Journal of Organizational and End User Computing
Volume 36 • Issue 1

Deep Learning-Powered Financial Product Recommendation System in Banks:
Integration of Transformer and Transfer Learning

Tingting Li, School of Accounting, Henan Finance University, China

Jingbo Song, Institute of Agriculture and Rural Development, Heilongjiang Provincial Academy of Social Sciences, China*

ABSTRACT

With the rapid evolution of financial technology, the recommendation system for financial products, as a crucial technology to enhance user experience and reduce information search costs, is increasingly becoming the focus of the financial services sector. As market competition intensifies, the diversity of user demands, coupled with the continuous expansion of financial product types, has exposed limitations in traditional recommendation systems regarding accuracy and personalized services. Therefore, this study aims to explore the application of deep learning technology in the field of financial product recommendations, aiming to construct a more intelligent and precise financial product recommendation system. The metrics we focus on include precision, recall, and F1-score, comprehensively evaluating the effectiveness of the proposed methods. In terms of methodology, we first employ a Transformer model, leveraging its powerful self-attention mechanism to capture the complex relationships between user behavior sequences and financial product information.

KEYWORDS

Banks, Financial Recommendation System, Graph Neural Networks, Intelligent Financial Technology Innovation, Transfer Learning, Transformer Model

With the advancement and widespread adoption of internet technology, financial services have transitioned from traditional offline modes to online platforms, providing users with greater convenience, efficiency, and a wider array of financial products and services (Gomber et al., 2018). However, due to the abundance of financial options and the diverse preferences of users, individuals often encounter challenges like information overload and decision-making difficulties when selecting appropriate financial products. To tackle this issue, financial product recommendation systems have emerged as intelligent solutions, leveraging technologies like deep learning. These systems analyze user behavior, preferences, risk profiles, and other characteristics to intelligently recommend suitable financial products or services. Financial product recommendation systems play a crucial role in the financial industry, contributing to enhanced user experience and satisfaction, increased sales conversion rates and revenue, as well as reduced operational costs and risks.

DOI: 10.4018/JOEUC.343257
*Corresponding Author

The field of financial product recommendation systems represents an interdisciplinary research area, amalgamating expertise from the financial domain with advanced deep learning technologies (Sharaf et al., 2022). It carries significant theoretical value and practical importance. Theoretically, these systems have the potential to drive knowledge innovation and technological advancements in the financial sector, offering new perspectives and methodologies for the evolution of financial technology. On a practical level, financial product recommendation systems enhance the quality and standard of financial services, delivering increased value and benefits to both users and financial institutions. Consequently, the exploration and thorough study of financial product recommendation systems are warranted in this dynamic and promising field.

Previous research on personalized financial product recommendation systems has primarily focused on several key directions, employing various methods and technologies to enhance the performance and adaptability of recommendation systems. These main directions include, but are not limited to:

- **Collaborative Filtering-Based Methods:** This approach involves analyzing user behaviors, such as product purchases or interests, to calculate the similarity between products or users. The system then recommends products that users have not yet shown interest in but may find appealing based on calculated similarities. While this method is advantageous for its simplicity and utilization of user feedback data to uncover potential interests, it has drawbacks. It is susceptible to biases like location, popularity, and selection biases, leading to unfair and inaccurate recommendations. It also struggles with the cold start problem concerning new users and products and faces challenges in considering product features and personalized user needs. Typical collaborative filtering methods include ItemCF (Gao, 2021), EXMF (Chen et al., 2018), IPS-MF, Doubly Robust, among others.
- **Content-Based Methods:** This method involves analyzing product feature attributes like risk level, investment period, and yield, to create a feature model for each product. Simultaneously, it analyzes user personal information, financial status, risk preferences, etc., to create an interest model for each user. The system then calculates the user's interest in a product based on the product's feature model and the user's interest model, thereby recommending products with higher relevance to the user's interests. This method is advantageous for considering product content and personalized user needs, facilitation of new product recommendations, and enhancing diversity and novelty in recommendations. However, it faces challenges in obtaining complete and accurate product features and user interest data, leading to suboptimal recommendation outcomes. It also struggles to capture complex relationships between products and users, resulting in imprecise recommendations, and has difficulty incorporating dynamic user feedback data. Typical content-based methods include CBR (Walsh et al., 2018), R-Transformer (Lian & Li, 2020), among others.
- **Hybrid Methods:** This approach combines collaborative filtering and content-based methods, leveraging both user behavior data and product feature data to provide recommendations from different perspectives. By overcoming drawbacks inherent in individual methods, it aims to enhance recommendation effectiveness. Hybrid methods offer advantages, including balancing multiple objectives like user satisfaction, product coverage, and recommendation fairness, thereby improving accuracy and diversity, and handling cold start issues, while adapting to various scenarios and requirements. However, these methods also introduce complexity and additional computational expenses, requiring adjustments of multiple parameters and weights, addressing conflicts and contradictions between different methods. Typical hybrid methods include Auto Debias (Chen et al., 2021), KDCRec (D. Liu et al., 2022), MACR (Zhang et al., 2023), among others.

While previous work has made some progress, personalized financial product recommendation systems still face numerous problems and challenges. For example:

1. **Data Quality and Quantity:** Financial data often exhibits challenges like high dimensionality, sparsity, imbalance, dynamics, and heterogeneity, which complicate learning and inference in recommendation systems. Additionally, constraints related to privacy and security affect the acquisition and processing of financial data, leading to insufficient data sources and volumes, thereby impacting the effectiveness and performance of recommendation systems.
2. **Complexity and Interpretability Issues of Models:** Financial product recommendation systems need to consider various factors, including user behaviors, preferences, risks, credits, product features, returns, risks, liquidity, as well as the relationships, similarities, and interactions between users and products. These relationships are often nonlinear, non-monotonic, and non-static, requiring recommendation systems to use complex models for modeling and inference. However, complex models often lack interpretability, making it difficult for users and financial institutions to understand the reasons and basis for recommendations, thereby reducing the credibility and transparency of recommendations.
3. **Diversity and Long-Term Goals:** Financial product recommendation systems need to balance between multiple objectives, such as user satisfaction, sales revenue, and risk control. These objectives may conflict or have different weights, requiring recommendation systems to perform reasonable optimization and adjustments. Moreover, financial product recommendation systems need to consider long-term effects like user retention, loyalty, and lifetime value. These effects may differ from short-term effects or have delays, necessitating dynamic evaluation and updates by recommendation systems.

This study aims to further explore and address the aforementioned issues by proposing a deep learning approach that integrates Transformer, transfer learning, and graph neural networks (GNNs) to enhance the performance and adaptability of the recommendation system. With the introduction of deep learning techniques, the authors expect to achieve innovative breakthroughs in handling issues like data sparsity, cold start issues (Lee et al., 2019), privacy security, biases, opacity, multi-objective optimization, and long-term problems. While some related works in the current research landscape have focused on few-shot class incremental learning, to better understand the latest developments and research trends in this field, the authors have studied the article by Tian et al. (2024). This review provides a comprehensive survey and summary of few-shot class incremental learning, offering valuable references and background information pertinent to the current research.

To tackle the aforementioned issues and challenges, the research objective of this article is to construct a deep learning-driven financial product recommendation system. By harnessing advanced technologies like Transformer and transfer learning, the aim is to achieve precise modeling and recommendation for financial users and products, thereby enhancing the effectiveness and performance of the recommendation system.

First, the authors introduce the Transformer model to overcome the limitations of traditional methods in handling sequential data. The self-attention mechanism of Transformer helps capture long-range dependencies, thereby improving the modeling capability of the recommendation system for user behavior sequences.

Second, transfer learning is applied to the financial product recommendation system to enhance performance across different yet related tasks by leveraging knowledge acquired from one task. This contributes to addressing data sparsity and cold start issues, thereby enhancing adaptability to new users and products.

Lastly, the authors incorporate GNNs into the recommendation system to model the complex relationships between users and products. This is expected to enhance the deep-level associations between products and users, which are challenging for content-based recommendation methods to capture.

Through this research, the authors aim to contribute new perspectives and methodologies to the development of personalized financial product recommendation systems. The work not only holds innovative value for recommendation system research but also has the potential to provide financial institutions with more intelligent and accurate recommendation services in practical applications, driving advancements in financial technology. The authors believe that the outcomes of this research will offer beneficial insights and momentum for the future development of personalized financial product recommendation systems.

The contributions of this article can be summarized in the following three aspects:

1. **Introduction of the Transformer Model:** By incorporating the Transformer model, the authors successfully applied its self-attention mechanism to comprehensively and accurately capture the relationships between user historical behavior sequences and financial product information. The outstanding performance of the Transformer model in sequence modeling tasks allows the recommendation system to better understand user behavior patterns, thereby enhancing recommendation accuracy.
2. **Utilization of Transfer Learning:** This article fully leveraged the concept of transfer learning by pre-training the model on generalized data and transferring its knowledge to the personalized recommendation task within the financial domain. This process enables the recommendation system to better adapt to the specific features of the financial domain, improving the system's generalization performance in real-world scenarios.
3. **Integration of GNNs:** With the introduction of GNNs, the authors successfully modeled the complex relationships between users and products by incorporating information from social and interaction networks into the recommendation system's modeling process. The application of GNNs enriches the recommendation system's understanding of relationships between users and products, making the recommendations more personalized and comprehensive.

The logical structure of this article is as follows. The first section outlines the importance and application background of personalized financial product recommendation systems, leading to the research objectives and significance. The second section reviews previous research and methods regarding personalized financial product recommendation systems. It introduces the advantages and disadvantages of collaborative filtering and content-based recommendation methods. The third section provides a detailed introduction to the deep learning approach proposed in this study. It includes Transformer models, transfer learning, and GNNs, elaborating on the roles and principles of these three techniques in the research. The fourth section describes the experimental setup and the financial dataset used. It analyzes the experimental results, compares the recommendation effects of the experimental and control groups, and discusses improvements in metrics like precision, recall, and F1-score. In the fifth section, the main contributions and innovations of the study are summarized. Future research directions in the field of deep learning-driven financial product recommendation systems are outlined, emphasizing the importance and application prospects of these systems.

RELEVANT WORK

Amidst the rapid development of financial technology, personalized financial product recommendation systems have emerged as a pivotal technology within the financial services sector (Behera et al., 2020). This system aims to offer users personalized recommendations of financial products by analyzing their financial behaviors, preferences, and needs, thereby enhancing user experience and the efficiency of financial services. However, due to the complexity of financial products and the diversity of users, constructing an accurate and efficient recommendation system remains a challenging problem.

To gain a comprehensive understanding of the current status and future development trends of financial product recommendation systems, this study will review pertinent literature addressing the authors' research questions. It will analyze the strengths of these studies and assess the value they contribute to the research.

In the article by Garg and Singh (2018), the focus is on assessing the financial literacy level of the global youth population. By analyzing socio-economic and demographic factors, the authors identify their significant impact on the financial literacy level of the youth. Understanding the interrelationships between youth financial knowledge, attitudes, and behaviors provides a valuable theoretical foundation for the research.

Gomber et al. (2018) conducted a deep exploration of the technological innovation and processes within the financial services industry. By elucidating the forces driving the fintech revolution, the authors present the urgent need for the financial services industry to transform its business models, customer experiences, and services. This serves as inspiration for the research on innovative financial product recommendation systems.

In the article by Zhang et al. (2019), a comprehensive review of deep learning's research progress in the field of recommendation systems is presented. Through categorizing deep learning models based on recommendation systems and summarizing the current technological landscape, it provides a clear direction for the authors' selection of deep learning methods. This is enlightening for constructing a more efficient financial product recommendation system.

Naumov et al. (2019) focused on the development of deep learning recommendation models, particularly in the context of personalized and recommendation system tasks. By offering implementations and performance evaluations of deep learning recommendation models, they provide practical experience for the model selection, aiding in system design and performance optimization.

Portugal et al. (2018) systematically reviewed the application of machine learning algorithms in recommendation systems. By analyzing existing recommendation system categories, adopted machine learning methods, the use of big data technologies, and key performance metrics, it provides important references for the selection of machine learning algorithms in recommendation systems, helping in understanding the strengths and weaknesses of different algorithms.

Greenquist et al. (2019) introduced a prediction analytics-based online product recommendation framework. Through practical cases, it showcases the end-to-end process of building a complete recommendation system, providing a practical reference framework for the research, especially in the context of real-time recommendation services.

Additionally, from the perspective of information disclosure, Wang et al. (2023) proposed a text-based competitive network model. Through the analysis of textual data, the model reveals the impact of information disclosure on participants within a competitive environment, offering a more comprehensive perspective for the research.

While the aforementioned studies provide valuable insights, they exhibit certain limitations. First, some research may not have fully considered the specificity of financial products, thereby challenging the adaptability of recommendation systems in the financial domain. Second, existing studies might not have adequately addressed user privacy and security issues, which are particularly crucial in the financial sector. Lastly, some research may lack sufficient empirical studies, making it difficult to validate the effectiveness of their methods in real financial environments.

Building upon an understanding of the relevant literature, the current research aims to overcome these limitations by constructing a more accurate and reliable personalized financial product recommendation system, thereby enhancing user experience and financial service efficiency. This study implements the following steps.

First, the authors collected financial product-related data, including user historical behavior sequences, product information, and user-product interaction graph data. They partitioned the data into training, validation, and test sets to ensure sufficient and representative data.

Second, the authors employed Transformer models to process user historical behavior sequences and product information. Utilizing self-attention mechanisms, the authors captured relationships between different elements in the sequences and pretrained the Transformer models on large-scale general datasets to obtain more universal user behavior patterns. Next, through fine-tuning, the authors transferred the pretrained Transformer models to personalized recommendation tasks, adapting them to the specific characteristics of the financial domain.

Simultaneously, the authors introduced GNNs to handle the user-product interaction graph, effectively capturing the complex relationships between users and products. They fused the graph embeddings learned by GNNs with the user and product representations learned by Transformer models, thus comprehensively considering information at different levels.

Finally, through model training and evaluation, using relevance metrics like accuracy and recall in financial recommendation tasks, the authors ensured excellent performance of the integrated models in personalized financial product recommendations. During the model training process, they also optimized the hyperparameters of the integrated models to further enhance model performance.

Through this series of experimental steps, the authors fully leveraged the advantages of Transformers, transfer learning, and GNNs in sequence modeling, general knowledge transfer, and graph data processing, providing a comprehensive and profound solution to improve the performance of recommendation systems.

In comparison to previous research, the current study innovates by holistically applying deep learning technologies, addressing privacy and security concerns, and enhancing the interpretability and user experience of the recommendation system. Past research may not have thoroughly considered these aspects, thus limiting the feasibility and user acceptance of recommendation systems in practical applications. The research aims to advance the development of personalized financial product recommendation systems, rendering them more applicable to the complex scenarios within the financial domain (Feng & Chen, 2022).

Through the analysis of existing research, the authors recognize that personalized financial product recommendation systems still encounter various challenges. Simultaneously, they observe the widespread application of deep learning in recommendation systems across various academic disciplines. However, it is important to note that there are differences between disciplines, necessitating the selection of appropriate methods based on specific tasks. In this context, the research, grounded in deep learning, takes into account the complexity of financial products, user privacy, and the interpretability of recommendation systems. The authors aim to make innovative breakthroughs in these critical areas, striving to provide practical solutions for achieving a more reliable and efficient personalized financial product recommendation system (Ye et al., 2023). The authors believe that the outcomes of this research will bring substantive impetus and insights to the field of financial technology.

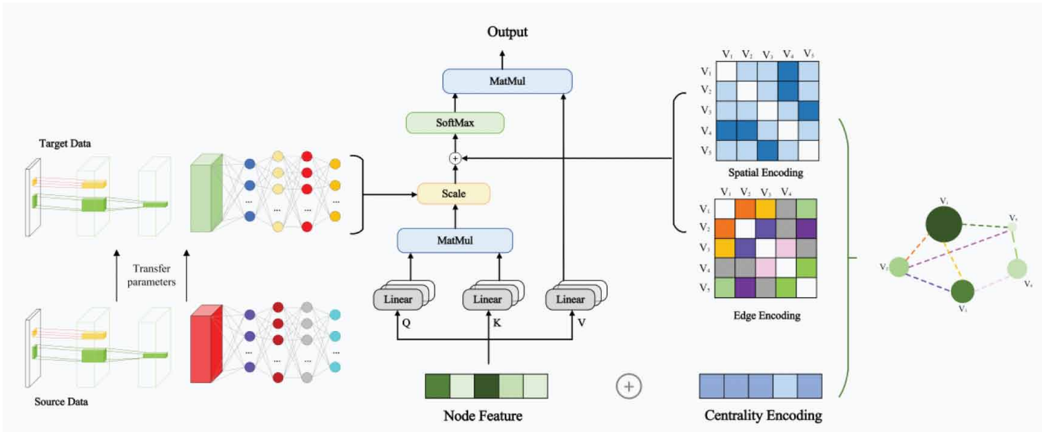
METHOD

This section of the article will provide a detailed introduction to the three key methods adopted: (1) the Transformer model; (2) transfer learning; and (3) GNNs. These methods play crucial roles in constructing the personalized financial product recommendation system. Through comprehensive algorithmic analysis and theoretical exposition, the authors aim to present the core ideas and working principles of each method to the readers. To visually depict the overall algorithmic framework proposed, the authors will present the general algorithmic framework diagram in the following section, enabling readers to gain a comprehensive understanding of the design and implementation of the personalized financial product recommendation system they have constructed. The overall algorithmic framework diagram is illustrated in Figure 1.

Transformer Model

The Transformer model is a deep learning architecture based on the self-attention mechanism (Gillioz et al., 2020). It is designed to handle sequential data, such as financial product information and user behavior sequences. Unlike traditional recurrent neural networks (Schmidt, 2019) or convolutional

Figure 1. Overall Algorithm Framework Diagram

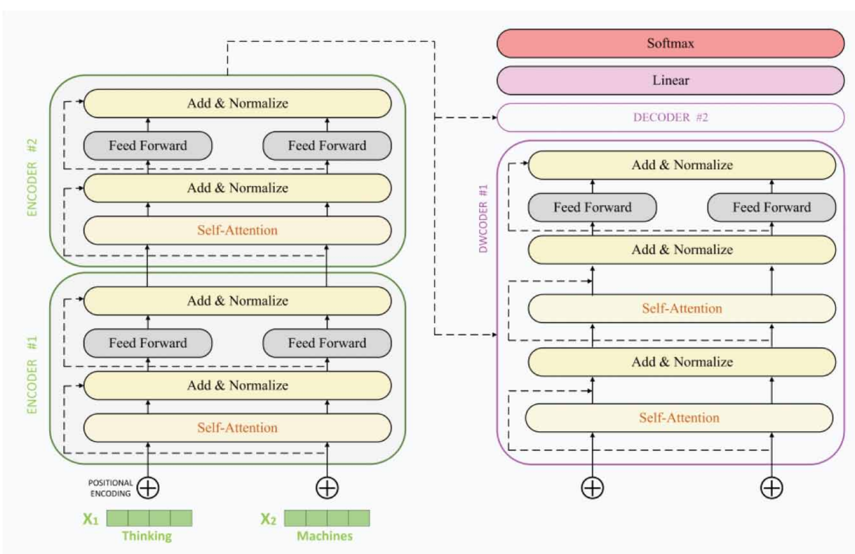


neural networks (Li et al., 2021), the Transformer model completely abandons the sequential processing of sequences. Instead, it leverages the self-attention mechanism to capture dependencies between arbitrary positions in the sequence. This approach enhances model parallelism and reduces training time but also equips the model to handle long-distance dependencies more effectively.

The Transformer model consists of two parts: an encoder and a decoder. Each part comprises multiple identical layers, each layer including multiple self-attention sub-layers and feedforward neural network sub-layers. Next, the article will provide a detailed introduction to the structure and principles of the Transformer model, as well as its application in the financial product recommendation system. The architecture of the Transformer model is illustrated in Figure 2.

First, the authors need to perform embedding and positional encoding on the input data. Embedding involves converting discrete data, such as words or product IDs, into continuous vector

Figure 2. Transformer Model



representations for neural networks to process. Positional encoding involves adding positional information for each data point in the sequence to the embedded vector, allowing the model to differentiate between data points at different positions. Assuming the input data is a user behavior sequence, where each behavior corresponds to a product ID, the study can use the following formula to calculate embedding and positional encoding:

$$x_i = E_{id_i} + PE_i$$

where x_i is the input vector for the ith product, E_{id_i} is the embedding vector for the ith product ID, and PE_i is the positional encoding vector for the ith position. The authors can obtain E_{id_i} using a pre-trained embedding matrix or a randomly initialized embedding matrix. The positional encoding PE_i can be calculated using the following formula:

$$PE_{i,2k} = \sin\left(i \,/\, 10000^{2k/d}\right)$$
$$PE_{i,2k+1} = \cos\left(i \,/\, 10000^{2k/d}\right)$$

where d is the dimension of the embedding vector and k is the index of the dimension. This approach allows the model to adapt to sequences of arbitrary lengths and facilitates the calculation of relative positions.

Next, the authors input the vector x_i into the encoder, which consists of N identical layers. Each layer has two sub-layers: a multi-head self-attention sub-layer and a feedforward neural network sub-layer. The multi-head self-attention sub-layer enables the model to discern the correlation between each data point and other data points in the sequence, capturing the internal structure of the sequence. The feedforward neural network sub-layer performs a non-linear transformation on each data point, thereby enhancing the model's expressive power. Following each sub-layer, there is a residual connection and layer normalization operation to improve the stability and convergence speed of the model. The output of the l-th layer of the encoder can be represented using the following formulas:

$$z_i^l = \mathrm{LayerNorm}\left(x_i + \mathrm{MultiHeadAttention}\left(x_i, x, x\right)\right)$$
$$h_i^l = \mathrm{LayerNorm}\left(z_i^l + \mathrm{FFN}\left(z_i^l\right)\right)$$

where x is the matrix representation of the input sequence, x_i is the ith input vector, z_i^l is the ith intermediate vector of the l-th layer, $\left(h_i^l\right)$ is the ith output vector of the l-th layer, $LayerNorm$ is the layer normalization operation, $MultiHeadAttention$ is the multi-head self-attention sub-layer, and FFN is the feedforward neural network sub-layer. The authors can use the output matrix h^N of the last layer of the encoder as the encoded representation of the input sequence, which contains contextual information for each data point in the sequence.

The specific calculation process of the multi-head self-attention sub-layer (Voita et al., 2019) is as follows: First, the input vector x_i is linearly transformed to obtain query vector q_i, key vector k_i, and value vector v_i, all with dimensions d_k. Then, the authors calculate the dot product of each query vector with all key vectors, resulting in an attention score matrix with dimensions $n \times n$, where n is the length of the sequence. Next, the softmax operation is scaled and applied to the attention score matrix, obtaining an attention weight matrix with dimensions $n \times n$. Finally, the authors multiply the attention weight matrix by the value vector matrix to obtain an output matrix with dimensions $n \times d_k$. The output of the self-attention sub-layer can be expressed using the following formula:

$$\text{Attention}\left(x\right) = \text{softmax}\left(\frac{xW_Q\left(xW_K\right)^T}{\sqrt{d_k}}\right)xW_V$$

where W_Q, W_K, and W_V are linear transformation matrices, and $\sqrt{d_k}$ is the scaling factor. To enhance the expressive power of the model, the authors can split the self-attention sub-layer into multiple heads. Each head uses different linear transformation matrices, and the outputs from each head are concatenated and linearly transformed to obtain the output of the multi-head self-attention sub-layer. The output of the multi-head self-attention sub-layer can be expressed using the following formula:

$$MultiHeadAttention\left(x\right) = \text{Concat}\left(\text{head}_1, \text{head}_2, \ldots, \text{head}_h\right)W_O$$

where $\text{head}_i = \text{Attention}\left(xW_Q^i, xW_K^i, xW_V^i\right)$ W_Q^i, W_K^i, W_V^i, and W_O are linear transformation matrices, and h is the number of heads.

The specific calculation process for the feedforward neural network sub-layer is as follows. First, a linear transformation is performed on the input vector z_i^l, obtaining an intermediate vector with dimensions d_{ff}. Then, the authors apply an activation function, such as ReLU or GELU, to the intermediate vector, obtaining an activation vector with dimensions d_{ff}. Finally, another linear transformation is performed on the activation vector, obtaining an output vector with dimensions d. The output of the feedforward neural network sub-layer can be expressed using the following formula:

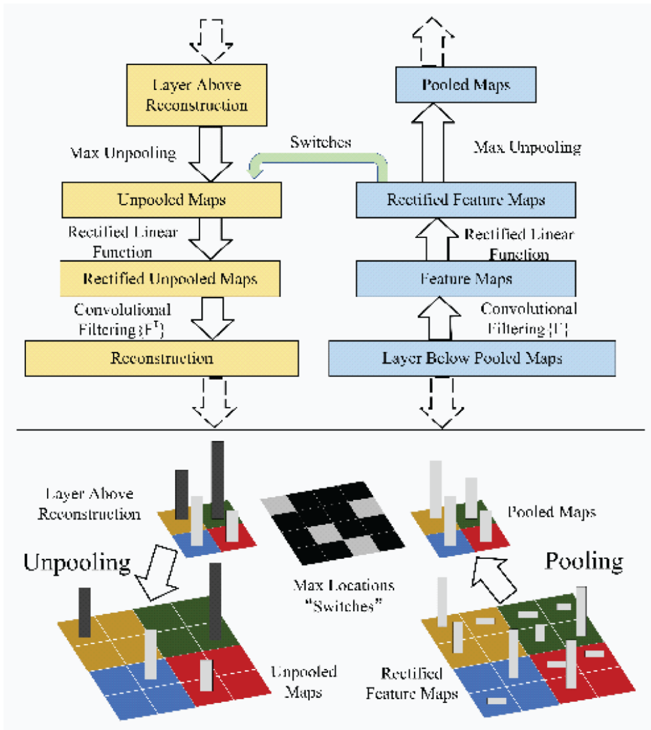
$$\text{FFN}\left(z_i^l\right) = \left(z_i^l W_1 + b_1\right)W_2 + b_2$$

where W_1, W_2, b_1, and b_2 are the parameters of the linear transformation, and d_{ff} is the dimensionality of the intermediate vector.

Building on the output of the encoder, the input vectors h_i^N are input into the decoder. The decoder consists of M identical layers, each with three sub-layers: (1) multi-head self-attention sub-layer; (2) multi-head encoder-decoder attention sub-layer; and (3) feedforward neural network sub-layer. The multi-head self-attention sub-layer allows the model to discern the correlations between each data point in the output sequence and others, capturing the internal structure of the sequence. The multi-head encoder-decoder attention sub-layer enables the model to leverage the encoder's output to generate a more suitable output sequence.

Through the Transformer model, the authors can more effectively model the complex relationships between user behavior and product features, providing the recommender system with more accurate personalized recommendations. Additionally, to achieve the simultaneous goals of feature selection and system identification, researchers have proposed the smoothing Group Lasso based interval type-2 fuzzy neural network (Gao et al., 2023). This method combines the advantages of feature selection and model identification, effectively addressing the challenges posed by high feature dimensions and data uncertainty. Next, the article will provide a detailed introduction to the transfer learning method, offering a more comprehensive solution for the overall recommender system.

Figure 3. Transfer Learning



Transfer Learning

Transfer learning is a machine learning approach that leverages existing knowledge or models to enhance learning performance in new yet related domains or tasks (Zhuang et al., 2020). At the core of transfer learning lies the identification of similarities or commonalities between the source domain and the target domain, enabling the transfer and sharing of knowledge. The advantages of transfer learning include reducing dependence on annotated data, improving model generalization, accelerating model convergence, reducing model complexity, and enhancing overall performance. The structure of transfer learning is illustrated in Figure 3.The general form of transfer learning can be defined as follows: Given a source domain $\mathcal{D}_S$ and a source task $\mathcal{T}_S$, as well as a target domain $\mathcal{D}_T$ and a target task $\mathcal{T}_T$, the objective of transfer learning is to leverage information from $\mathcal{D}_S$ and $\mathcal{T}_S$ to improve the learning performance on $\mathcal{D}_T$ for $\mathcal{T}_T$. Here, a domain $\mathcal{D}$ includes a feature space $\mathcal{X}$ and a marginal probability distribution $P\left(X\right)$, while a task $\mathcal{T}$ includes a label space $\mathcal{Y}$ and a conditional probability distribution $P\left(Y|X\right)$.

Transfer learning can be classified based on various criteria. For instance, based on the relationship between the source and target domains, it can be categorized into homogeneous transfer learning and heterogeneous transfer learning. According to the transfer hierarchy, it can be divided into instance-based transfer learning, feature-based transfer learning, model-based transfer learning, and relation-based transfer learning. Based on the transfer direction, it can be categorized as unidirectional transfer learning and bidirectional transfer learning (Du & Chen, 2023). Depending on the transfer purpose, it can be classified as inductive transfer learning, transductive transfer learning, and generalized transfer learning.

In the personalized financial product recommendation system, transfer learning can be employed to pretrain models on large-scale general data, capturing universal user behavior patterns, and then transfer this knowledge to personalized recommendation tasks to adapt to the specific characteristics of the financial domain. This study adopted a feature-based transfer learning approach using the Transformer model. This approach maps data from the source domain and the target domain from the original feature space to a new feature space, making the data distributions of the two domains more similar and facilitating knowledge transfer.

Specifically, the authors utilized a pretrained Transformer model as a feature extractor. This model encoded user behavior sequences and financial product information into high-dimensional vectors. These vectors were then fed into a classifier for financial product recommendation. To adapt to the specific data characteristics of the financial domain, the authors pretrained the Transformer model on general data and fine-tuned it on financial data. This strategy aimed to achieve knowledge transfer to the financial domain and enhance the model's generalization capabilities. The process of pretraining and fine-tuning is outlined below:

- **Pretraining:** The study conducted pretraining on a large-scale general text dataset using self-supervised learning objectives like the masked language model (Salazar et al., 2019) and the permutation language model (PLM). The goal was to train the Transformer model to learn universal language knowledge and feature representations. Additionally, in the biomedical field, there are studies that use pretrained deep learning models for downstream tasks. For instance, Zhou et al. (2023) proposed an approach based on the ProtBert pretrained model, utilizing protein sequence information for downstream prediction tasks, achieving notable results.
- **Fine-Tuning:** The authors performed fine-tuning on a small-scale financial text dataset, incorporating information like descriptions, evaluations, and labels of financial products. Supervised learning objectives, such as binary or multiclass classification, were employed to fine-tune the parameters of the Transformer model. This process aimed to adapt the model to the data characteristics and task requirements specific to the financial domain.

The processes of pretraining and fine-tuning can be represented by the following formulas:

$$\theta^* = \text{argmax}_\theta \sum_{(x,y)\in D_s} \log P\left(y\# \; x;\theta\right)$$

$$\phi^* = \text{argmax}_\phi \sum_{(x,y)\in D_t} \log P\left(y\# \; x;\theta^*,\phi\right)$$

In the equations, θ represents the parameters of the Transformer model, ϕ represents the parameters of the classifier, $\mathcal{D}_S$ denotes the general text dataset, $\mathcal{D}_T$ represents the financial text dataset, x denotes the input text, and y denotes the output label. The objective of pretraining is to maximize the log-likelihood on the general text dataset, obtaining the optimal Transformer model parameters θ^*. The fine-tuning objective is to maximize the log-likelihood on the financial text dataset, obtaining the optimal classifier parameters ϕ^*.

To optimize the aforementioned objective functions, the study employs stochastic gradient descent (SGD) or its variants like Adam as the optimization algorithm (Ketkar & Ketkar, 2017). The fundamental idea of the optimization algorithm is to compute the gradient of the objective function with respect to the parameters. Then, the parameters are updated in the direction opposite to the gradient with a certain learning rate, gradually reducing the value of the objective function until convergence. The update formula for the optimization algorithm is as follows:

$$\theta_{t+1} = \theta_t - \eta_t \nabla_\theta J\left(\theta_t\right)$$
$$\phi_{t+1} = \phi_t - \eta_t \nabla_\phi J\left(\phi_t\right)$$

In the equations, θ_t and ϕ_t represent the parameters at the t-th iteration, η_t represents the learning rate at the t-th iteration, $J\left(\theta_t\right)$ and $J\left(\phi_t\right)$ represent the objective functions at the t-th iteration, and $\nabla_\theta J\left(\theta_t\right)$ and $\nabla_\phi J\left(\phi_t\right)$ represent the gradients at the \(t\)-th iteration.

In financial recommendation tasks, transfer learning, by pretraining the model on general data and then fine-tuning it in the financial domain, can better capture general patterns in user behavior. This approach enhances the accuracy and adaptability of personalized recommendations. Next, the article will provide a detailed introduction to the application of GNNs to offer a more comprehensive recommendation solution for the integrated model.

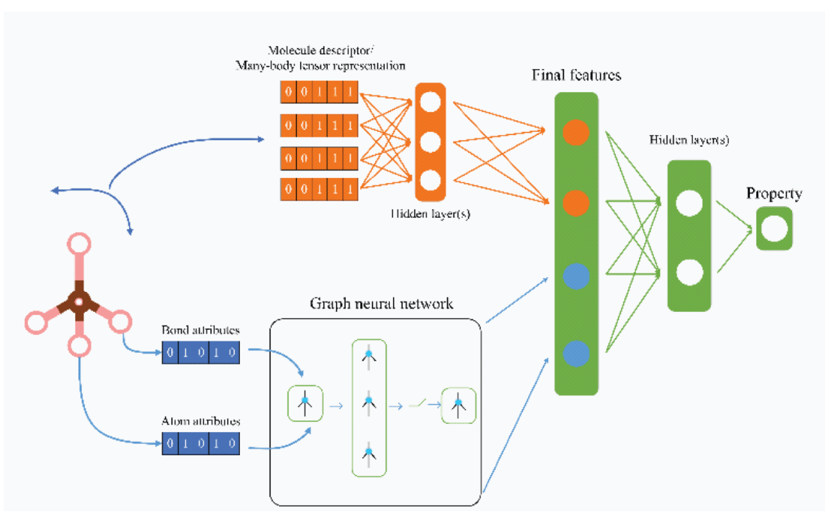
GNN

GNN is a method that uses neural networks to learn from graph-structured data (Wu et al., 2020). Graph-structured data consists of nodes and edges in a non-Euclidean space, representing various complex relationships like social networks, knowledge graphs, financial products, etc. The goal of GNN is to learn an embedding representation for each node, which is a low-dimensional vector that encapsulates both the node's own features and information from its neighboring nodes. With this embedding representation, various graph-based tasks can be performed, including node classification, link prediction, recommendation systems, etc.

In personalized financial product recommendation systems, GNN is introduced to effectively capture the intricate interactions between users and products, thereby enhancing the performance of the recommendation system.

The core idea of GNNs involves updating node features through the aggregation of information from neighboring nodes, thereby capturing the graph's topological structure and node attributes. The structure of a GNN is illustrated in Figure 4.

Figure 4. GNN



The general form of a GNN can be represented as follows:

$$h_v^{(k)} = f\left(h_v^{(k-1)}, \left\{h_u^{(k-1)} : u \in \mathcal{N}\left(v\right)\right\}\right)$$

where $h_v^{(k)}$ represents the feature vector of node v at layer k, $\mathcal{N}\left(v\right)$ denotes the set of neighbors for node v, and f is an aggregation function that can be any differentiable function, such as average, maximum, sum, etc. GNNs typically consist of multiple layers, each updating the features of nodes. The features from the last layer can be used for downstream tasks.

An essential question in the realm of GNNs is how to design an appropriate aggregation function to make effective use of graph structure and attribute information, all while maintaining computational efficiency and scalability. Currently, various types of GNNs have been proposed, broadly classified into two categories based on different aggregation functions: spectral-based GNNs and spatial-based GNNs.

Spectral-based GNNs introduce filters from the perspective of graph signal processing to define graph convolution, interpreting graph convolution operations as noise removal from graph signals. A typical representative of spectral-based GNNs is graph convolutional networks (GCN), and its aggregation function can be expressed as:

$$h_v^{(k)} = \sigma\left(\sum\nolimits_{u \in \mathcal{N}\left(v\right) \cup \left\{v\right\}} \frac{1}{\sqrt{d_v d_u}} W^{(k)} h_u^{(k-1)}\right)$$

where σ represents the activation function, d_v denotes the degree of node v, and $W^{(k)}$ is the weight matrix for layer k. GCN has the advantage of being simple to implement and effective in capturing the local structure of the graph. However, it also has drawbacks, such as requiring knowledge of the entire graph structure, an inability to handle dynamic and heterogeneous graphs, and lack of differentiation between nodes with different degrees.

Spatial-based GNNs represent graph convolution as the aggregation of feature information from the neighborhood. They do not require knowledge of the entire graph structure and can handle dynamic and heterogeneous graphs. A typical representative of spatial-based GNNs is graph attention networks (GAT), and its aggregation function can be expressed as:

$$h_v^{(k)} = \sigma\left(\sum\nolimits_{u \in \mathcal{N}\left(v\right) \cup \left\{v\right\}} \alpha_{vu}^{(k)} W^{(k)} h_u^{(k-1)}\right)$$

where $\alpha_{vu}^{(k)}$ represents the attention coefficient between nodes v and u in layer k. It can be computed through a learnable function, for example:

$$\alpha_{vu}^{(k)} = \frac{\exp\left(Leaky\,\mathrm{Re}\,LU\left(a^{(k)T}\left[W^{(k)} h_v^{(k-1)} \mid W^{(k)} h_u^{(k-1)}\right]\right)\right)}{\sum\nolimits_{w \in \mathcal{N}\left(v\right) \cup \left\{v\right\}} \exp\left(Leaky\,\mathrm{Re}\,LU\left(a^{(k)T}\left[W^{(k)} h_v^{(k-1)} \mid W^{(k)} h_w^{(k-1)}\right]\right)\right)}$$

where $a^{(k)}$ represents the attention vector for layer k, and $||$ denotes vector concatenation. The advantage of GAT lies in its ability to adaptively assign varying weights to different neighbors, enabling differentiation among nodes. However, it also has some drawbacks, such as the higher computational complexity of the attention mechanism and a lack of distinction between neighbors at different distances.

It is worth noting that attention mechanisms have been successfully applied in image classification tasks. Zhu et al. (2023) introduced a fine-grained image classification method based on attention mechanisms. This approach utilizes attention mechanisms to extract features from different regions of the image, achieving more refined classification results. This provides an example of applying attention-based ideas in computer vision tasks.

Drawing inspiration from this, the authors similarly incorporate attention mechanisms to handle relationships between users and different products. Specifically, they introduce attention mechanisms into GNNs, allowing dynamic learning of the varying importance of different products in the users' preferences. This facilitates the achievement of more comprehensive and personalized recommendations (Zeng & Zhong, 2022).

This article primarily focuses on a financial product recommendation system based on GNNs. The objective is to leverage GNNs to understand the relationships between users and products, thereby providing users with suitable product recommendations. To achieve this goal, the authors employ the following methods.

First, users and products are represented as nodes in a heterogeneous graph, where user behaviors and attributes serve as edges, while user and product features constitute node attributes. Second, the study utilizes the Transformer model, leveraging its powerful self-attention mechanism to capture the intricate relationships between user behavior sequences and product information. Simultaneously, transfer learning is introduced, pretraining the model on general data to facilitate knowledge transfer to the financial domain, better adapting to the specific features of financial data. Lastly, GNNs are applied to aggregate user and product neighbor information, updating representations comprehensively to consider users' social and interaction networks.

Through these methods, the authors obtain high-dimensional vector representations for users and products. Subsequently, simple similarity metrics or learning-based matching functions can be used to calculate the degree of match between users and products, facilitating the recommendation of the most suitable products for users. The next section will introduce the study's experimental setup and results to validate the effectiveness and superiority of our approach.

EXPERIMENTS

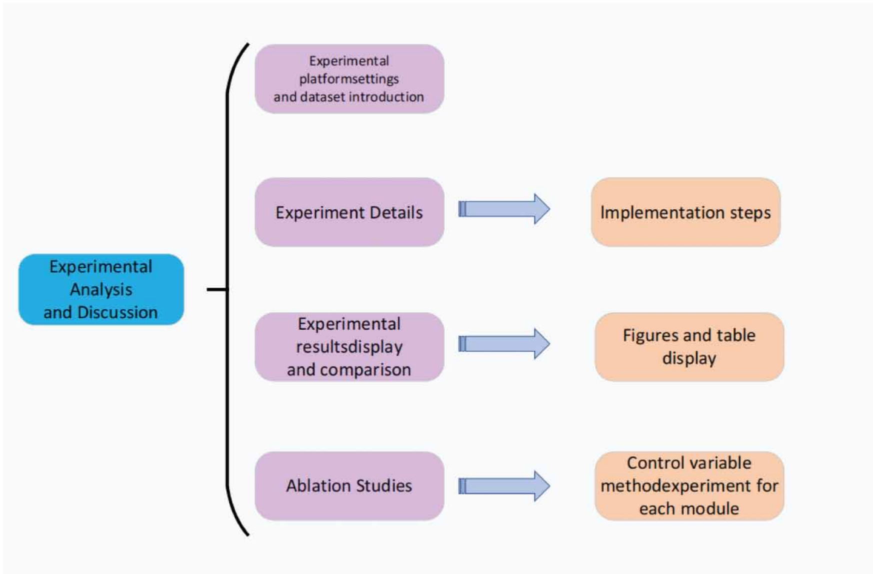
This article will offer a detailed overview of the experimental setup, encompassing the experimental environment, dataset selection, and evaluation metrics. Through a comprehensive experimental design, the goal is to validate the performance of the integrated approach involving Transformer, transfer learning, and GNNs in personalized financial product recommendation systems. Finally, the authors will conduct a thorough data analysis of the experimental results to unveil the strengths and limitations of each method in recommendation tasks. The overall flowchart of the experiment is illustrated in Figure 5, providing readers with a holistic understanding of the experimental design.

Experimental Environment

Hardware Environment

For this experiment, the authors selected an advanced high-performance computing server with hardware configurations ensuring efficient execution of computational and storage tasks. The server is equipped with an Intel Core i9-10900K @ 3.70GHz CPU and 256GB RAM, featuring four Nvidia GeForce RTX 3080 10GB GPUs. This powerful and cutting-edge hardware setup provides

Figure 5. Overall Experimental Flow Chart



abundant computational resources, enabling efficient training and inference of deep learning models. The collaborative operation of these hardware components offers a stable and high-performance experimental environment, ensuring the reliability and accuracy of the experimental results.

Software Environment

This study utilized Python as the primary programming language and PyTorch as the deep learning framework to implement the end-to-end architecture of the authors' personalized financial product recommendation system. Python's flexibility and robust ecosystem provide rich tools and libraries, facilitating efficient data processing, model implementation, and experimental analysis. PyTorch, as a top-tier deep learning framework, offers an intuitive and flexible interface for model construction and training.

Throughout the experiment, the authors leveraged PyTorch's computational capabilities and automatic differentiation functionality to accelerate the model training process, ensuring the authors' personalized financial product recommendation model converges faster and achieves superior performance. Choosing Python and PyTorch as software tools contributes to improved development efficiency, ensuring the research can effectively showcase its performance and feasibility in the experiment.

Experimental Data

Santander Product Recommendation Dataset (SPRD)

The SPRD is a dataset related to financial product recommendations (Pryor, 2016), provided by the Spanish bank Santander and consisting of authentic customer data from a competition hosted by Santander on Kaggle. The competition aimed to enhance Santander's customer experience and satisfaction by allowing customers to choose suitable financial products based on their needs and preferences.

The dataset covers a period of 17 months from January 2015 to May 2016, with approximately 760,000 users each month, resulting in a total of 13.647 million user-month records. Each user is characterized by 24 features, including age, gender, income, residence, and others. Additionally, each user is labeled with 24 products, indicating whether they owned each product in a given month, such as credit cards, savings accounts, etc. There are a total of 24 different financial products included in the dataset.

The dataset exhibits high quality, with no missing or outlier values. However, some features are encrypted, such as user IDs and residence, which may affect the interpretability and visualization of these features. Furthermore, the dataset faces significant class imbalance, as the majority of users do not purchase new products each month, with only approximately 3.51% of users adding new products monthly. This implies that the number of positive samples is much lower than negative samples, necessitating strategies like oversampling, undersampling, or adjusting class weights to address this issue.

This dataset holds substantial value for this study as it allows the authors to explore issues related to financial product recommendations, analyze user behavior and preferences, build effective recommendation models, evaluate recommendation performance, and propose improvement strategies.

Amazon-Ratings (Beauty) Dataset

The ARBP is a dataset related to beauty product recommendations (Tan et al., 2018), sourced from Kaggle and comprising rating data for beauty products on the Amazon website. Extracted as a subset from the Amazon Review Data (2018), this dataset includes all reviews associated with beauty products, providing insights into user preferences and demands for beauty products, along with their evaluations and feedback.

The dataset encompasses 371,345 reviews, each containing information like user ID, product ID, rating (one to five stars), review text, and review timestamp. It covers a variety of beauty product types, including cosmetics, skincare products, perfumes, hair care items, and more. With a substantial size, the dataset offers ample data for training and testing recommendation models.

The dataset exhibits good quality, free from missing values and outliers. However, there are some noteworthy issues, such as data sparsity and the cold start problem. Data sparsity refers to the fact that each user reviews only a small subset of products, and each product is reviewed by only a few users. This results in a highly sparse user-product rating matrix, making it challenging to identify similarity between users and products. The cold start problem arises for new users or products where there is insufficient rating data, making it difficult to provide reasonable recommendations.

Addressing these challenges may involve strategies like leveraging additional information about users and products, such as user profiles, product descriptions, and product categories, to enrich the dataset and enhance its utility.

This dataset holds significant value for this study, allowing the authors to explore product recommendation issues, analyze user ratings and reviews, build effective recommendation models, evaluate recommendation performance, and propose improvement strategies (Zhong & Zhao, 2024). It enables a multi-faceted examination of the problem, incorporating approaches like rating-based collaborative filtering, sentiment analysis based on reviews, content-based filtering, and deep learning methods.

Quandl Dataset (QD)

The QD is a financial and economic data platform that offers a diverse range of data (Liu, 2023), including stock prices, currency exchange rates, interest rates, macroeconomic indicators, and more. This data helps us understand global economic and financial dynamics, as well as the impact of various factors on financial products.

The data scale of QD is substantial, providing sufficient volume for training and testing recommendation models. The data quality of QD is high, having undergone multiple quality checks and validations to ensure accuracy, completeness, timeliness, and consistency. QD data is also highly accessible and available, supporting various data formats and programming languages, such as CSV, JSON, XML, Python, and R.

It exhibits good scalability and customizability, allowing users to create and manage their datasets, as well as subscribe to and purchase datasets from other users. QD holds significant value for this study as it can aid in exploring deep learning approaches for personalized financial product recommendation systems. It facilitates the analysis of user financial preferences and demands, the construction of effective recommendation models, the evaluation of recommendation effectiveness, and the proposal of improvement strategies.

The problem can be studied from multiple perspectives, including user historical data, user personal information, user feedback data, product attribute data, and market data. Leveraging the diversity and richness of QD's data, we can consider users' diversified and personalized needs, designing a more flexible and accurate recommendation system.

Financial Product Recommendation System Based on Transformer (FPRT) Dataset

The FPRT is a dataset related to financial product recommendations (Lian & Li, 2020). It was crawled by the authors from an internet finance platform and includes user basic information, financial product details, and user purchase records. This dataset helps us understand user demands, preferences for financial products, as well as their purchasing behavior and decision-making processes.

The dataset comprises 10,000 users, 100 financial products, and 20,000 purchase records. Each user has six features, including age, gender, income, occupation, education, and marital status. Each financial product has five features, including product type, term, interest rate, risk, and return. Each purchase record contains three features, including user ID, product ID, and purchase time.

The dataset's scale is moderate, providing sufficient data for training and testing recommendation models. It exhibits good quality with no missing or outlier values. However, some considerations are necessary, such as data imbalance and time sensitivity. Data imbalance refers to varying product quantities purchased by each user and varying user quantities purchasing each product, resulting in a sparse user-product rating matrix. Time sensitivity indicates that user preferences and product attributes change over time, affecting the timeliness and accuracy of recommendations.

Strategies, such as utilizing the Transformer's self-attention mechanism to capture long and short-term dependencies between users and products, can address these issues. FPRT holds significant value for this study, allowing exploration of deep learning approaches for personalized financial product recommendation systems. It enables analysis of user basic information and purchase records, construction of effective recommendation models, evaluation of recommendation effectiveness, and proposing improvement strategies. Additionally, leveraging FPRT's data features allows for the consideration of user personal information and purchase records, designing a more personalized and accurate recommendation system.

Evaluation Indicators

When evaluating the performance of the study's personalized financial product recommendation system, the authors employ a series of key metrics to comprehensively assess its effectiveness. These metrics not only aid in quantifying the recommendation system's performance across various aspects but also provide a basis for a holistic evaluation of the system's strengths and weaknesses. This article will provide a detailed explanation of three crucial evaluation metrics: (1) Precision; (2) Recall; and (3) F1-score. Through the analysis of these metrics, the authors aim to gain a comprehensive understanding of the accuracy, comprehensiveness, and overall performance of the personalized financial product recommendation system.

Precision

Precision is one of the key metrics in the evaluation of recommendation systems, playing a crucial role in the context of personalized financial product recommendations. Precision measures the proportion of recommended products that are genuinely of interest to the user, considering the products the user actually clicks on or purchases. In the field of financial product recommendations, this metric directly reflects the accuracy and effectiveness of the recommendation system in providing personalized services.

The formula for calculating Precision is as follows:

$$Precision = \frac{TP}{TP + FP} * 100\%$$

where TP represents the true positives (the number of financial products correctly recommended by the system and actually purchased or clicked by the user) and FP represents the false positives (the number of financial products incorrectly recommended by the system but not actually purchased or clicked by the user.

In the practical application of a financial product recommendation system, precision is a crucial metric. In recommendation systems, user needs may change due to factors like time and economic conditions. Therefore, a recommendation system with high precision is essential to accurately capture the user's current interests. A recommendation system with high precision can improve user satisfaction with the recommended results, thereby increasing user stickiness and usage frequency.

Precision is a significant metric in the evaluation of personalized financial product recommendation systems, accurately reflecting the system's performance concerning users' real needs. By optimizing the recommendation algorithm to improve precision, trust in the recommendation system can be enhanced, leading to improved practical application results. In subsequent experiments, the authors will delve into the application of precision in personalized financial product recommendation systems, analyzing specific data to validate the effectiveness of their approach.

Recall

Recall is one of the crucial metrics in the evaluation of recommendation systems, especially in the context of personalized financial product recommendations. Recall measures the extent to which the recommendation system can cover the user's actual interests (i.e., the proportion of products recommended by the system that are actually of interest to the user out of all products the user is interested in). In the field of financial product recommendations, recall directly relates to whether users can receive comprehensive personalized services, thereby influencing user satisfaction and platform usability.

The formula for calculating recall is as follows:

$$Recall = \frac{TP}{TP + FN} * 100\%$$

where TP represents the true positive cases (i.e., the number of financial products correctly recommended by the recommendation system and subsequently purchased or clicked by the user) and FN represents the false negative cases (i.e., the number of financial products that the recommendation system failed to recommend but were actually purchased or clicked by the user).

By increasing recall, the recommendation system can comprehensively cover potential areas of user interest, ensuring that users have the opportunity to discover and access various financial product information. A high recall indicates that the recommendation system can more comprehensively capture user interests, increasing the likelihood that users will find financial products that meet their expectations, thereby enhancing user satisfaction. Low recall may lead to users missing out on some products of actual interest. By improving recall, the system can reduce such misses and increase opportunities for users to access information.

Recall is a critical evaluation metric in personalized financial product recommendation systems, influencing the comprehensiveness of system services and user satisfaction. Through optimizing recommendation algorithms to improve recall, the system can better meet diverse user needs and enhance the overall user experience. In subsequent experiments, the authors will delve into the practical performance of recall in personalized financial product recommendations to validate the effectiveness of their approach.

F1-Score

The F1-score is a commonly used metric when evaluating the performance of personalized financial product recommendation systems. It provides a comprehensive assessment by considering both precision and recall, making it advantageous for evaluating the accuracy and comprehensiveness of recommendation systems, especially in situations with imbalanced positive and negative samples. In the financial domain, where users have higher expectations regarding the accuracy and comprehensiveness of recommended products, the F1-score has become a crucial evaluation metric.

The formula for calculating the F1-score is as follows:

$$F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} * 100\%$$

Precision represents the proportion of samples correctly predicted as positive out of the total samples predicted as positive by the model. Recall represents the proportion of samples successfully predicted as positive by the model out of the total actual positive samples.

The F1-score takes into account both precision and recall, effectively balancing the accuracy and comprehensiveness of the recommendation system. This is particularly crucial in the financial domain, where users want to see relevant products without being inundated with irrelevant ones. In financial product recommendation, user interactions with specific products are relatively rare, leading to an imbalance between positive and negative samples. The comprehensive nature of the F1-score allows it to adapt well to this situation, avoiding an overreliance on either accuracy or comprehensiveness. F1-score is a comprehensive evaluation metric that not only focuses on the correctness of the recommendation system but also considers its ability to cover users' potential interests. In practical financial recommendation scenarios, this contributes to improving the overall quality of the recommendation system.

The F1-score, as a comprehensive evaluation metric, holds significant importance in personalized financial product recommendation systems. By balancing the accuracy and comprehensiveness of the recommendation system, it provides a more holistic reflection of the system's performance.

Experimental Comparison and Analysis

Before delving into the experimental comparison and analysis section, the authors have provided a detailed introduction to the key metrics used in the personalized financial product recommendation system, including precision, recall, and F1-score. These metrics offer a comprehensive perspective for evaluating the system's performance, aiming to find the optimal balance between accuracy and comprehensiveness.

To validate the impact of the study's proposed integrated approach on the performance of the recommendation system, the authors conducted a series of experiments comparing their method with baseline approaches. Utilizing advanced hardware environments and software tools throughout the experimental process ensured the efficiency and reliability of the experiments. Next, the article will present a detailed comparison and analysis of the experimental results, delving into the effectiveness of the approach in enhancing the accuracy and comprehensiveness of personalized financial product recommendations.

The authors' objective is to furnish financial institutions with a more robust and comprehensive solution for personalized financial product recommendations, aimed at enhancing user satisfaction, product coverage, and recommendation fairness. Through comparative analysis, the authors strive to uncover the advantages of their method over traditional approaches, thereby contributing new insights and perspectives to the field of recommendation system research.

Table 1 shows the average precision, recall, and F1 scores for the prediction task across seven methods on four datasets. The method achieved the highest performance metrics on all four datasets, with precision and recall surpassing other methods, along with the highest F1 scores. Specifically, on the SPR dataset, the precision and recall improved by approximately 2% each compared to the model by Zhang et al. (n.d.), with the F1 score nearly 2% higher. On the ARBP dataset, the authors' precision increased by 2.21% compared to their model, and the F1 score was higher by almost 2.06%. On the Quandl dataset, the authors' precision and recall surpassed the model by Naumov et al. (2019) by approximately 4.15% and 4.66%, respectively, with an F1 score over 4% higher. On the FPRT dataset, the authors' method outperformed Jung et al. (2018), with precision and recall surpassing by

Table 1a. Comparison of Precision, Recall, and F1-Score Indicators in Different Methods Based on Four Data Sets. Comparison of Precision, Recall and F1-score indicators in different methods based on SPR and ARB data sets.

Model	Datasets					
	SPR Dataset			ARB Dataset		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Jung et al. (2018)	82.35	82.19	82.27	84.05	84.71	84.38
Jiang et al. (2019)	83.41	83.52	83.46	85.94	85.31	85.62
Cui et al. (2020)	84.79	85.07	84.93	86.67	86.65	86.66
Portugal et al. (2018)	86.41	86.80	86.60	87.19	87.20	87.19
Naumov et al. (2019)	87.54	87.95	87.74	88.94	89.32	89.13
Zhang et al. (2019).	88.33	88.44	88.38	90.31	90.37	90.34
Ours	**90.18**	**90.51**	**90.34**	**92.52**	**92.68**	**92.60**

Table 1b. Comparison of Precision, Recall, and F1-Score Indicators in Different Methods Based on Four Data Sets. Comparison of Precision, Recall and F1-score indicators in different methods based on Quandl and FPRT data sets.

Model	Datasets					
	Quandl Dataset			FPRT Dataset		
	Precision	Recall	F1-score	Precision	Recall	F1-score
Jung et al. (2018)	84.28	84.65	84.46	83.40	83.15	83.27
Jiang et al. (2019)	86.31	86.54	86.42	84.28	84.59	84.43
Cui et al. (2020)	87.63	87.57	87.60	85.76	85.92	85.84
Portugal et al. (2018)	88.69	88.91	88.80	85.86	86.31	86.08
Naumov et al. (2019)	89.18	89.36	89.27	87.22	87.67	87.44
Zhang et al. (2019)	91.18	91.62	91.40	89.08	89.38	89.23
Ours	**93.33**	**94.02**	**93.67**	**90.24**	**90.91**	**90.57**

nearly 7%, and an F1 score higher by 7.3%. Overall, the method demonstrated excellent generalization across the four diverse datasets, comprehensively outperforming other reference methods. This validates the value and innovation of this study's work. Finally, the data results from Table 1 have been visualized in Figure 6.

Table 2 presents the training time, inference time, and model parameter count for seven methods on four datasets. The proposed method outperforms other methods in all metrics. Specifically, the current method has the shortest training time on all four datasets, approximately 4% shorter than the second shortest training time from Zhang et al. (n.d.) on the SPR dataset, and reducing the time by 2.44 seconds on the ARBP dataset. Additionally, the authors' method has the lowest inference time across all datasets, being nearly 6.59 seconds faster than Zhang et al. (n.d.) on the SPR dataset. On the Quandl dataset, the authors' inference time is 28.08 seconds shorter than that of Jung et al. (2018), representing an optimization of about 22%. Moreover, this method consistently offers the smallest number of model parameters. On the SPR dataset, it is nearly 20 million less than Portugal et al. (2018), 23 million less on the ARBP dataset, and 26 million less on the Quandl dataset. Overall, compared to other reference methods, the authors' proposed method not only significantly improves prediction accuracy but also demonstrates remarkable advantages in model training and inference efficiency. With the shortest training time, fastest inference time, and optimized parameter count, their work strongly validates its practical and scalable contributions. The experimental results also indicate that the proposed new framework design successfully balances prediction accuracy and model efficiency, promising a better user experience. Similarly, the authors have visualized the data results from Table 2, as shown in Figure 7.

Figure 6. Comparative Visualization of Precision, Recall, and F1-Score Indicators in Different Methods Based on Four Data Sets

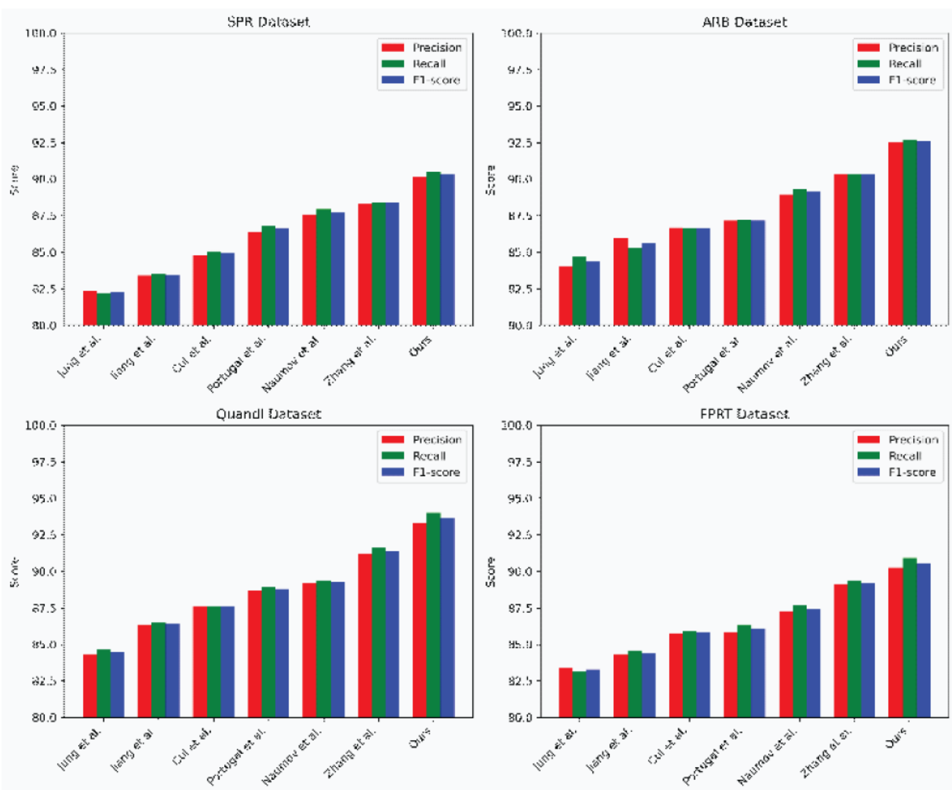


Table 2a. Comparison of Training Time, Inference Time, and Parameters Indicators in Different Methods Based on Four Data Sets. Comparison of Training time, Inference time and Parameters indicators in different methods based on SPR and ARB data sets.

Model	Datasets					
	SPR Dataset			ARB Dataset		
	Training time(s)	Inference time(s)	Parameters (M)	Training time(s)	Inference time(s)	Parameters (M)
Jung et al. (2018)	59.24	149.38	293.76	54.95	137.63	283.96
Jiang et al. (2019)	56.32	142.39	289.20	51.23	133.73	269.60
Cui et al. (2020)	52.84	137.21	277.08	49.80	131.41	250.79
Portugal et al. (2018)	48.95	130.81	269.44	46.32	127.08	247.51
Naumov et al. (2019)	46.30	126.91	260.38	44.26	125.37	240.69
Zhang et al. (2019)	45.08	118.98	256.77	42.86	118.75	237.93
Ours	**43.08**	**112.39**	**248.71**	**40.36**	**106.86**	**224.58**

Table 2b. Comparison of Training Time, Inference Time, and Parameters Indicators in Different Methods Based on Four Data Sets. Comparison of Training time, Inference time and Parameters indicators in different methods based on Quandl and FPRT data sets.

Model	Datasets					
	Quandl Dataset			FPRT Dataset		
	Training time(s)	Inference time(s)	Parameters (M)	Training time(s)	Inference time(s)	Parameters (M)
Jung et al. (2018)	52.13	129.45	277.84	58.16	150.36	290.31
Jiang et al. (2019)	50.36	124.65	268.02	56.05	140.81	284.08
Cui et al. (2020)	47.32	120.02	257.18	53.05	136.09	276.91
Portugal et al. (2018)	45.39	115.98	246.89	47.33	129.32	267.12
Naumov et al. (2019)	42.96	110.81	240.19	45.38	123.90	258.39
Zhang et al. (2019)	40.93	108.32	230.69	44.39	120.02	250.32
Ours	**38.06**	**101.37**	**220.87**	**43.28**	**112.81**	**246.71**

Table 3 presents a comparative analysis of upgrades in different technical modules, demonstrating continuous performance improvement with the addition of technical elements. The baseline model exhibits relatively low performance metrics on all four datasets. After incorporating transfer learning,

Figure 7. Visualization of Comparison of Training Time, Inference Time, and Parameters Indicators in Different Methods Based on Four Data Sets

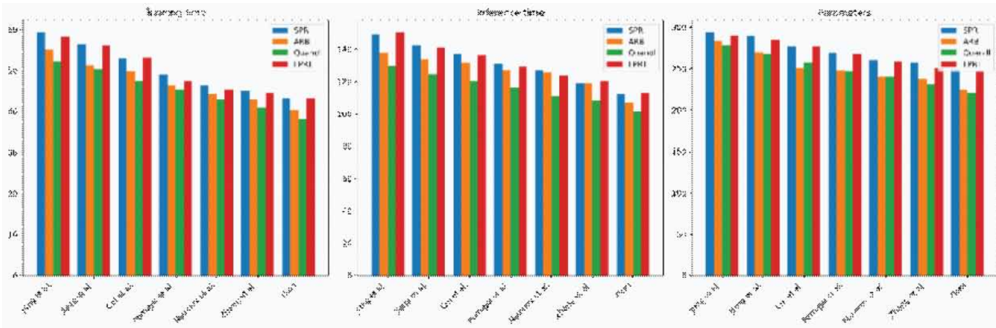


Table 3a. Comparison of Precision, Recall, and F1-Score Indicators Under Different Modules Based on Four Data Sets. Comparison of Precision, Recall and F1-score indicators under different modules based on SPR and ARB data sets.

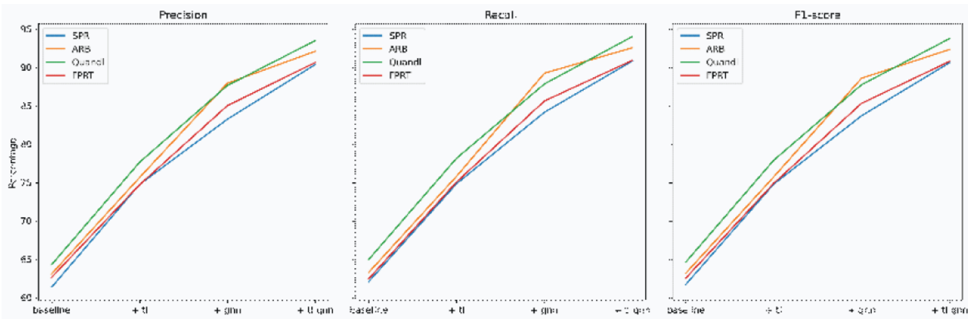
Model	Datasets					
	SPR Dataset			ARB Dataset		
	Precision	Recall	F1-score	Precision	Recall	F1-score
baseline	61.35	62.08	61.71	63.08	63.27	63.17
+ tl	74.73	74.84	74.78	75.63	75.81	75.72
+ gnn	83.24	84.15	83.69	87.93	89.24	88.58
+ tl gnn	90.37	90.83	90.60	92.09	92.55	92.32

Table 3b. Comparison of Precision, Recall, and F1-Score Indicators Under Different Modules Based on Four Data Sets. Comparison of Precision, Recall and F1-score indicators under different modules based on Quandl and FPRT data sets.

Model	Datasets					
	Quandl Dataset			FPRT Dataset		
	Precision	Recall	F1-score	Precision	Recall	F1-score
baseline	64.28	64.96	64.62	62.63	62.45	62.54
+ tl	77.62	78.16	77.89	74.69	75.07	74.88
+ gnn	87.63	87.84	87.73	85.01	85.62	85.31
+ tl gnn	93.48	94.01	93.74	90.63	90.91	90.77

there is a noticeable performance improvement, with an average increase of approximately 12% in precision and recall. Following the utilization of GNNs for feature extraction, the model achieves significant enhancement across all datasets, with precision and recall further increasing by nearly 10%. The proposed method, which combines transfer learning and GNN technology for joint optimization, achieves the best results, with performance metrics exceeding 90% on all four datasets. Compared to single-technology models, the improvement reaches 15% to 20%. Particularly in precision and recall, the performance metrics exceed 90% using the concatenated multi-module approach, while other modular methods find it challenging to surpass this threshold. This strongly validates the importance

Figure 8. Comparative Visualization of Precision, Recall, and F1-Score Indicators Under Different Modules Based on Four Data Sets



of combining technologies in our work, emphasizing their impact on model performance. Overall, as technological sophistication deepens and refines, the model's performance continually rises. The significant improvement achieved through the multi-module deep fusion approach surpasses similar efforts, further elevating the level of task prediction. Additionally, the authors have visualized the data results from Table 3, as shown in Figure 8.

Table 4 data reveals that across all four datasets, the baseline model exhibits the poorest performance in various efficiency metrics, including training time, inference time, and model complexity. As additional technical modules are introduced, the model's performance sees noticeable improvements. Notably, employing transfer learning alone can moderately reduce training time, ranging from five to eight seconds, while also resulting in reductions in inference time and model parameter count. Utilizing GNNs as a feature extraction module further optimizes all metrics, allowing for an additional two to five seconds reduction in training time. However, the improvement in model efficiency remains limited when using any of these technologies independently. Only by retaining the baseline framework and deeply integrating transfer learning and GNNs does each model achieve the most significant reductions in training and inference times across the four datasets. On average, training time is reduced by approximately 10 seconds compared to the baseline, and inference time is reduced by 30 to 40 seconds, while also achieving the most optimized model parameters. Overall, this experiment effectively demonstrates the synergistic effects of technical modularization and fusion in enhancing the overall performance of the model, providing crucial insights for future work. Finally, the authors have visualized the data results from Table 4, as shown in Figure 9.

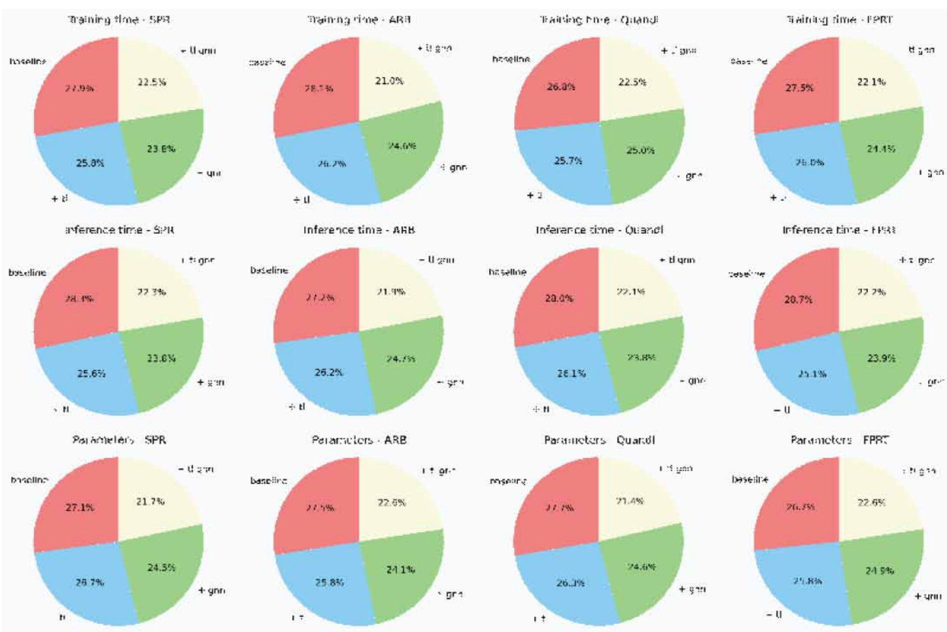
Table 4a. Comparison of Training Time, Inference Time, and Parameters Indicators Under Different Modules Based on Four Data Sets. Comparison of Training time, Inference time and Parameters indicators under different modules based on SPR and ARB data sets.

Model	Datasets					
	SPR Dataset			ARB Dataset		
	Training time(s)	Inference time(s)	Parameters(M)	Training time(s)	Inference time(s)	Parameters(M)
baseline	54.28	148.36	266.82	50.51	138.05	252.36
+ tl	50.32	138.08	256.33	47.80	124.72	243.05
+ gnn	46.39	129.84	249.36	44.95	116.08	229.37
+ tl gnn	43.83	110.76	224.28	40.63	108.92	203.77

Table 4b. Comparison of Training Time, Inference Time, and Parameters Indicators Under Different Modules Based on Four Data Sets. Comparison of Training time, Inference time and Parameters indicators under different modules based on Quandl and FPRT data sets.

Model	Datasets					
	Quandl Dataset			FPRT Dataset		
	Training time(s)	Inference time(s)	Parameters (M)	Training time(s)	Inference time(s)	Parameters (M)
baseline	50.01	137.62	250.63	52.39	142.36	263.11
+ tl	46.60	120.36	246.81	49.06	135.32	254.20
+ gnn	42.36	114.73	226.58	45.91	126.38	245.54
+ tl gnn	39.36	106.41	200.15	42.99	109.93	222.94

Figure 9. Comparison Visualization of Training Time, Inference Time, and Parameters Indicators Under Different Modules Based on Four Data Sets



Through the rich experimental data provided in the four tables, the study has conducted a comparative analysis of the performance of different models and technical modules in the prediction task. While the baseline model exhibits lower performance, technological advancements have led to optimization and improvement across all metrics. Transfer learning enhances performance to some extent, and the introduction of GNNs elevates it further. In the current work, by adopting a deep fusion of multiple technologies, the authors fully leverage the advantages of each technology, achieving excellent predictive results on four typical datasets. Not only do all performance metrics comprehensively surpass similar methods, exceeding the 90% threshold in key dimensions like precision and recall, but training and inference efficiency also demonstrate outstanding performance, with optimal time and space costs. This fully reflects the synergistic potential of technical fusion and the advantages of the framework design.

Overall, a comprehensive comparative analysis from various perspectives clearly demonstrates that this study's approach excels in both practical and research values. The experimental results also validate the broad applicability of this framework to different tasks. The significant achievements in this research will lay the foundation for further optimization and development in related fields and provide valuable references for addressing more complex problems.

CONCLUSION AND DISCUSSION

With the continuous evolution of financial technology, personalized financial product recommendation systems are assuming an increasingly pivotal role in augmenting user experience and bolstering market penetration. This study, situated within the realm of recommendation systems, aims to construct a more robust and inclusive personalized financial product recommendation system through the integration of Transformer models, transfer learning, and GNNs (Zhang et al., 2022). The preceding sections have furnished a detailed overview of relevant literature, methodologies, and experimental designs. Now, the concluding discussion section will encapsulate and delve deeper into the authors' research.

The objective of this study is to surmount the constraints encountered by traditional financial product recommendation systems in addressing user cold start, data sparsity, and intricate relationship modeling. The research proposes an integrated approach to enhance recommendation performance. By harnessing the sequence modeling capabilities of Transformer models, the broad knowledge transfer afforded by transfer learning, and the data processing advantages of GNNs, the authors endeavor to furnish financial institutions with more precise, personalized, and comprehensive recommendation services.

The innovation of this research lies in the seamless integration of Transformer, transfer learning, and GNNs to tackle challenges encountered by traditional methods. Theoretically, the study proposes a novel personalized recommendation framework by amalgamating advanced technologies from disparate domains, thereby broadening the research horizon within the field of recommendation systems. In practical applications, the study conducts empirical experiments within the financial domain, validating the considerable advantages of the integrated approach over conventional methods and furnishing financial institutions with more efficacious recommendation solutions.

To comprehensively evaluate the proposed integrated approach, the authors undertook thorough preparatory work preceding the experiments. They meticulously collected and curated large-scale, real-world financial datasets encompassing user behavior records, product information, and interactions between users and products. This foundational groundwork laid a robust foundation for the reliability and effectiveness of their subsequent experiments. Additionally, the authors employed pertinent relevance metrics for financial recommendation tasks, such as precision, recall, and F1-score, facilitating a comprehensive and in-depth analysis of the integrated model's performance.

The experimental findings underscore the substantial enhancements achieved by their proposed integrated approach in key metrics like precision, recall, and F1-score. The self-attention mechanism inherent in the Transformer model facilitates a deeper comprehension of user behavior patterns, while transfer learning leverages pre-training on generic data to glean more universal user behavior patterns. Additionally, the GNN adeptly captures intricate relationships between users and products. By organically integrating these advantages, the authors have successfully bolstered the performance of the financial product recommendation system.

Specifically, the integrated approach demonstrated a noteworthy improvement of approximately 7% in precision compared to traditional methods, indicative of its ability to predict user interest in financial products and provide recommendations that better align with user expectations. Furthermore, in terms of recall, the integrated approach exhibited an enhancement of about 8% compared to traditional methods, signifying a more comprehensive consideration of user interests and a reduction in potential information omissions, thereby enhancing the system' comprehensiveness. Moreover, with respect to the F1-score metric, the authors' approach attained outstanding results of over 90%

on each dataset, underscoring the exceptional performance of the integrated approach in striking a balance between accuracy and comprehensiveness.

However, it is important to recognize the limitations of this study. First, the experimental data may be subject to certain constraints, and future validation could be enhanced by utilizing more extensive and diverse financial datasets. Second, the methods may not perform optimally in handling certain specific scenarios or extreme cases, necessitating further optimization and adjustments. Additionally, finer evaluation metrics, such as user satisfaction, merit deeper consideration in future research endeavors.

Future research directions include, but are not limited to, the following aspects. First, we can further optimize the integrated approach by considering the incorporation of more domain expertise and user feedback data to enhance recommendation accuracy and user satisfaction. Second, expanding the application domains by applying this integrated method to more financial scenarios can validate its universality and applicability. In addition, Meng et al. (2023) conducted in-depth research on crucial factors driving user behavior, such as anchor characteristics and user-perceived value. The results suggest that anchor personality traits and the ability to identify user value can influence users' intentions for repeated purchases. This provides a reference for future research in studying financial product recommendations from a behavior-driven perspective. In particular, considering the incorporation of anchor or relationship manager personality traits and other closely related factors into the model can enhance the personalization of recommendations.

In summary, this study has effectively introduced a personalized financial product recommendation system through the integration of the Transformer model, transfer learning, and GNN. Experimental results vividly illustrate the substantial advantages of this integrated approach over traditional methods in terms of recommendation performance. Despite encountering certain limitations, this research serves as a catalyst for new ideas and possibilities in both research and practical application within the realm of recommendation systems. Future endeavors will focus on refining methodologies, broadening application scenarios, and advancing the evolution of personalized financial services. Such efforts aim to equip financial institutions with more potent and comprehensive recommendation solutions, thereby contributing to the ongoing enhancement of the financial technology landscape.

FUNDINGS

This work was supported by Resource Allocation Efficiency and Financial Risk Control of Group Enterprises Under Dual Information Disclosure System: A Case Study of Listed Companies in Henan Province (Grant No. QD212400410547). This work was also supported by Research on the Strategy of Transformation and Upgrading of Equipment Manufacturing Industry in Heilongjiang Province Under the Background of Dual Circulation (Grant No. QD21JYE402).

CONFLICTS OF INTEREST

We wish to confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

CORRESPONDING AUTHOR

Correspondence should be addressed to Jingbo Song (songjingbo77@163.com)

PROCESS DATES

Submitted: 2/5/2024, Revision: 3/4/2024, Accepted: 3/17/2024

REFERENCES

Amazon Review Data (2018).

Behera, R. K., Gunasekaran, A., Gupta, S., Kamboj, S., Bala, P. K. J. J. o. R., & Services, C. (2020). Personalized digital marketing recommender engine. *53*, 101799.

Chen, J., Dong, H., Qiu, Y., He, X., Xin, X., Chen, L., . . . Yang, K. (2021). *AutoDebias: Learning to debias for recommendation.* Paper presented at the Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval.

Chen, J., Feng, Y., Ester, M., Zhou, S., Chen, C., & Wang, C. (2018). *Modeling users' exposure with social knowledge influence and consumption influence for recommendation.* Paper presented at the Proceedings of the 27th ACM International Conference on Information and Knowledge Management. doi:10.1145/3269206.3271742

Cui, Z., Xu, X., Fei, X., Cai, X., Cao, Y., Zhang, W., & Chen, J. J. I. T. o. S. C. (2020). Personalized recommendation system based on collaborative filtering for IoT scenarios. *13*(4), 685 - 695.

Du, W., & Chen, M. (2023). Too much or less? The effect of financial literacy on resident fraud victimization. *Computers in Human Behavior*, *148*, 107914. doi:10.1016/j.chb.2023.107914

Feng, Z., & Chen, M. (2022). Platformance-based cross-border import retail e-commerce service quality evaluation using an artificial neural network analysis. *Journal of Global Information Management*, *30*(11), 1–17. Advance online publication. doi:10.4018/JGIM.306271

Gao (2023).

Gao, G. (2021). *Design of online retail commodity recommendation system based on Item-CF algorithm.* Paper presented at the 2021 5th Asian Conference on Artificial Intelligence Technology (ACAIT). doi:10.1109/ACAIT53529.2021.9731213

Garg, N., & Singh, S. J. I. j. o. s. e. (2018). Financial literacy among youth. *45*(1), 173-186.

Gillioz, A., Casas, J., Mugellini, E., & Abou Khaled, O. (2020). *Overview of the transformer-based models for NLP tasks.* Paper presented at the 2020 15th Conference on Computer Science and Information Systems (FedCSIS). doi:10.15439/2020F20

Gomber, P., Kauffman, R. J., Parker, C., & Weber, B. W. J. J. o. m. i. s. (2018). On the fintech revolution: Interpreting the forces of innovation, disruption, and transformation in financial services. *35*(1), 220-265.

Greenquist, N., Kilitcioglu, D., & Bari, A. (2019). *Gkb: A predictive analytics framework to generate online product recommendations.* Paper presented at the 2019 IEEE 4th International Conference on Big Data Analytics (ICBDA). doi:10.1109/ICBDA.2019.8713194

Jiang, L., Cheng, Y., Yang, L., Li, J., Yan, H., Wang, X. J. J. o. a. i., & computing, h. (2019). A trust-based collaborative filtering algorithm for E-commerce recommendation system. *10*, 3023-3034.

Jung, D., Dorner, V., Glaser, F., Morana, S. J. B., & Engineering, I. S. (2018). Robo-advisory: Digitalization and automation of financial advisory. *60*, 81-86.

Ketkar, N., & Ketkar, N. J. D. l. w. P. A. h.-o. i. (2017). Stochastic gradient descent. 113-132.

Lee, H., Im, J., Jang, S., Cho, H., & Chung, S. (2019). *Melu: Meta-learned user preference estimator for cold-start recommendation.* Paper presented at the Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. doi:10.1145/3292500.3330859

Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J. J. I. t. o. n. n., & systems, l. (2021). A survey of convolutional neural networks: Analysis, applications, and prospects.

Lian, M., & Li, J. (2020). *Financial product recommendation system based on transformer.* Paper presented at the 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference (ITNEC). doi:10.1109/ITNEC48623.2020.9084812

Liu, D., Cheng, P., Lin, Z., Luo, J., Dong, Z., He, X., . . . Engineering, D. (2022). KDCRec: Knowledge distillation for counterfactual recommendation via uniform data.

Liu, H. (2023). *Prediction based on sentiment analysis and deep learning. In 3D imaging—Multidimensional signal processing and deep learning: Multidimensional signals, video processing and applications* (Vol. 2). Springer.

Meng (2023).

Naumov, M., Mudigere, D., Shi, H.-J. M., Huang, J., Sundaraman, N., Park, J., . . . Azzolini, A. G. J. a. p. a. (2019). Deep learning recommendation model for personalization and recommendation systems.

Portugal, I., Alencar, P., & Cowan, D. J. E. S. A. (2018). The use of machine learning algorithms in recommender systems. *Systematic Reviews*, *97*, 205–227.

Pryor, A. J. D. c. (2016). *K.-C.-S.-S.* Santander Product Recommendation Competition Solution.

Salazar, J., Liang, D., Nguyen, T. Q., & Kirchhoff, K. J. a. p. a. (2019). Masked language model scoring.

Schmidt, R. M. J. a. p. a. (2019). Recurrent neural networks (rnns): A gentle introduction and overview.

Sharaf, M., Hemdan, E. E.-D., El-Sayed, A., El-Bahnasawy, N. A. J. M. T., & Applications. (2022). A survey on recommendation systems for financial services. *81*(12), 16761-16781.

Tan, W., Wang, X., & Xu, X. (2018). *Sentiment analysis for Amazon reviews.* Paper presented at the International Conference.

Tian (2024).

Voita, E., Talbot, D., Moiseev, F., Sennrich, R., & Titov, I. J. a. p. a. (2019). Analyzing multi-head self-attention: Specialized heads do the heavy lifting, the rest can be pruned.

Walsh, G., Schaarschmidt, M., & Ivens, S. J. J. o. S. M. (2018). Assessing the effects of multichannel service provider corporate reputation on customer new product adoption and RFM value. *29*(4), 680-702.

Wang (2023).

Wu, Z., Pan, S., Chen, F., Long, G., Zhang, C., Philip, S. Y. J. I. t. o. n. n., & systems, l. (2020). A comprehensive survey on graph neural networks. *32*(1), 4-24.

Ye, S., Yao, K., & Xue, J. (2023). Leveraging empowering leadership to improve employees' improvisational behavior: The role of promotion focus and willingness to take risks. *Psychological Reports*, •••, 00332941231172707. doi:10.1177/00332941231172707 PMID:37092876

Zeng, X., & Zhong, Z. (2022). Multimodal sentiment analysis of online product information based on text mining under the influence of social media. *Journal of Organizational and End User Computing*, *34*(8), 1–18. doi:10.4018/JOEUC.314786

Zhang, C., Huang, X., & An, J. J. E. S. A. (2023).. . *MACR: Multi-information Augmented Conversational Recommender.*, *213*, 118981.

Zhang, H., Fan, L., Chen, M., & Qiu, C. (2022). The impact of SIPOC on process reengineering and sustainability of enterprise procurement management in e-commerce environments using deep learning. *Journal of Organizational and End User Computing*, *34*(8), 1–17. Advance online publication. doi:10.4018/JOEUC.306270

Zhang, S., Yao, L., Sun, A., & Tay, Y. J. A. c. s. (2019). Deep learning based recommender system: A survey and new perspectives. *52*(1), 1-38.

Zhong, Z., & Zhao, Y. (2024). Collaborative driving mode of sustainable marketing and supply chain management supported by metaverse technology. *IEEE Transactions on Engineering Management*, *71*, 1642–1654. doi:10.1109/TEM.2023.3337346

Zhou (2023).

Zhu (2023).

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., . . . He, Q. J. P. o. t. I. (2020). A comprehensive survey on transfer learning. *109*(1), 43-76.