, illustrating its three-step core process of "message generation-aggregation-update" and its mathematical implementation. The paper focuses on comparing the mechanism with typical models such as GCN and GAT. Drawing on empirical examples from fields such as bioinformatics, financial risk control, and retail supply chain, the paper reveals the unique advantages of this mechanism in capturing entity associations and discovering hidden patterns. In protein function prediction, the message passing-based DeepFRI model achieves accurate mapping between structure and function; in financial fraud detection, aggregating multi-hop transaction information significantly improves the ability to identify hidden criminal networks. The paper also analyzes current challenges such as oversmoothing and large-scale graph adaptation, proposing solutions such as residual connection optimization and neighbor sampling. This research demonstrates that the message passing mechanism provides a modeling approach that captures the relational nature of structured data, demonstrating strong potential for application in complex, cross-domain problems. Future applications will be further overcome by combining self-supervised learning with large-scale model technologies.

Keywords: Graph neural networks; Message passing mechanism; Structured data; Node embedding; Aggregation operator.

1. Introduction

In the era of big data, structured data has become the core carrier of critical information. From the spatial connections of biomolecules to the transaction networks of financial accounts, from user interactions on social platforms to product associations in retail systems, this type of data generally exhibits a complex entity-relationship topology. Traditional deep learning models such as CNNs and RNNs, constrained by the Euclidean space assumption, face inherent limitations when processing such irregularly structured data. CNNs' fixed convolution kernels struggle to adapt to flexible node connections, while RNNs' sequential dependency modeling loses critical multi-directional correlation information. Forced mapping also results in loss of edge semantics and waste of computational resources. The emergence of graph neural networks overcomes this limitation. Their core advantage lies in their natural adaptability to graph structures, and the message passing mechanism is the fundamental building block of GNNs for mining structural information. Just as convolution operations are used in CNNs and temporal recursion in RNNs, message passing simulates the information exchange process between nodes, allowing each node to integrate features from its neighbors, thereby generating a representation vector that combines local details with global connections. This mechanism essentially transforms implicit relationships in a graph into learnable, explicit features, offering a novel approach to solving complex modeling problems with structured data. Existing research has demonstrated the application value of the message passing mechanism in various fields: in bioinformatics, it can decode the amino acid connectivity patterns of proteins to predict function; in finance, it can identify money laundering networks by

aggregating transaction chain information; and in retail, it can optimize inventory forecasting by leveraging product co-occurrence relationships. This article systematically analyzes the working principles and application potential of the message passing mechanism from five perspectives: theoretical applicability, core architecture, variant optimization, empirical applications, and challenges. The discussion is based on real-world models and case studies, providing a technical reference for structured data modeling.

2. Theoretical Applicability of Graph Neural Networks for Structured Data

Structured data is essentially a collection of entities organized by explicit relational rules. Its core characteristic is the synergy between entity independence and relational dependency. From a mathematical perspective, this type of data can be naturally mapped into a graph structure $G = (V, E, X)$: nodes V correspond to entities, edges E represent entity relationships, and the matrix X records node attributes. For example, in financial transaction data, accounts are nodes, transaction behaviors are edges, and account balances and transaction amounts are node attributes. In protein structure data, amino acids are nodes, peptide bonds are edges, and chemical properties are attribute features. Traditional data processing methods have significant limitations in modeling this type of data. Methods such as SVM based on statistical learning can only process independent samples and cannot capture information about the associations between entities. While classic deep learning models can extract complex features, they have shortcomings in structural adaptation: when CNN processes molecular graphs, it must forcibly grid irregular structures, resulting in the loss of key edge

information such as bond types. When LSTM analyzes social networks, it has difficulty processing multi-directional interactions in parallel, and long-distance dependency transmission is prone to information attenuation. These shortcomings have limited the effectiveness of traditional methods in mining deep relationships in structured data. The design philosophy of graph neural networks is highly aligned with the essential characteristics of structured data. Their adaptability is primarily reflected in three aspects: First, topological consistency [1]. GNNs directly take graphs as input and can preserve the connections between entities without structural transformation. For example, in knowledge graph completion tasks, entity embeddings can be learned directly based on triple structures. Second, relationship modeling flexibility. Edge attribute encoding allows for adaptability to diverse types of relationships. Whether it's transaction amounts in weighted graphs or differences in entity types in heterogeneous graphs, both can be incorporated into feature learning through message passing. Third, feature aggregation is hierarchical. By stacking multiple layers of message passing units, nodes can gradually integrate multi-hop neighbor information, achieving structural awareness from local to global scale. This aligns with the "direct-indirect" hierarchical nature of structured data. This inherent adaptability allows GNNs to break through the bottlenecks of traditional methods and provide an end-to-end solution for structured data, from representation learning to task solving. The message passing mechanism is the core technical path that enables this adaptability.

3. Core Architecture and Mathematical Implementation of the Message Passing Mechanism

The message passing mechanism draws on the logic of information exchange in human society: individuals exchange information with surrounding objects to update their cognition and thereby understand the overall environment. In GNNs, this process is abstracted into a standardized three-stage process: "message generation - information aggregation - state update." Core elements include node features, neighbor sets, edge attributes, and three key operations: message generation, aggregation, and update. Its permutation invariance, receptive field scalability, and differentiability provide theoretical support for model training and structural feature mining. Message generation is the starting point of the mechanism, and its core is to transform neighbor features into transferable "information units." The strategies employed by different models vary significantly. GCN uses linear mapping to process neighbor features and mitigates feature bias caused by node degree differences through degree matrix normalization, balancing efficiency and adaptability to small and medium-sized graphs. GAT introduces self-attention, allowing the model to autonomously calculate neighbor association weights to focus on important nodes. It typically uses 8 or 16 attention groups to enhance representations and requires sparsification to manage video memory. MPNN is specifically designed for data with edge attributes, such as molecular graphs, and can incorporate edge information such as chemical bond types into messages, making it suitable for scenarios with attributed edges. Information aggregation requires integrating all neighbor messages, with the core requirement being permutation invariance [2]. Common aggregation operators

fall into two categories: basic statistical operators offer simple computations. Mean aggregation, for example, can reduce interference from nodes with extreme degrees when node degrees vary significantly, resulting in more stable performance. Among advanced learning operators, LSTM aggregation relies on gating to capture sequential dependencies, making it suitable for dynamic graph time series scenarios. Attention aggregation uses weights to filter key messages and strengthen important relationships. GIN uses a combination of summation and multi-layer perceptrons, theoretically approaching the upper bound of the graph isomorphism test, providing an important reference for operator design. Node updates require integrating aggregated information with native features to generate new representations. The most commonly used approach is to concatenate aggregated information with native features and then update them through a nonlinear transformation. GraphSAGE and GIN both adopt this design. To mitigate information attenuation in deep models, residual connections reuse the original node features by adding the aggregated results. DeepGCN overcomes the limitations of traditional GNN layers in this way. Gated updates can autonomously control the flow of information, excelling in processing time-series structured data.

4. Typical Variants and Adaptive Optimization of Message Passing Mechanisms

Structured data types are diverse, and a single message passing model cannot be suitable for all scenarios. Researchers have proposed mechanism variations for specialized types such as heterogeneous graphs, dynamic graphs, very large graphs, and high-dimensional sparse graphs. The core approach is to adjust the granularity, timing, and scope of message passing based on data characteristics, forming a multi-dimensional technical system. The core challenge of heterogeneous graphs is the diverse types of nodes and edges, making it difficult to handle diverse semantic associations. RGCN designs an independent linear transformation matrix for each relationship to achieve relationship-aware message generation. In the knowledge graph completion task, it significantly improves link prediction accuracy by modeling multiple semantic relationships. In more complex heterogeneous scenarios, HetGNN adopts a "type-specific aggregation" strategy, first aggregating messages for different types of neighbors and then fusing them, successfully classifying authors' research directions in academic networks. In dynamic graphs, the relationships between nodes and edges change over time, requiring message passing to be temporally adaptable. TGN introduces a memory module to store historical messages, designs time-aware attention weights, and incorporates timestamp features into message updates to effectively capture short-term fluctuations in traffic flow prediction. GGNN uses GRU gated units to control the retention and discarding of temporal messages [3]. In video scene graph generation, it successfully models the dynamic changes in the interactive relationships between objects. Ultra-large-scale graphs face the dual pressures of computational complexity and video memory usage, and conventional full-graph aggregation is inefficient. GraphSAGE proposed an inductive learning framework that reduces complexity from $O(N)$ to $O(K)$ (K is the sampling size) by randomly sampling a fixed

number of neighbors instead of full aggregation. This framework has been applied to user embedding learning in large-scale recommendation systems. PinSAGE has been further optimized, using importance sampling to select highly influential neighbors and combining graph partitioning technology to achieve distributed training, supporting Pinterest's recommendation of hundreds of millions of items. In high-dimensional, sparsely structured data, message passing is prone to losing feature information due to "over-compression." D-MPNN, a molecular prediction framework, uses bilinear transformations to expand the message dimension and design separate message generation and update networks. In QM9 molecular property prediction, it reduces the prediction error of fine-grained properties such as atomic energy by 12% compared to traditional MPNNs. DeepFRI, a protein function prediction model, integrates evolutionary and structural information, implements message passing through multi-layer convolutions, and then aggregates multi-scale features through attention-weighted aggregation. The model achieves an F1 score of 87% for enzyme function prediction, far exceeding traditional sequence alignment methods.

5. Empirical Application of the Message Passing Mechanism in Structured Data

The application value of the message passing mechanism lies in its ability to model complex relationships and its adaptability across scenarios. It has been implemented in fields such as bioinformatics, financial risk control, and retail supply chain. The following analysis uses real-world examples to analyze its practical logic and effectiveness. In bioinformatics, structural modeling of molecules and proteins is a core challenge, and the message passing mechanism provides a breakthrough solution. Molecules can be modeled as "atom-chemical bond" graphs [4]. MPNN accurately predicts 13 physicochemical properties on the QM9 dataset, with a prediction error of less than 0.1 eV for the HOMO-LUMO energy gap, making it a tool for virtual screening of drug molecules. For protein function prediction, DeepFRI uses a GCN architecture to convert protein three-dimensional structures into graphs, aggregating spatial information about amino acid residues. This allows for inferring function before DNA sequencing, facilitating the detection of unknown pathogens. Compared to traditional BLAST methods, its prediction recall rate is 35% higher and has been used for functional annotation of the human proteome. Financial risk control transaction networks are heterogeneous structured data, and the message passing mechanism can mine hidden associations. Traditional methods focus on single transactions and ignore inter-account connections. The GAT model, however, constructs an account-transaction graph and uses attention to aggregate multi-hop features to identify money laundering networks. After deployment at a state-owned bank, the false positive rate for suspicious transactions dropped from 18% to 5.2%, and the bank identified three money laundering rings with hundreds of accounts. In credit assessment, the RGCN model, which models a ternary network of "enterprise-guarantee-bank," improved the accuracy of default prediction for small and medium-sized enterprises by 21% and reduced bad debt rates. In the retail supply chain, product association networks are crucial, and message passing mechanisms optimize inventory and

recommendations. Traditional ARIMA and LSTM methods rely solely on individual product sales, ignoring complementarity/substitution relationships. Amazon uses GraphSAGE to construct a product graph and integrate sales of related products, resulting in an 18% increase in fresh produce inventory turnover and a 24% reduction in out-of-stock rates. Walmart builds a heterogeneous "user-product-category" graph and uses GAT to distinguish the importance of associations. This results in a 30% increase in click-through rate for personalized recommendations compared to traditional collaborative filtering, achieving an upgrade from "single-point prediction" to "network collaboration." Message passing mechanisms are also highly effective in spatiotemporal structured data such as transportation and social networks. In traffic flow prediction, ST-GNN transforms road networks into spatiotemporal graphs, incorporating time-series features into message passing to capture traffic flow patterns. Application in one city achieved an 89% accuracy rate for morning rush hour congestion predictions and a 22% improvement in route planning efficiency. In social network analysis, GIN uses deep message passing to detect communities, identifying hidden interest groups within the Facebook social graph. Its modularity index reaches 0.83, providing support for precision marketing [5].

6. Existing Challenges and Optimization Paths for the Message Passing Mechanism

While the message passing mechanism holds significant potential for structured data processing, it still faces key theoretical and engineering challenges when dealing with complex, large-scale, and highly dynamic data. Identifying these challenges and exploring optimization directions are key to its further development. Oversmoothing is a core theoretical bottleneck: as the number of network layers increases, node features gradually converge and lose their differentiation. The root cause is feature averaging caused by multiple message aggregations, which is more pronounced in dense graphs. When the number of GCN layers exceeds 6, the cosine similarity of node embeddings often exceeds 0.9, resulting in a performance drop of over 40% in tasks such as community detection. Current optimizations fall into three categories: GCN-Res uses residual connections to preserve original features; JK-Net relies on skip connections to adaptively fuse features from different layers; and DropEdge breaks down over-aggregation by randomly dropping edges. On the Cora dataset, a 10-layer GCN maintains a classification accuracy of over 83%, a 27% improvement over the baseline. Scalability for large-scale graphs is an engineering challenge: traditional message passing requires traversing the entire graph adjacency matrix. A single forward computation on a billion-node graph has a complexity of $O(10^8)$, requiring over 100GB of graphics memory. The solution focuses on two key points: Sampling has been upgraded from GraphSAGE's uniform sampling to PinSAGE's importance sampling, and then to Cluster-GCN's graph partitioning. The latter splits the graph into subgraphs for independent training, making training time scale linearly with node size, achieving an 8x speedup on the 2.4 million-node ogbn-products dataset. For parallel computing, the DGL framework's distributed module supports computation on billion-node graphs through node partitioning and feature sharding, improving training efficiency by two orders of

magnitude on Alibaba Cloud clusters compared to single machines. Inadequate adaptability to dynamic and heterogeneous data limits its application. The addition and deletion of nodes and edges in dynamic graphs frequently changes the topology, making it difficult for traditional static mechanisms to update embeddings in real time. Although TGN incorporates temporal memory, it still experiences a latency of over 30% to handle sudden topological changes. Delta-GNN reduces latency from seconds to milliseconds by incrementally updating local node features. Heterogeneous graphs require unified modeling of multiple entity and relationship types [6]. Models like RGCN use independent parameters for each relationship, leading to a surge in parameters and overfitting when there are more than 100 relationships. HGT, relying on meta-path-guided attention weight sharing, reduces the number of parameters by 60% on the DBLP academic heterogeneous graph while maintaining 91% node classification accuracy. Poor interpretability hinders its application in high-reliability domains. The "black box" nature of the model makes it difficult to trace feature transfer. For example, in molecular drug prediction, it is impossible to explain the impact of atomic features on activity prediction. Two related studies exist: GAT uses attention weight heatmaps to locate key accounts for fraud detection; and GNNExplainer generates sparse subgraphs to identify key nodes and edges, achieving 85% feature attribution accuracy on the PubMed dataset. Future research focuses on cross-modal fusion and self-supervised learning. MM-GNN aligns medical cross-modal features, achieving an 89% lung cancer diagnosis accuracy. GraphCL, pre-trained through contrastive learning, achieves a classification accuracy of 78% on the Citeseer dataset, which has only 10% labels, a 15% improvement over supervised learning. GraphGPT enhances semantic understanding through graph-text alignment, paving the way for complex data reasoning.

7. Conclusion

This paper examines the principle architecture and application potential of the message passing mechanism in graph neural networks for structured data processing. Combining theoretical analysis with empirical cases, it identifies the mechanism's core value and development direction. The study found that the message passing mechanism, through a standardized "generate-aggregate-update" process, can effectively mine entity associations from structured data. Essentially, it transforms the topological features of the graph structure into learnable node embeddings, providing a fundamentally relevant technical solution for non-Euclidean data modeling. Theoretically, the message passing mechanism demonstrates strong adaptability and scalability. An analysis of typical models such as GCN, GAT, and GraphSAGE reveals that their core differences lie in the design of message generation and aggregation. This modularity enables flexible adaptation to diverse data types, including heterogeneous, dynamic, and large-scale graphs. A unified theoretical framework lays the foundation for

mechanism optimization, and strategies such as residual connections and attention weighting effectively mitigate inherent issues such as oversmoothing and overcompression. Empirically, this mechanism has been applied in numerous fields: breaking through the limitations of molecular property prediction accuracy in bioinformatics, revealing hidden transaction networks in financial risk management, and improving efficiency through product association mining in retail supply chains. These practices demonstrate its ability to accurately address the core pain point of "difficulty in modeling relationships" in structured data, providing a universal solution for complex cross-domain problems. While the current mechanism faces three major challenges: oversmoothing, scalability, and interpretability, a clear path for optimization is emerging: residual and skip connections mitigate feature convergence, sampling and distributed computing enhance large-scale data processing capabilities, and attention visualization and causal reasoning enhance model transparency. Future breakthroughs in more complex scenarios, such as brain neural network modeling and quantum chemistry computation, combined with self-supervised learning, cross-modal fusion, and large language model technologies, are expected. In summary, the message passing mechanism is not only the technical core of graph neural networks but also a revolutionary tool for structured data processing. With the improvement of theoretical frameworks and advancements in engineering technology, it will play a key role in unlocking the value of data relationships and promoting the implementation of AI, providing the core driving force for intelligent upgrades across various industries.

References

- [1] Zhong Zhe, Li Chuntao, Pang Jian. Hierarchical message passing graph neural network [J]. *Data Mining and Knowledge Discovery*, 2023, 37(1): 381-408.
- [2] He Hua, Yu Xiao, Zhang Jian, et al. Message passing meets graph neural network: a new paradigm for massive MIMO systems [J]. *IEEE Transactions on Wireless Communications*, 2023, 23(5): 4709-4723.
- [3] Ma Tianqi. Research on message passing mechanism and generalization strategy for SAT solution based on graph neural network [D]. Huazhong University of Science and Technology, 2024. DOI: 10.27157/d.cnki.ghzku.2024.006096.
- [4] Balcilar M, Héroux P, Gauzere B, et al. Breaking the limit of message passing graph neural network [C] // *International Conference on Machine Learning*. PMLR, 2021: 599-608.
- [5] Zhu Meiqi. Research on the message passing mechanism of graph neural network [D]. Beijing University of Posts and Telecommunications, 2022. DOI:10.26969/d.cnki.gbydu.2022.002655.
- [6] Ma Tianqi. Research on the message passing mechanism and generalization strategy of SAT solution based on graph neural network [D]. Huazhong University of Science and Technology, 2024. DOI:10.27157/d.cnki.ghzku.2024.006096.