



基于改进Transformer和超图模型的股票趋势预测方法研究

郝剑龙, 刘志斌, 张宸, 孙琪炜, 常新功

引用本文:

郝剑龙, 刘志斌, 张宸, 孙琪炜, 常新功. 基于改进Transformer和超图模型的股票趋势预测方法研究[J]. 智能系统学报, 2024, 19(5): 1126–1135.

HAO Jianlong, LIU Zhibin, ZHANG Chen, et al. Stock trend prediction method based on improved Transformer and hypergraph model[J]. *CAAI Transactions on Intelligent Systems*, 2024, 19(5): 1126–1135.

在线阅读 View online: <https://dx.doi.org/10.11992/tis.202308017>

您可能感兴趣的其他文章

用于关系抽取的注意力图长短时记忆神经网络

Attention graph long short-term memory neural network for relation extraction

智能系统学报. 2021, 16(3): 518–527 <https://dx.doi.org/10.11992/tis.202008036>

地理位置和时间感知的表示学习框架

A geography and time aware representation learning framework

智能系统学报. 2021, 16(5): 909–917 <https://dx.doi.org/10.11992/tis.202104011>

结合卷积特征提取和路径语义的知识推理

Knowledge-based inference on convolutional feature extraction and path semantics

智能系统学报. 2021, 16(4): 729–738 <https://dx.doi.org/10.11992/tis.202008007>

标签传播时间序列聚类的股指期货套期保值策略研究

Research on the stock index futures hedging strategy using label propagation time series clustering

智能系统学报. 2019, 14(2): 288–295 <https://dx.doi.org/10.11992/tis.201707023>

基于递归神经网络的风暴潮增水预测

Prediction of storm surge based on recurrent neural network

智能系统学报. 2017, 12(5): 640–644 <https://dx.doi.org/10.11992/tis.201706015>

一种基于知识树和约束的柔性活动动态细化方法

A dynamic refinement approach for flexible activity based on knowledge tree and constraints

智能系统学报. 2017, 12(2): 158–165 <https://dx.doi.org/10.11992/tis.201603009>

DOI: 10.11992/tis.202308017

网络出版地址: <https://link.cnki.net/urlid/23.1538.TP.20240828.0933.008>

基于改进 Transformer 和超图模型的股票趋势预测方法研究

郝剑龙, 刘志斌, 张宸, 孙琪炜, 常新功

(山西财经大学 信息学院, 山西 太原 030006)

摘要: 股票预测是一项令人痴迷又极具挑战的任务。近年来, 融合关系信息的股票时序预测方法取得一些进展, 但仍存在如下问题: 首先, 基于图神经网络的方法仅考虑股票之间简单的成对关系, 而未考虑股票间的高阶协同关系。其次, 现有方法采用预定义图的方式直接给出股票间的静态关系, 无法建模股票间潜在的动态变化关系。为了解决上述问题, 提出一种端到端的动态超图卷积神经网络股票趋势预测框架。该框架基于改进的 Transformer 提取股票的时序信息, 通过静态超图和动态超图将股票间的协同关系信息引入到时序建模中。在中国 A 股和美股市场数据集上的实验结果表明, 与当前先进模型相比, 本文模型的预测性能具有显著优势。

关键词: Transformer; 趋势感知; 注意力机制; 动态超图; 协同关系; 股票趋势预测; 时序预测; 混合模型

中图分类号: TP301.6 **文献标志码:** A **文章编号:** 1673-4785(2024)05-1126-10

中文引用格式: 郝剑龙, 刘志斌, 张宸, 等. 基于改进 Transformer 和超图模型的股票趋势预测方法研究 [J]. 智能系统学报, 2024, 19(5): 1126-1135.

英文引用格式: HAO Jianlong, LIU Zhibin, ZHANG Chen, et al. Stock trend prediction method based on improved Transformer and hypergraph model[J]. CAAI transactions on intelligent systems, 2024, 19(5): 1126-1135.

Stock trend prediction method based on improved Transformer and hypergraph model

HAO Jianlong, LIU Zhibin, ZHANG Chen, SUN Qiwei, CHANG Xingong

(School of Information, Shanxi University of Finance and Economics, Taiyuan 030006, China)

Abstract: Stock forecasting is an interesting but extremely difficult task. Stock time series prediction methods that incorporate relationship information have progressed in recent years, but the following issues remain. First, graph neural network-based methods consider simple pairwise relationships between stocks but ignore higher-order collaborative relationships. Second, most existing methods utilize the static relationship among stocks with predefined graphs, and modeling the potential changes in dynamic relationships among stocks is challenging. To address the abovementioned issues, a Dynamic HyperGraph Convolutional neural Network (DHGCN) framework for the end-to-end stock trend prediction is proposed. The temporal information of stocks is captured by an improved Transformer model, and the collaborative relationship information is integrated into the time series modeling by the static and dynamic hypergraphs. Experiments on the real-world datasets of the Chinese A-share market and the US stock market show that the prediction performance of the proposed model is significantly superior to that of the contemporary advanced models.

Keywords: Transformer; trend forecasting; attention mechanism; dynamic hypergraph; collaborative relationship; stock trend prediction; time series forecasting; hybrid model

收稿日期: 2023-08-16. 网络出版日期: 2024-08-28.

基金项目: 国家自然科学基金青年项目 (62003198); 山西省基础研究计划自然科学面上项目 (202203021221218).

通信作者: 郝剑龙. E-mail: haojianlong2012@sxufe.edu.cn.

©《智能系统学报》编辑部版权所有

随着中国资本市场的发展, 股票投资逐渐进入大众视野。股票投资的相关研究对于个人和投资机构而言, 可以通过预测分析协助投资者作出

更加理性的决策, 获得可观的投资回报; 对于国家和社会而言, 可以通过探寻证券市场风险波动的内在机理, 为政府监管部门的政策制定提供智能决策辅助, 防患化解重大金融风险。因此, 对股票市场进行预测分析具有重要的现实意义。

早期的股票预测方法主要基于历史股价序列, 通过传统统计学方法构建自回归模型。然而, 这类方法对于高度非线性、非平稳的金融数据等拟合能力较差, 比如常见的差分整合自回归移动平均 (auto-regressive integrated moving average, ARIMA) 模型在小规模的可平稳化处理的时间序列预测任务中表现优异, 但是模型超参数的确定需要人工参与, 并且得出的预测结果往往滞后于实际的价格变动, 大大限制了模型在实际场景中的应用^[1]。随着大数据, 人工智能等技术的不断发展, 基于机器学习的预测方法受到了广泛的研究和关注^[2-3]。

近年来, 随着计算机算力的进一步提升以及可用于训练学习的数据爆发式增长, 深度学习技术可以利用多层人工神经网络实现任意非线性函数的逼近, 其在图像处理、语音识别、自然语言处理等领域的性能逐渐超越了传统的机器学习方法。例如 Hoseinzade 等^[4]提出了一种基于卷积神经网络的模型, 提取信号不同尺度上的抽象特征, 丰富了信号的特征表示, 实验结果表明, 深度学习方法优于浅层方法。Zhang 等^[5]提出了基于状态频率内存 (state frequency memory, SFM) 的方法, 将长短期记忆网络 (long short-term memory, LSTM) 存储单元的隐藏状态分解为多个频率分量, 以发现多频交易模式。Ding 等^[6]提出的融合多尺度高斯先验的 Transformer^[7]模型在股票趋势预测中取得了显著的结果。

但是上述股票预测方法仅考虑单一的股票时间序列, 忽视了股票之间复杂的关系。研究表明, 金融市场时间序列概念背后的空间维度, 即股票间关系信息, 在很大程度上会影响模型预测的准确性^[8]。鉴于金融市场数据信息的非欧几里得特性, 研究人员引入了图神经网络方法来学习实体间的分布表示。图神经网络 (graph neural network, GNN)^[9-10]是建模实体间关系最常见模型之一, 它采用邻接矩阵来表示实体之间成对相关, 并用于分析未来趋势。在股票预测建模中, 图中的顶点和边分别表示股票及其关系, 通过顶点表示学习来捕获每对股票之间的关系。例如, Chen 等^[11]根据市场投资事件构建了公司实体关系图, 并通过图卷积神经网络整合公司股票间关系信息, 获得了更准确的预测结果。Feng 等^[12]收集行业关系元数据和公司股权信息构建静态关系

图, 并通过 LSTM 来学习序列特征, 然后将学习到的序列特征输入到图卷积注意力网络对股票进行排名预测, 提高了模型预测效果。

然而, 在真实市场中, 股票之间的关系往往比成对关系更复杂, 而普通图结构只能描述两点之间的关系, 无法对高阶关系进行建模。例如“行业内联动”现象: 同一行业内的相关个股受到同一外部事件冲击时, 股价变化具有同一的趋势^[13]。“行业间联动”现象: 在供应链中, 上游公司的经营调整会对相关下游公司造成影响, 在股价上也会表现出对应的变化趋势。这些协同关系自然地将股票作为集体联系在一起, 而不是成对关系。GNN 不足以对股票之间的这种协同关系进行建模。作为图的扩展, 超图具有比图更强的关系建模能力^[14]。超图引入关联矩阵来表示这种协同关系, 顶点和超边分别是关联矩阵的行索引和列索引。在超图中, 超边可以连接属于同一组内的多支股票, 表明它们具有共同的内在属性。同时, 一支股票可以连接到多个超边, 这意味着它具有多个特征。超图模型已经在许多领域取得了进展, 包括物体识别^[15]、交通预测^[16]和推荐系统^[17]等。在股票预测方面, Sawhney 等^[18]使用超图学习来捕获股票间的复杂关系。Cui 等^[19]提出了超图注意力网络用于股票趋势预测。Li 等^[20]提出了一种基于超图的股票投资组合强化学习方法, 通过嵌入超图注意力模块来显式建模股票之间的高阶关系从而优化投资选择。

目前基于图或超图神经网络模型的股票预测方法已经取得了不错的效果, 但是仍存在 2 个待解决的问题: 1) 现有图方法仅对股票进行成对关系的建模, 忽略了股票间的高阶协同关系。2) 现有图或超图的方法仅通过预定义信息给出了股票之间的静态关系, 而未对股票随时间变化的动态关系进行建模。为了解决上述问题, 本文提出了一种基于动态超图卷积神经网络的股票趋势预测框架。在该框架中, 首先基于改进的 Transformer 模型编码股票时序信息, 增强模型对股票的短期趋势感知能力。然后, 通过构建静态超图和动态超图将股票间协同关系信息引入到时序建模中。最后, 预测层融合静态超图和动态超图的信息进行趋势预测。

1 相关理论

1.1 超图建模

在使用超图 $\mathcal{G} = (\mathbf{V}, \mathbf{E}, \mathbf{W})$ 建模时, 单支股票由超图中的顶点 v 表示, 且 $v \in \mathbf{V}$, 股票间的多元关系由超边 e 表示。超边集中的每条超边 $e \in \mathbf{E}$ 都被赋

予一个正的权重 W_{ee} , 初始时权重都为 1, 表示每条超边权重相同, $W = I \in \mathbf{R}^{|V| \times |E|}$ 。文献 [21] 已经证明, 当且仅当每条超边连接 2 个顶点时, 超图退化为普通图。

1.2 超图关联矩阵

超图中的关联矩阵揭示了顶点和超边之间的关系, 对于一个没有孤立点的无向超图的关联矩阵 $H \in \mathbf{R}^{|V| \times |E|}$, 其定义为

$$H_{ij} = \begin{cases} 1, & v_i \in e_j \\ 0, & v_i \notin e_j \end{cases}$$

对于超图中一个顶点, 其度被定义为与该顶点相关联的所有超边权重之和:

$$D_{ii} = \sum_{e=1}^{|E|} W_{ee} H_{ie}$$

超边的度被定义为

$$B_{ee} = \sum_{i=1}^{|V|} H_{ie}$$

式中 $D \in \mathbf{R}^{|V| \times |V|}$ 和 $B \in \mathbf{R}^{|E| \times |E|}$ 都是对角矩阵。

2 模型框架

本文提出了一种改进 Transformer 和动态超图卷积神经网络的框架, 用于端到端的股票趋势预测。该框架包括时序特征提取、关系建模和趋势预测 3 个部分, 如图 1 所示。

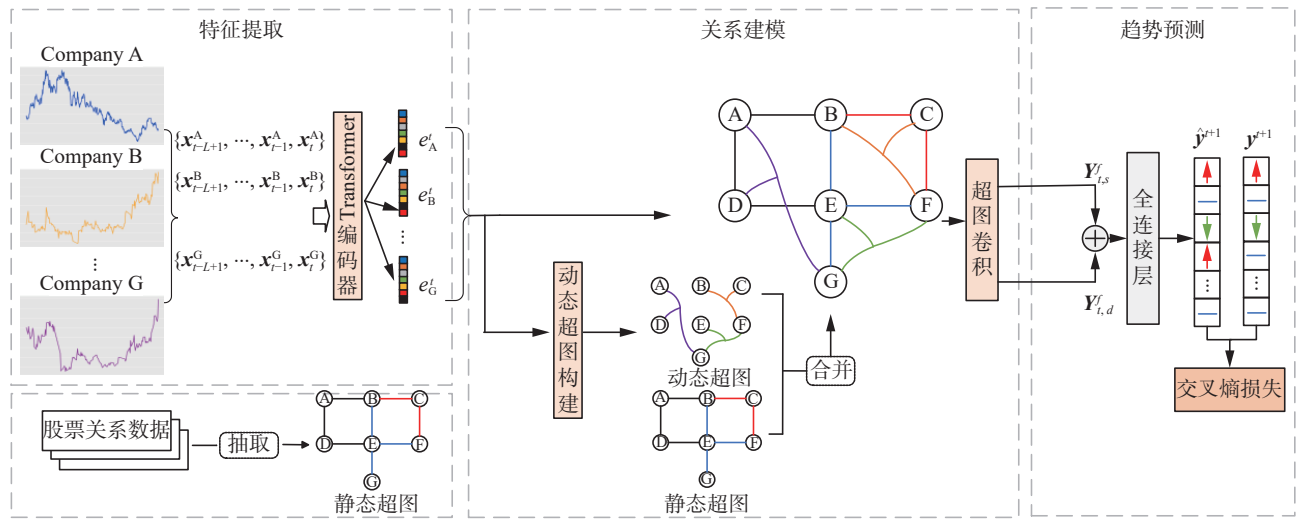


图 1 基于改进 Transformer 和动态超图卷积神经网络的整体框架

Fig. 1 Overall framework based on improved Transformer and dynamic hypergraph convolutional neural network

2.1 问题描述

给定含有 N 支股票的股票集 $S = \{s_1, s_2, \dots, s_N\}$, 在 t 时收集每支股票过去 L 天交易记录 $X_t \in \mathbf{R}^{N \times L \times F}$ 。对于 $X_t^{s_i} = \{[x_{t-L+1}^{s_i}, \dots, x_{t-1}^{s_i}, x_t^{s_i}], s_i \in S\} \in \mathbf{R}^{L \times F}$, 其中 $x_t^{s_i} \in \mathbf{R}^F$ 是股票 s_i 在第 t 个交易日时的价格属性集合, 例如开盘价、收盘价等, F 是原始特征的数量。

股票趋势预测是一个典型的分类预测任务, 给定历史交易记录 $X \in \mathbf{R}^{N \times L \times F}$ 和由行业关系构造的静态超图 \mathcal{G}_s , 模型旨在学习一个函数 $f(X, \mathcal{G}_s, \theta)$, θ 是模型参数, 函数 $f(\cdot)$ 将 X 映射为第 $t+1$ 个交易日的涨跌预测概率分布。

2.2 时序特征提取

股票的历史价格记录在预测其未来趋势方面起着关键作用 [22]。本文设计了一种只有编码器结构的 Transformer 变体从每支股票的时间序列价格数据中捕获其时间动态特征。

Transformer 编码器通过多头自注意力机制对序列的长期依赖关系进行建模取得较好效果, 但由于标准的自注意力机制的计算过程是基于点对

点进行查询 (query)-键 (key) 匹配, 这可能存在 2 个问题: 1) 标准的自注意力计算可能受局部噪声点影响, 导致学习到的特征嵌入掺杂了噪声点的信息。2) 股票时间序列往往存在周期性趋势, 而点对点的计算无法感知短期的周期性趋势。如图 2 所示, 实线框标注出宁德时代 (300750.SZ, 简称 CATL) 公司股票在一段时间内出现相似的波动, 这对于未来的股票趋势预测具有重要的意义。



图 2 宁德时代股票价格周期性波动规律示例

Fig. 2 Example description of the periodic fluctuation pattern of CATL's stock price

为解决上述问题, 本文通过设计趋势感知模块来增强 Transformer 模型对于短期趋势的感知能力。具体来说, 在日期 t 时 N 支股票过去 L 天的历史交易记录 $X_t \in \mathbf{R}^{N \times L \times F}$ 经过位置编码后得到 $U_t \in \mathbf{R}^{N \times L \times F_p}$, 将其作为多头注意力模块的输入。为便于说明, 以多头注意力处理某支股票 i 为例, 股票 i 经位置编码后的输入特征为 $U_t^i \in \mathbf{R}^{L \times F_p}$, 首先将 U_t^i 变换成 R 个不同的值 (value) 矩阵 $V_{t,r}^i = U_t^i W_r^v$, 其中 $W_r^v \in \mathbf{R}^{F_p \times F_p}$ 表示可学习的参数矩阵, $r \in \{1, 2, \dots, R\}$ 表示多头注意力中的头索引。为了使输出的短期趋势特征和状态特征 U_t^i 窗口长度对齐, 如图 3 所示, 在卷积操作之前进行填充操作, 然后使用大小为 $1 \times c$ 一维卷积核沿时序维度提取短期趋势特征, 得到短期趋势特征 $T_t^i \in \mathbf{R}^{L \times F_p}$, 公式表示为

$$T_t^i = \text{1d_conv}(\text{padding}(U_t^i))$$

式中: 短期趋势特征 T_t^i 表示当前股票价格特征 U_t^i 的前 c 日股价趋势, 将二者拼接起来 $[U_t^i : T_t^i]$, 经过线性变化 W_r^q 和 W_r^k 得到 R 个不同的 Query 矩阵和 Key 矩阵, 公式表示为

$$Q_{t,r}^i = [U_t^i : T_t^i] W_r^q$$

$$K_{t,r}^i = [U_t^i : T_t^i] W_r^k$$

式中 W_r^q 和 $W_r^k \in \mathbf{R}^{2F_p \times 2F_p}$ 。最后由 $Q_{t,r}^i, K_{t,r}^i \in \mathbf{R}^{L \times 2F_p}$ 和 $V_{t,r}^i \in \mathbf{R}^{L \times F_p}$ 计算多头注意力得到输出 $O_{t,r}^i$:

$$O_{t,r}^i = \text{softmax}\left(\frac{Q_{t,r}^i K_{t,r}^{i^T}}{\sqrt{2F_p}}\right) M V_{t,r}^i$$

式中 $M \in \mathbf{R}^{L \times L}$ 是非零元素都为 1 的下三角掩码矩阵。最后平均池化所有头输出得到 $Z_t^i \in \mathbf{R}^{L \times F_p}$:

$$Z_t^i = \frac{1}{R} \sum_{r=1}^R O_{t,r}^i$$

同理 N 支股票实体最终的编码输出是 $Z_t \in \mathbf{R}^{N \times L \times F_p}$ 。

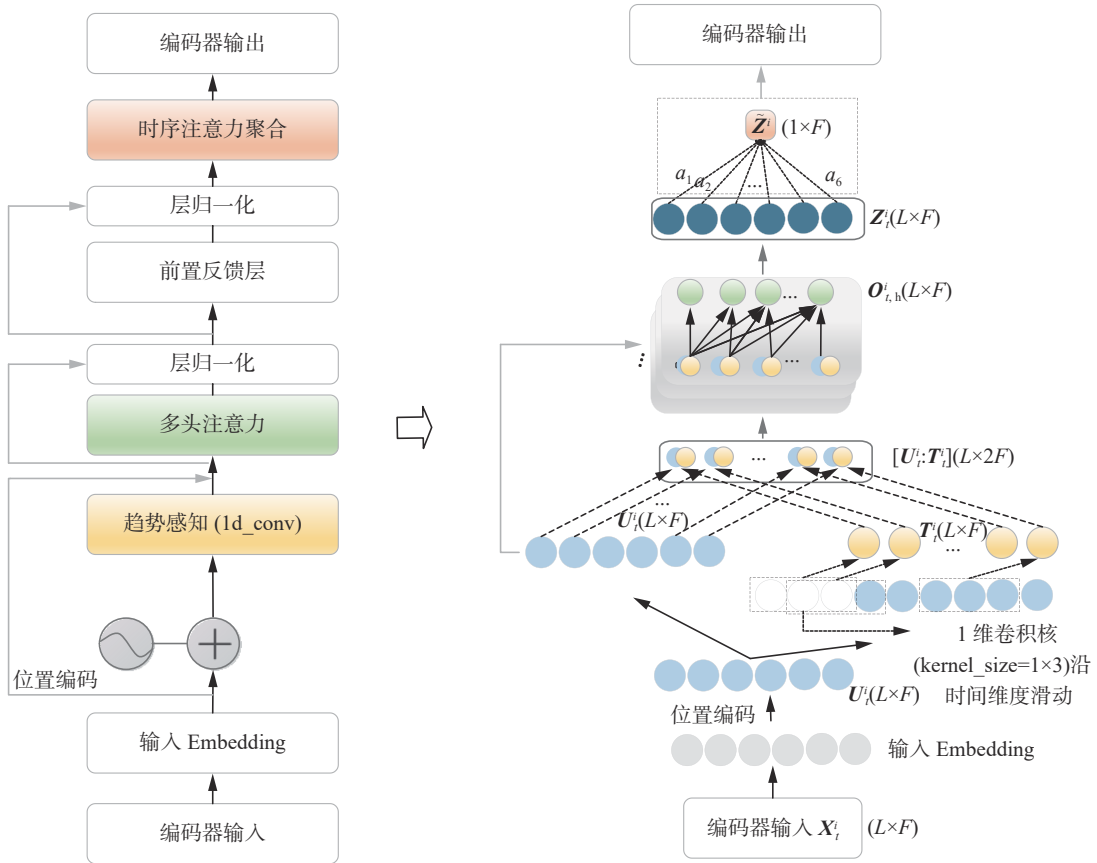


图 3 改进的 Transformer 编码器整体结构

Fig. 3 Overall structure of improved Transformer encoder

由于股票存在日历效应^[19], 即过去不同时刻的价格信息对预测股票未来趋势的重要性不同, 为此把多头注意力的模块输出作为时序注意力聚合模块的输入, 有选择地强调过去交易日的关键状态。具体而言, 某个交易日的状态重要性是通过其与 Query 向量进行相似度衡量得出的, 选择

最近一个交易日的状态作为 Query 向量, 则时间注意力权重 α_t 定义为

$$\alpha_t = \frac{\exp(\text{sim}(z_t^i, z_L^i))}{\sum_{m=0}^{L-1} \exp(\text{sim}(z_m^i, z_L^i))}$$

以股票 i 为例, 特征为 $\mathbf{z}^i = \mathbf{Z}_i[i, :, :]$, 股票 i 第 L 天状态特征为 $\mathbf{z}_L^i = \mathbf{z}^i[L-1, :]$, 式中 $\text{sim}(\mathbf{z}_t^i, \mathbf{z}_L^i)$ 为相关度评分函数, 用于衡量在时间 t 状态特征 $\mathbf{z}_t^i \in \mathbf{R}^{F_p}$ 与 Query 向量 $\mathbf{z}_L^i \in \mathbf{R}^{F_p}$ 相关程度, 定义为

$$\text{sim}(\mathbf{z}_t^i, \mathbf{z}_L^i) = (\mathbf{P}_q \mathbf{z}_t^i)^T (\mathbf{P}_q \mathbf{z}_L^i)$$

最后, 对线性变换后状态特征 $\mathbf{P}_v \mathbf{z}_t^i$ 进行加权聚合, 得到新的状态特征 $\tilde{\mathbf{z}}^i$, 表示股票 i 在时间窗口大小为 L 的全局时间动态特征:

$$\tilde{\mathbf{z}}^i = \sum_{t=0}^{L-1} \alpha_t \mathbf{P}_v \mathbf{z}_t^i$$

上述 \mathbf{P}_q 、 \mathbf{P}_k 和 \mathbf{P}_v 是需要学习的参数矩阵, 同理可得 N 支股票经过时序注意力聚合模块输出 $\tilde{\mathbf{Z}}_t \in \mathbf{R}^{N \times F_p}$ 。

2.3 超图构建

上述时序建模将不同股票视为相互独立的个体, 然而, 股票之间存在着密切关系, 并且关系密切的股票集合更容易出现一定的波动规律。因此, 本节通过构建超图来挖掘股票间成组的协同关系。

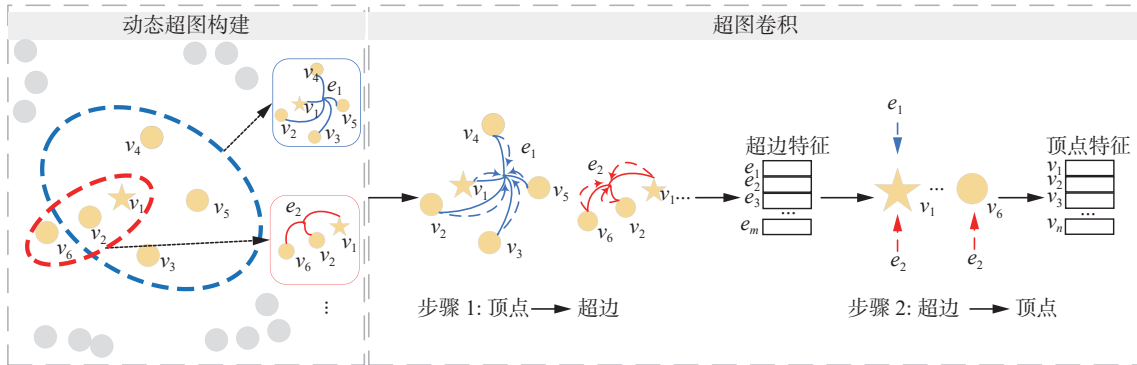


图 4 动态超图构建和超图卷积

Fig. 4 Dynamic hypergraph construction and hypergraph convolution

本文通过以下步骤创建动态超图:

1) 使用改进的 Transformer 编码后的时序信息 $\tilde{\mathbf{Z}}_t \in \mathbf{R}^{N \times F_p}$ 作为动态超图构建模块的输入, 对于每条超边 e , 将其特征初始化为每个顶点 v 的嵌入表示, 即初始化时有 N 条超边;

2) 每个顶点寻找离自身最近的一条超边, 排除使用自身特征初始化的超边;

3) 每个顶点加入离自己最近的那条超边, 同时删除没有包含任何顶点的超边, 各条超边加入其特征初始化所使用的顶点。

在每个时间步都执行上述流程, 当股票特征嵌入随着网络的加深而不断演化时, 超边集也在动态调整。

2.4 超图卷积

基于空间域图理论的超图卷积将超图学习视

2.3.1 静态超图构建

属于同一行业的股票往往会受到相同外部事件的影响, 它们的价格往往会呈现相似的趋势。使用一条超边 $e \in E$, 连接同一个行业的所有股票, 这样由不同行业构成的多条超边形成超边集, 即静态超图。

2.3.2 动态超图构建

动态超图用来捕获公司间动态变化的关联关系。为便于说明定义如下符号: $p(e)$ 表示某条超边 e 所拥有的顶点组成的集合, $a(v)$ 表示顶点 v 所加入的超边组成的集合, 分别表示为

$$p(e) = \{v_1, v_2, \dots, v_{n_e}\}$$

$$a(v) = \{e_1, e_2, \dots, e_{m_v}\}$$

式中: n_e 和 m_v 分别表示超边 e 中的顶点个数和包含顶点 v 的超边数目。

由图 4 可知, 对于顶点 $v_2 \sim v_5$ 而言, 它们的最近点是 v_1 (图中的星状顶点), 这 5 个点构成一条超边 (蓝色虚线框)。同理对于 v_6 和 v_1 , 它们的最近点是 v_2 , 这 3 个点构成一条超边 (红色虚线框)。

为存在相邻关系的相关顶点之间的信息交换^[21]。超图空间域卷积通过一个闭环的消息传递循环机制将时序信息 $\mathbf{X}_t^{(0)}$ 更新为新特征 $\mathbf{X}_t^{(i+1)}$, 超图卷积的初始输入 $\mathbf{X}_t^{(0)} = \tilde{\mathbf{Z}}_t \in \mathbf{R}^{L \times F_p}$ 。闭环的消息传递循环机制涉及 2 个阶段定向的消息流传播, 通过矩阵形式描述如下:

第 1 阶段, $\mathbf{U}_t^{(0)} = \mathbf{W}\mathbf{B}^{-1}\mathbf{H}^T\mathbf{X}_t^{(0)}$ 表示每个顶点特征 $\mathbf{X}_t^{(0)} \in \mathbf{R}^{N \times F_p}$ 汇聚为超边特征 $\mathbf{U}_t^{(0)} \in \mathbf{R}^{|E| \times F_p}$ 。

第 2 阶段, $\mathbf{X}_t^{(i+1)} = \sigma(\mathbf{D}^{-1}\mathbf{H}\mathbf{U}_t^{(0)}\boldsymbol{\theta})$ 表示每个顶点通过汇聚超边特征进行特征更新, $\sigma(\cdot)$ 表示激活函数。

超图卷积的完整更新规则为

$$\mathbf{X}_t^{(i+1)} = \text{LeakyReLU}(\mathbf{D}^{-1}\mathbf{H}\mathbf{W}\mathbf{B}^{-1}\mathbf{H}^T\mathbf{X}_t^{(0)}\boldsymbol{\theta})$$

式中: $\boldsymbol{\theta}$ 是一个可学习的参数矩阵, LeakyReLU 是

修正线性单元激活函数, D 、 B 、 H 和 W 已在上文第 1 节相关理论中定义。

2.5 融合注意力机制的超图卷积

仅通过关联矩阵的元素 0 或 1 来决定股票直接隶属哪个分组缺乏合理性, 即便可以确定隶属哪个分组, 但其隶属度是动态变化的。如图 5 所示, 根据股票机构的分类标准, 股票 300750.SZ 属于电池 (BK1033) 和新能源 (399808.SZ) 板块, 但其股价波动规律不仅与电池和新能源板块相似, 还与汽车 (BK1029) 板块相似。

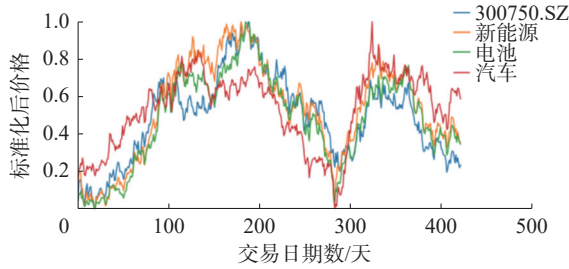


图 5 宁德时代股票价格波动与相关行业价格波动规律示例

Fig. 5 Example description of the fluctuation pattern of CATL's stock price and related sectors

本节引入融合注意力机制的超图卷积, 通过隶属度来详细、合理地描述股票之间随时间演化的关系。首先对关联矩阵 H 嵌入注意力机制^[23], 通过超边-顶点注意力机制来自适应地学习同一超边内不同顶点的重要性。给定某个顶点 $v_i \in V$ 及其所加入的某条超边 $e_j \in E$, 使用顶点 v_i 的时序特征 \tilde{z}^i 以及超边 e_j 的特征 \tilde{m}^j 来计算一个注意力系数 H_{ij} , 该系数表示顶点 v_i 对于超边 e_j 的重要程度。

超边 e_j 的特征 \tilde{m}^j 通过以下步骤计算。首先计算超边平均特征 \tilde{c} , 它是由超边 e_j 包含的顶点嵌入平均池化和非线性变换得到:

$$\tilde{c} = \tanh \left(\left(\frac{1}{|p(e_j)|} \sum_{i=1}^{|p(e_j)|} \tilde{z}^i \right) W_c \right)$$

式中 W_c 是可学习的参数矩阵。基于超边平均特征 \tilde{c} , 可以计算该条超边内每个顶点的注意力权重。对于超边包含的每个顶点, 计算每个顶点特征和超边平均特征 \tilde{c} 之间的余弦相似度, 顶点越相似于超边平均特征 \tilde{c} 获得的权重越高。最后, 超边 e_j 特征是其包含的顶点嵌入加权和 \tilde{m}^j :

$$\tilde{m}^j = \sum_{i=1}^{|p(e_j)|} a_i \tilde{z}^i = \sum_{i=1}^{|p(e_j)|} \left(\frac{\cos(\tilde{z}^i, \tilde{c})}{\sum_{k=1}^{|p(e_j)|} \cos(\tilde{z}^k, \tilde{c})} \right) \tilde{z}^i$$

其余各条超边的特征计算同理。最后由顶点特征 \tilde{z}^i 和超边特征 \tilde{m}^j 得到:

$$H_{ij}^{\text{att}} = \frac{\exp(\sigma(\text{sim}(\tilde{P}\tilde{z}^i, \tilde{P}\tilde{m}^j)))}{\sum_{k \in N_i} \exp(\sigma(\text{sim}(\tilde{P}\tilde{z}^k, \tilde{P}\tilde{m}^j)))}$$

式中: $\sigma(\cdot)$ 为修正线性单元激活函数, \tilde{P} 是可学习参数矩阵。 N_i 表示顶点 v_i 的邻居集, 它可以通过上文定义的超图关联矩阵来获取。 $\text{sim}(\cdot)$ 是相似性计算函数, $\text{sim}(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{r}^T[\mathbf{x}_i \parallel \mathbf{x}_j]$, \mathbf{r} 是可学习的权重向量, $[\cdot \parallel \cdot]$ 表示 2 个向量拼接。此时关联矩阵的元素是介于 0 到 1 的连续值, 不再是离散值。本节使用 H^{att} 替换超图卷积更新公式中的 H 。

静态超图卷积和动态超图卷积采用相同的卷积方式, 二者唯一区别在于关联矩阵 H 不同。2 种不同类型关联矩阵应用超图卷积后输出 $X_{t,s}^{(l+1)}$ 和 $X_{t,d}^{(l+1)}$, 经过线性变换和 LeakyReLU 函数得到 2 种股票趋势表征, 分别是静态趋势表征 $Y_{t,s}$ 和动态趋势表征 $Y_{t,d}$, 公式表示为

$$Y_{t,s} = \sigma(X_{t,s}^{(l+1)} W_s)$$

$$Y_{t,d} = \sigma(X_{t,d}^{(l+1)} W_d)$$

2.6 趋势预测层

将 2 种不同的股票表征 $Y_{t,s}$ 和 $Y_{t,d}$ 相加得到 $Y_t = Y_{t,s} + Y_{t,d}$, 并将 Y_t 输入全连接层得到预测概率分布:

$$\hat{\mathbf{y}}^{t+1} = \text{FC}(Y_t) = \text{FC}([Y_{t,s} + Y_{t,d}])$$

采用交叉熵损失函数作为优化目标:

$$\mathcal{L}(\hat{\mathbf{y}}_i^{t+1}, \mathbf{y}_i^{t+1}) = - \sum_b \mathbf{y}_{i,b}^{t+1} \ln \hat{\mathbf{y}}_{i,b}^{t+1}$$

式中: $\hat{\mathbf{y}}_i^{t+1}$ 是股票 s_i 在接下来的 $t+1$ 交易日中预测概率分布, 选择概率最大的作为预测的趋势方向; \mathbf{y}_i^{t+1} 是表示股票 s_i 真实趋势方向的独热向量; $\mathbf{y}_{i,b}^{t+1}$ 和 $\hat{\mathbf{y}}_{i,b}^{t+1}$ 分别是 \mathbf{y}_i^{t+1} 和 $\hat{\mathbf{y}}_i^{t+1}$ 中第 b 个元素。根据股票收盘价的涨跌幅得到标签, 将涨幅不小于 0.55% 的标记为“上涨”, 跌幅不小于 0.50% 的标记为“下跌”, 其余为“持平”。

3 实验及结果分析

3.1 数据集

采用中国 A 股股票数据集和美股股票数据集来评估本文模型在不同市场下的鲁棒性。

A 股数据集 收集了 2013 年 1 月 4 日—2019 年 12 月 31 日期间 A 股市场中的 2 433 支股票, 通过保留交易日超过 98% 的股票, 最终筛选得到 758 支股票^[19]。收集的 2 种类型数据, 包括股票历史价格数据和行业关系数据。

1) 股票历史价格数据: 收集 758 支股票的历史价格数据, 对于突发事件造成的数据缺失使用前一日的的数据填充。股票历史价格数据涉及 6 个

基本价格特征 (开盘价、最高价、最低价、收盘价、成交量和交易额) 和 4 个技术指标特征 (5、10、20 和 30 日移动平均线)。最后这些数据经过标准化处理后得到 10 维特征作为模型的输入。

2) 行业关系数据: 按照申万行业划分标准对所有股票进行分组, 分为 104 个行业类别。

美股数据集 收集了 2013 年 1 月 2 日—2017 年 8 月 12 日期间 1 026 只纳斯达克交易所 (NASDAQ) 股票和 1 737 只纽约证券交易所 (NYSE) 股票^[12]。对于这些股票, 同样收集 2 种类型数据: 股票历史价格数据和行业关系数据。

3.2 训练设置

本文模型基于 PyTorch 框架实现, 并使用网格搜索来确定最佳超参数。为了保证实验结果的公平性, 首先将每个交易日的股票价格数据嵌入到 16 维空间中, 然后输入到不同的方法中。模型使用 Adam 优化器进行训练, 初始学习率设置为 0.003, dropout 设置为 0.4, 最大 epoch 设置为 600。对于改进的 Transformer 在集合 {16, 32, 64} 内搜索特征嵌入维度大小, 在区间 (2, 8) 内搜索一维卷积核大小, 在集合 {16, 32, 64} 内搜索超图神经网络特征嵌入维度大小。所有超参数使用验证集进行了优化, 在时间窗口大小为 25, 一维卷积核大小为 6, Transformer 网络和超图神经网络的特征嵌入维度大小为 32 时, 模型取得最佳效果。

实验环境为: Windows10 操作系统、AMD Ryzen 7 5800H with Radeon Graphics 3.20 GHz、Nvidia GeForce GTX 3060 GPU。

3.3 评价指标及基准模型

选取分类任务中广泛使用的 4 个评价指标: 准确率 (accuracy)、精确率 (precision)、召回率 (recall) 和调和平均值 (F1-score)。

$$I_{\text{Accuracy}} = \frac{N_{\text{TP}} + N_{\text{TN}}}{N_{\text{TP}} + N_{\text{TN}} + N_{\text{FP}} + N_{\text{FN}}}$$

$$I_{\text{Precision}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FP}}}$$

$$I_{\text{Recall}} = \frac{N_{\text{TP}}}{N_{\text{TP}} + N_{\text{FN}}}$$

$$F1\text{-score} = 2 \times \frac{I_{\text{Recall}} \times I_{\text{Precision}}}{I_{\text{Recall}} + I_{\text{Precision}}}$$

3.4 实验结果与分析

为了验证模型的性能, 本文将其与一些现有的模型进行了比较。基准模型包括: LSTM、DARNN^[24]、SFM^[5]、GCN^[11]、TGC^[12]、HATS^[25]、STHGCN^[18]、HGTAN^[19]。

DARNN^[24]: 该方法是 LSTM 的扩展, 采用双重注意力机制, 在每个时间步自适应地提取相关输入特征, 并在所有时间步内选择相关的编码器

隐藏状态。

SFM^[5]: 该方法是 LSTM 的扩展, 将 LSTM 存储单元的隐藏状态分解为多个频率分量, 以发现多频交易模式。

GCN^[11]: 该方法是基于图的方法, 使用 LSTM 对股票的历史价格数据进行编码, 然后将结果输入到 GCN 中, 以根据股票之间的关系进行学习。

TGC^[12]: 该方法是基于图的方法, 在 LSTM 基础上, 通过设计时间图卷积模块, 根据股票之间的关系进行学习并生成股票的关系嵌入。

HATS^[25]: 该方法是基于图的方法, 设计了一种分层图注意力网络, 该网络有选择地聚合不同类型的关系数据来学习股票表征。

STHGCN^[18]: 该方法是基于超图的方法, 通过股票的行业归属关系预定义静态超图, 并引入门控时间卷积来捕获股票价格特征中的时间依赖关系。

HGTAN^[19]: 该方法是基于超图的方法, 通过股票的行业归属关系和主动型基金持股关系预定义静态超图, 并引入三重注意力机制来引导信息流在超图上传播。

实验结果如表 1 所示。

表 1 在 A 股数据集上 DHGCN 与现有股票预测方法性能比较

Table 1 Performance comparison between DHGCN and existing stock prediction methods on the A-share datasets %

模型	Accuracy	Precision	Recall	F1-score
LSTM	35.09	38.09	34.37	35.90
DARNN	37.68	37.81	35.17	36.43
SFM	34.95	24.82	33.34	28.22
GCN	37.44	39.07	34.49	36.62
TGC	38.42	39.35	35.72	37.44
HATS	38.85	38.70	35.06	36.78
STHGCN	38.81	36.57	35.11	35.75
HGTAN	40.02	41.77	39.03	40.32
DHGCN	40.03	40.78	40.53	40.59

由对比实验结果可以得出以下结论:

1) 考虑股票间关系的方法如 GCN、TGC、HATS、STHGCN、HGTAN 和 DHGCN 在多个指标上优于 LSTM、DARNN 和 SFM, 说明融合关系信息可以提高模型预测精度。

2) HGTAN 和 DHGCN 都是基于超图的模型, 二者预测结果都优于基于图的模型, 这也说明超图具有捕获股票间高阶协同关系的能力。

3) DHGCN 的性能超越了其他基准模型, 它在 Recall 和 F1-score 指标上分别超过次优模型 3.8% 和 0.6%。此外, DHGCN 模型的实验结果优

于同是基于超图的模型 (STHGCN 和 HGTAN), 这可能归因于 DHGCN 采用的是数据驱动的方式自动构建动态超图, 而后者则需要先验信息进行预先定义。

表 2 给出了不同方法在美股数据集上的预测性能。可以看到, DHGCN 在多个评估指标中取

得最好或持平结果, 尤其是在 NYSE 数据集上, 4 项指标分别优于次好模型 (HGTAN) 1.26%、5.70%、1.53% 和 13.63%。此外本文还注意到通过引入股票间的关系信息, 基于图或超图的模型性能都得到了显著提高, 这一结果与 A 股数据集上的结果一致。

表 2 在美股数据集上 DHGCN 与现有股票预测方法性能比较

Table 2 Performance comparison between DHGCN and existing stock prediction methods on the US stock datasets %

模型	NASDAQ				NYSE			
	Accuracy	Precision	Recall	F1-score	Accuracy	Precision	Recall	F1-score
LSTM	37.22	34.64	36.56	35.52	45.73	36.22	38.04	37.08
DARNN	40.46	37.05	37.76	37.40	47.98	41.41	39.53	40.44
SFM	33.41	11.13	33.33	16.68	45.73	15.24	34.48	21.13
GCN	39.75	40.82	38.78	39.76	45.99	35.89	37.38	36.31
TGC	39.98	38.24	38.08	38.16	47.95	41.94	38.54	40.15
STHGCN	40.11	39.84	39.09	39.46	47.08	39.48	37.57	38.47
HGTAN	40.67	38.11	38.86	38.48	48.25	41.02	39.84	40.42
DHGCN	40.67	42.69	41.58	41.05	48.86	43.36	40.45	45.93

3.5 模型消融实验

1) Transformer: 基于标准的多头注意力机制编码时序信息用于趋势预测。

2) I-Transformer (improved-Transformer): 在 1) 基础上改进多头注意力机制和增添时序注意力聚合模块用于趋势预测。

3) I-Transformer+static: 在 2) 基础上融合静态关系信息用于趋势预测。

4) I-Transformer+dynamic: 在 2) 基础上融合动态关系信息用于趋势预测。

5) I-Transformer+static+dynamic(DHGCN): 在 2) 基础上融合静态和动态关系信息用于趋势预测。

表 3 给出了不同 DHGCN 变体的性能。可以看到, 改进的 Transformer 模型 (I-Transformer) 在 4 项指标上均高于 Transformer, 这表明感知股票历史时段相类似的价格波动规律对于股票趋势预测是有效的。未考虑实体间关系信息的模型 (Transformer 和 I-Transformer) 预测性能低于其他考虑了股票间实体关系信息的模型, 说明捕捉股票之间的关系信息是有价值的。此外, 融合静态和动态关系信息的模型 (DHGCN) 比仅考虑静态关系信息的模型 (Transformer+static) 或仅考虑动态关系信息的模型 (Transformer+dynamic) 表现更好。这也证明了静态关系和动态关系都是重要的, 二者缺一不可。DHGCN 取得了最好的预测效果, 证实了融合不同模块信息能够增强特征嵌入表示趋势的能力。

表 3 DHGCN 模型在 A 股数据集上消融研究结果

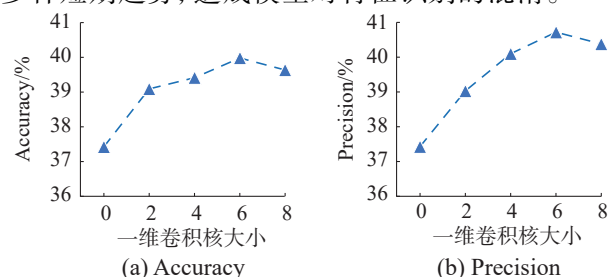
Table 3 DHGCN model ablation research results on the A-share datasets %

模型	Accuracy	Precision	Recall	F1-score
Transformer	35.91	35.45	34.79	34.45
I-Transformer	37.16	36.43	35.86	36.92
I-Transformer+static	39.02	38.19	38.75	37.93
I-Transformer+dynamic	39.28	38.93	38.69	39.04
DHGCN	40.03	40.78	40.53	40.59

3.6 模型超参数分析

本文通过调整一维卷积核大小 c 来了解 DHGCN 在 4 项指标上的表现, 使用 3.2 节中提到的默认设置来固定其他超参数。

图 6(a)~(d) 给出了改变一维卷积核大小 c , 结果呈现出先升后降的趋势。在 $c=6$ 时获得了最好的结果。当 c 过小时, 模型无法学习当前和历史时段是否存在类似波动规律, 尤其是当 $c=0$ 时, 改进的 Transformer 网络退化为标准的 Transformer 网络; 当 c 过大时, 模型提取到的趋势特征包含了多种短期趋势, 造成模型对特征识别的混淆。



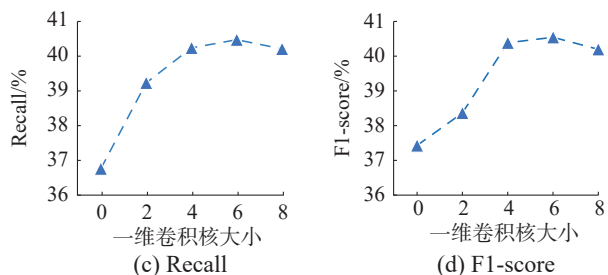


图 6 改进的 Transformer 网络中一维卷积核大小对模型整体性能的影响

Fig. 6 Impact of one-dimensional convolutional kernel size of the improved Transformer network on the overall performance

4 结束语

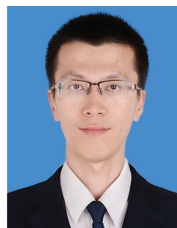
本文提出了一种基于改进的 Transformer 和动态超图卷积神经网络的股票趋势预测框架 DHGCN。首先使用改进后 Transformer 编码序列的时间维度信息,然后通过静-动态超图和超图卷积建模序列的空间维度信息,将不同序列关系信息引入到时序建模中,最后趋势预测层融合静-动态信息用于股票趋势预测。对比实验结果表明,在 A 股数据集和美股数据集上,本文模型的预测性能都有明显的优势。此外本文还对模型进行了消融实验,验证了所提模型各个组件的有效性。在未来工作中,将探索融合多源信息,如在线金融新闻和社交媒体信息等用于股票综合预测。

参考文献:

- [1] 次必聪, 张品一. 基于 ARIMA-LSTM 模型的金融时间序列预测 [J]. 统计与决策, 2022, 38(11): 145–149.
- [2] 徐浩然, 许波, 徐可文. 机器学习在股票预测中的应用综述 [J]. 计算机工程与应用, 2020, 56(12): 19–24.
- [3] 李斌, 龙真. 中国股票市场可预测性研究: 基于机器学习的视角 [J]. 管理科学学报, 2023, 26(10): 138–158.
- [4] HOSEINZADE E, HARATIZADEH S. CNNpred: CNN-based stock market prediction using a diverse set of variables[J]. *Expert systems with applications*, 2019, 129: 273–285.
- [5] ZHANG Liheng, AGGARWAL C, QI Guojun. Stock price prediction via discovering multi-frequency trading patterns[C]//Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. Halifax: ACM, 2017: 2141–2149.
- [6] DING Qianggang, WU Sifan, SUN Hao, et al. Hierarchical multi-scale Gaussian transformer for stock movement prediction[C]//Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence. Yokohama: IJCAI, 2020: 4640–4646.
- [7] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. *Advances in neural information processing systems*, 2017, 30: 5998–6008.
- [8] FENG Shibo, XU Chen, ZUO Yu, et al. Relation-aware dynamic attributed graph attention network for stocks recommendation[J]. *Pattern recognition*, 2022, 121: 108119.
- [9] 吴博, 梁循, 张树森, 等. 图神经网络前沿进展与应用 [J]. 计算机学报, 2022, 45(1): 35–68.
- [10] 吴国栋, 查志康, 涂立静, 等. 图神经网络推荐研究进展 [J]. 智能系统学报, 2020, 15(1): 14–24.
- [11] CHEN Yingmei, WEI Zhongyu, HUANG Xuanjing. Incorporating corporation relationship via graph convolutional neural networks for stock price prediction[C]//Proceedings of the 27th ACM International Conference on Information and Knowledge Management. Torino: ACM, 2018: 1655–1658.
- [12] FENG Fuli, HE Xiangnan, WANG Xiang, et al. Temporal relational ranking for stock prediction[J]. *ACM transactions on information systems*, 2019, 37(2): 1–30.
- [13] HE Pinglin, SUN Yulong, ZHANG Ying, et al. COVID-19's impact on stock prices across different sectors—an event study based on the Chinese stock market[J]. *Emerging markets finance and trade*, 2020, 56(10): 2198–2212.
- [14] FENG Yifan, YOU Haoxuan, ZHANG Zizhao, et al. Hypergraph neural networks[C]//Proceedings of the AAAI conference on artificial intelligence. Honolulu: AAAI, 2019: 3558–3565.
- [15] WEI Jinfeng, WANG Yunxin, GUO Mengli, et al. Dynamic hypergraph convolutional networks for skeleton-based action recognition[EB/OL]. (2021–12–20)[2023–08–16]. <https://doi.org/10.48550/arXiv.2112.10570>.
- [16] WANG Jingcheng, ZHANG Yong, WEI Yun, et al.

- Metro passenger flow prediction *via* dynamic hypergraph convolution networks[J]. *IEEE transactions on intelligent transportation systems*, 2021, 22(12): 7891–7903.
- [17] XIA Lianghao, HUANG Chao, XU Yong, et al. Hypergraph contrastive collaborative filtering[C]//Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. Madrid: SIGIR, 2022: 70–79.
- [18] SAWHNEY R, AGARWAL S, WADHWA A, et al. Spatiotemporal hypergraph convolution network for stock movement forecasting[C]//2020 IEEE International Conference on Data Mining. Sorrento: IEEE, 2020: 482–491.
- [19] CUI Chaoran, LI Xiaojie, ZHANG Chunyun, et al. Temporal-relational hypergraph tri-attention networks for stock trend prediction[J]. *Pattern recognition*, 2023, 143: 109759.
- [20] LI Xiaojie, CUI Chaoran, CAO Donglin, et al. Hypergraph-based reinforcement learning for stock portfolio selection[C]//2022 IEEE International Conference on Acoustics, Speech and Signal Processing. Singapore: IEEE, 2022: 4028–4032.
- [21] GAO Yue, FENG Yifan, JI Shuyi, et al. HGNN⁺: general hypergraph neural networks[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2023, 45(3): 3181–3199.
- [22] JEANBLANC M, YOR M, CHESNEY M. Mathematical methods for financial markets[M]. London: Springer Science & Business Media, 2009.
- [23] BAI Song, ZHANG Feihu, TORR P H S. Hypergraph convolution and hypergraph attention[J]. *Pattern recognition*, 2021, 110: 107637.
- [24] QIN Yao, SONG Dongjin, CHEN Haifeng, et al. A dual-stage attention-based recurrent neural network for time series prediction[C]//Proceedings of the 26th International Joint Conference on Artificial Intelligence. Melbourne: IJCAI, 2017: 2627–2633.
- [25] KIM R, SO C H, JEONG M, et al. Hats: a hierarchical graph attention network for stock movement prediction [EB/OL]. (2019–08–07)[2023–08–16]. <https://doi.org/10.48550/arXiv.1908.07999>.

作者简介:



郝剑龙, 讲师, 博士, 主要研究方向为深度学习、金融时序分析。主持国家自然科学基金青年项目 1 项, 发表学术论文 10 余篇。E-mail: haojianlong2012@sxufe.edu.cn。



刘志斌, 硕士研究生, 主要研究方向为金融时序分析、数据挖掘。E-mail: soberlau34@163.com。



常新功, 教授, 博士, 中国计算机学会高级会员, 主要研究方向为图神经网络、数据挖掘、进化算法。发表学术论文 30 余篇。E-mail: c_x_g@126.com。