# 基于注意力机制的Dilated CNN-BiLSTM的股票走势预测

摘要——基于价量数据预测未来股价走势是一种重要的机器学习量化选股策略，长期以来在研究和实践中得到广泛应用。由于价量数据伴随交易活动产生，本质上是一类时间序列数据。为了建模价量数据与未来股价走势之间的关系，大多数现有研究采用了循环神经网络等时序模型。然而，时序模型在捕捉价格与交易量走势形态方面存在局限性，影响了其性能表现。为解决这一问题，我们提出利用卷积神经网络对图表化的价量数据与未来股价进行建模，从而更有效地预测未来股价。包括使用扩张卷积来捕获图表中的时序信息，并通过双向长短期记忆网络（BiLSTM）对捕获到的特征进行处理。相较于单纯的卷积神经网络（CNN）和BiLSTM，我们的模型将时序信息的捕获和处理分开，可以更好的利用图表中的信息。在价量图表的构建上，我们使用了多种方法，包括matpoltlib手工绘制和mplfinance绘制。并尝试以5日和20日窗口期的K线图作为模型输入来对股价走势进行预测，相较于baseline中简单的K线图，我们的方法在数据集上的相关系数获得了显著的提升。

## 研究背景

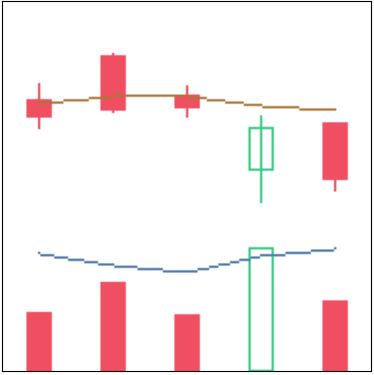
### **1.1 基于价量数据的机器学习量化选股策略**

基于价量数据预测未来股价走势是一种重要的机器学习量化选股策略，长期以来受到广泛的研究与应用。由于个股的价量数据随交易活动动态生成，其本质上是时间序列数据。因此，大多数研究自然选择采用循环神经网络（RNN）或Transformer等时序模型来建模价量数据与未来股价走势之间的关系。

在这些方法中，模型通常以价量数据的一维或多维数组作为输入，输出则是股价的未来走势。尽管时序模型能够在一定程度上捕捉价量序列中的价格、交易量波动及其复杂的交互关系，但它们在识别价格与交易量走势的形态及变化特征方面仍存在局限性。

以人类视角来看，在预测股价未来走势时，通常不会仅依赖价量序列中的数字变化，因为仅从数值波动中难以全面理解其内在规律。因此，为了更好地捕捉价格和交易量的形态及趋势，人们倾向于参考如K线图、移动平均线、成交量柱状图和MACD指标等图表（如图1），而非单纯的一组数字序列。

因此，我们舍弃了使用时序模型对序列数据进行建模的传统方法。转而采用扩张卷积和双向长短期记忆网络对标准化的K线图和未来股价走势进行建模，以实现对未来股价走势的预测。对此，本方法首先构建了包含蜡烛图、移动平均价、交易量的标准化图表（如图2）。然后，设计了能捕捉图表中价量数据走势形态的卷积神经网络来对当前股票市场及其未来股价走势进行建模。

**图1 同花顺上某股的月行情数据 图2 标准化K线图**

## **相关工作**

### **2.1 机器学习方法**

2001年，Gençay等[1]研究学者系统性的介绍了小波变换和其他滤波方法在金融和经济学领域的应用，并进一步探讨了如傅里叶分析、卡尔曼滤波在内的各种滤波技术，与它们在金融时间序列分析中的作用。但应用成本较高，复杂的滤波方法也易出现过拟合风险。2014年，Marković等[2]基于支持向量机（SVM），通过技术指标和宏观经济指标的分析方法进行特征选择，从而建立预测贝尔格莱德证券交易所（BELEX15）股票市场指数的趋势预测模型。但预测模型单一，仅适合短期股票预测。Zi R等[3]提出一种基于甲醛随机森林蚁群算法的预测模型，从而缓解基准模型在金融分析应用中精度低、误差大的问题，但整体模型效果仍不尽如人意。

综上所述，尽管通过机器学习的方法预测股票模型不断发展和完善，但在准确性、适用性和复杂性方面仍存在挑战。伴随着深度学习的发展逐渐而被替代。

### **2.2深度学习方法**

20世纪80年代，深度学习算法走入人们视野。研究学者不断发展相关算法并应用于股票预测领域。

**卷积神经网络（Convolutional Nerural Networks，CNN）**。Mehtab等[4]通过使用具有前行验证的卷积神经网络构建基于深度学习的回归模型来增强模型的预测能力。针对其参数进行了微调，以便验证损失随着迭代次数的增加而稳定下来，并收敛训练和验证精度。但可解释性差。Ma D等[5]提出了一种带有图卷积网络（Graph Convolutional Networks，GCN）的正则化集成框架，用于股票价格预测。该框架采用多个多图卷积网络构建集成图卷积网络，阐明不同时间尺度的交易数据中的时空关系。从交易数据的每个维度序列中提取特征，从而能够捕获不同时间尺度的特征并减轻数据中的非平稳性和噪声。但仍然有一定的过拟合风险。

**循环神经网络（Recurrent Neural Network，RNN）。**2018年，Jahan等[6]应用基本的RNN算法，处理股票时间序列数据，并结合有监督机器学习算法来预测单只股票。Wang等人[7]提出了一种结合门控循环单元（Gated Recurrent Unit，GRU）和隔离森林（IForest）的混合模型用于股票价格预测。利用GRU来有效地解决标准RNN在处理长序列数据时遇到的梯度消失问题，并结合了隔离森林算法来检测异常值，以提高预测的鲁棒性。但GRU模型对超参数的选择较为敏感。此外，模型的复杂性可能导致计算成本较高，限制系统应用。Teng等[8]提出了一个多尺度局部线索和基于层次注意力的LSTM模型（MLCA-LSTM）来捕捉潜在的价格趋势模式。 通过分层注意力机制和多分支 LSTM 结构，时间依赖和多尺度交互共同参与和聚合。但模型需要大量数据来避免过拟合，周期长，对参数选择具有较大的依赖性，这也使得LSTM在股票预测领域的应用上逐步被BiLSTM所代替。

**transformer模型。**Li等[9]基于transformer提出了一种名为Spacetimeformer的新模型，用于股票市场预测。该模型不仅考虑了时间序列数据，还纳入了空间维度，即不同股票之间的相互关系，有助于更全面地理解市场动态。但需要需要大量高质量的时空数据和计算资源才能发挥最佳效果。Gou等[10]提出了一种基于非平稳Transformer模型的股票价格预测方法。利用Transformer的注意力机制，模型能够捕捉到数据中的长期依赖关系和复杂的交互模式。通过引入系列平稳化模块和非平稳注意力模块来处理股票价格时间序列数据中的非平稳性，提高预测的准确性，更好地适应市场变化。但推理成本高，模型可解释性不足。

## **研究方法**

### **3.1 数据预处理**

为了生成更加直观和信息丰富的图表，我们需要对原始数据进行预处理，计算出一些常见的技术指标。这些指标能够帮助更好地揭示股票价格和成交量的趋势特征，为后续建模提供更有意义的输入。以下是主要计算的指标和方法：

#### 3.1.1 移动平均线（MA）

移动平均线通过平滑股票价格或成交量的短期波动，能够有效捕捉趋势信息：

· **MA5**：计算过去5个交易日的简单移动平均线，用于短期趋势判断。

· **MA10**：计算过去10个交易日的简单移动平均线，用于中期趋势分析。

#### 3.1.2 移动平均收敛发散指标（MACD）

MACD 是一种基于指数平滑移动平均（EMA）的指标，用于揭示价格的短期和长期趋势：

* EMA12：过去12天的指数平滑移动平均线（短期）。

EMA26：过去26天的指数平滑移动平均线（长期）。

DIF：即短期EMA与长期EMA之差：

DIF = EMA12 − EMA26

* DEA：DIF的9日移动平均值。
* MACD：DIF与DEA之间的差值：

MACD = 2 × (DIF - DEA)

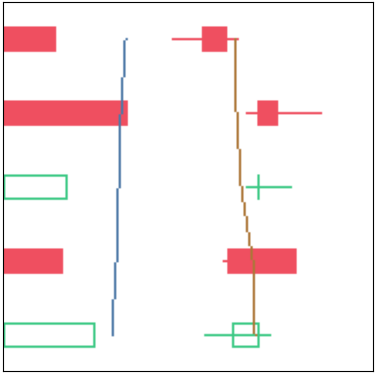
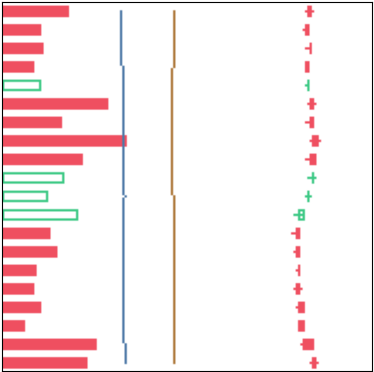
#### 3.1.3 成交量指标

为了捕捉成交量的变化趋势，我们计算成交量的移动平均值（如 MA5 和 MA10），用于衡量市场的活跃程度。

### **3.2 K线图构建方法**

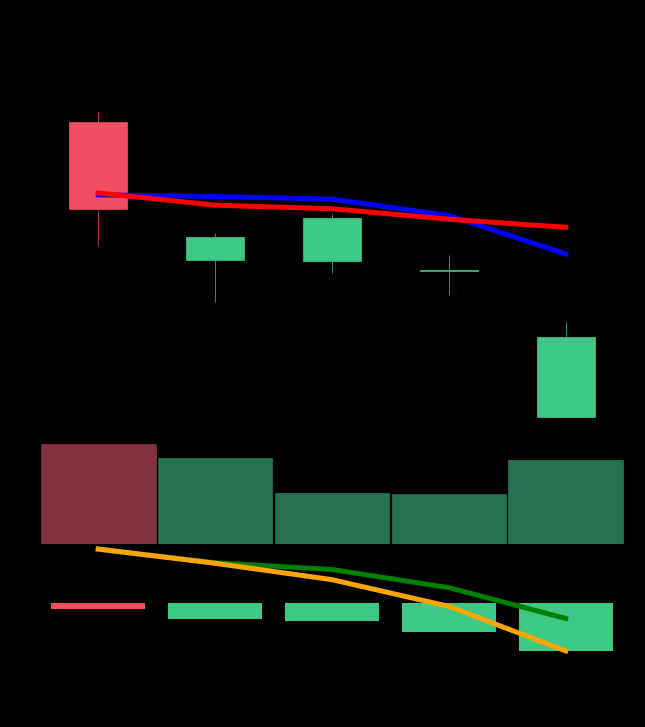
作为模型输入的K线图包含全部的股票信息，因此K线图的绘制至关重要。我们分别采用matpoltlib手工绘制和mplfinance绘制进行尝试。其中matpoltlib手工绘制较为简单，mplfinance绘制较为复杂。因为服务器和模型大小的原因，我们选择了matpoltlib手工绘制的图表作为我们模型的输入。对于模型更为复杂的模型可以使用mplfinance绘制的图表。除此之外，我们还分别尝试了以5日和20日窗口期的股票数据进行绘制。

#### 3.2.1 matpoltlib手工绘制

** **

**图3 plt绘制5日K线图 图4 plt绘制20日K线图**

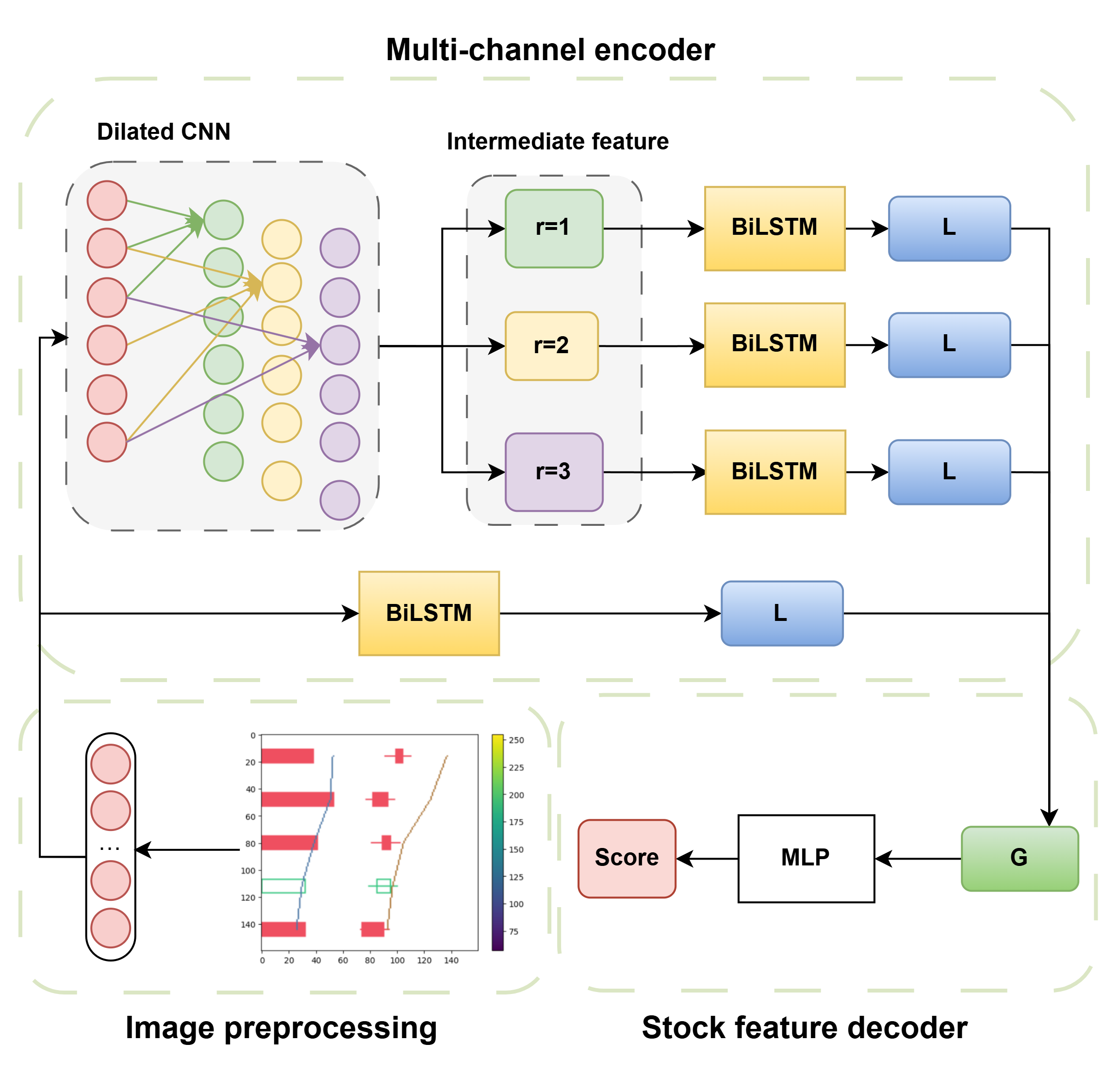
#### 3.2.1 mplfinance绘制

** **

**图5 mpf绘制5日K线图 图6 mpf绘制20日K线图**

### **3.3 基于注意力机制的Dilated CNN-BiLSTM**

如图7所示，本文提出的多通道Dilated CNN-BiLSTM模型主要由3部分组成:图片预处理、多通道编码和股票特征解码。



#### **图 7 基于注意力机制的Dilated CNN-BiLSTM。其中L为局部注意力机制，G为全局注意力机制。**

#### 3.3.1 图片预处理

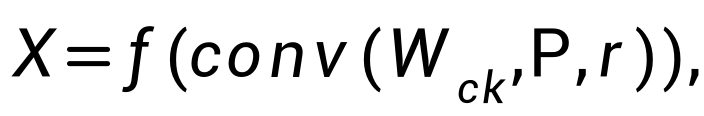
经过前述操作，我们将形状为 (160, 160, 3) 的K线图作为输入。首先，对图片数据进行归一化处理，以保证特征值的分布更加均匀。随后，通过嵌入层提取图片中的关键特征，将复杂的视觉信息转化为有意义的特征表达。经过处理后，图片被映射为股票趋势的特征向量P，从而为后续的分析和预测奠定了基础。

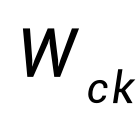
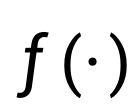
#### 3.3.2 多通道编码

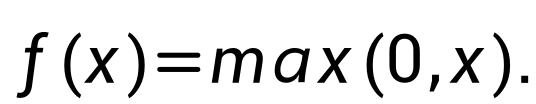
股票的价格波动复杂多变，要细致分析其变化趋势极具挑战性。为此，多通道编码器通过结合适当的 CNN 扩张率，捕捉高层次的上下文信息。利用多通道结构，模型能够提取不同尺度的高层特征以及原始上下文信息。每个通道的核心结构均采用 dilated CNN-BiLSTM。

首先，为了捕捉局部特征并通过这些局部特征推断整体趋势，我们引入CNN。然而，传统 CNN[11]存在无法同时提取相同卷积核下不同尺度特征的局限性。此外，股票涨跌的趋势不仅仅由局部特征决定。因此，我们采用dilated CNN[12]，通过扩张卷积获取不同尺度的特征信息。然而，dilated CNN的网格效应可能导致信息遗漏。针对这一问题，我们设计了多通道框架，通过多通道提取不同尺度特征，并融合这些特征以增强模型的表现。

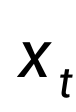
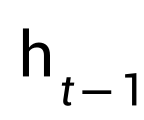
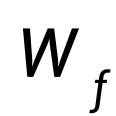
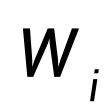
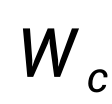
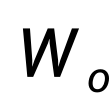
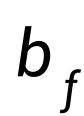
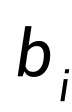
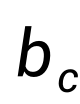
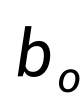
对每个序列向量P，使用一个扩张率为r的卷积来提取高层特征，并使用一个双曲型单元作为激活函数。最后dilated CNN的输出是一个新的向量X可被描述为：

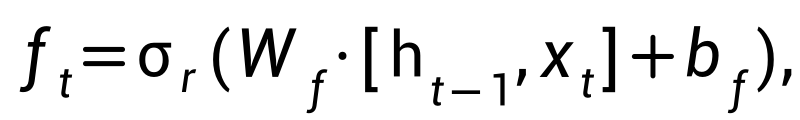


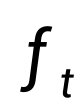
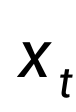
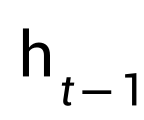
其中是扩张卷积中核的权重，是线性整流函数[13]：

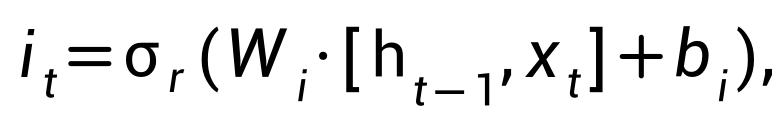


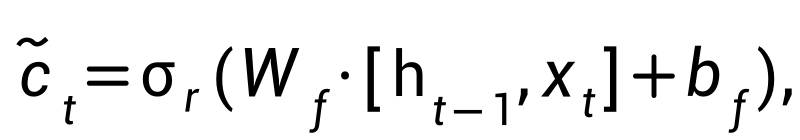
为了捕捉股票趋势中的长期依赖关系，我们在dilated CNN操作之后引入了BiLSTM[14]，用于获取高层次的上下文特征。BiLSTM可以全面分析时间序列中双向的依赖信息，从而更准确地理解长期趋势。

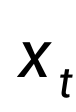
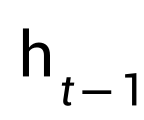
双向长短时记忆网络（BiLSTM）是由两个LSTM层组合而成，这两个层分别是前向LSTM和后向LSTM。它们共同作用，使得网络能够同时捕获序列数据的前向和后向依赖关系。BiLSTM的输出是一个维度为k×2l的向量，其详细表示可以通过下述公式进行阐述，其中是第t个LSTM单元的输入，是前一个LSTM单元的输出；Σ表示逻辑回归函数Sigmoid；\*是元素级乘法；，，，和，，，是第t个LSTM单元的权重和偏置。

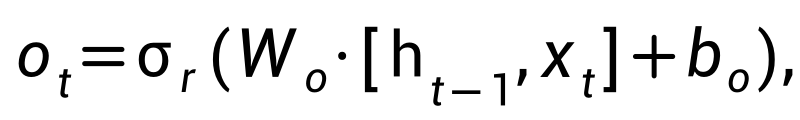


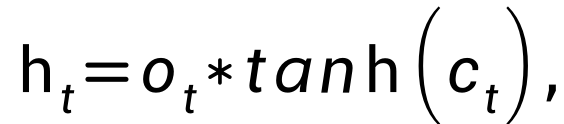
其中是LSTM的遗忘门，它可以通过和确定那些信息需要丢弃。

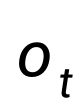
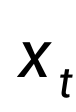
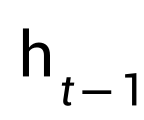




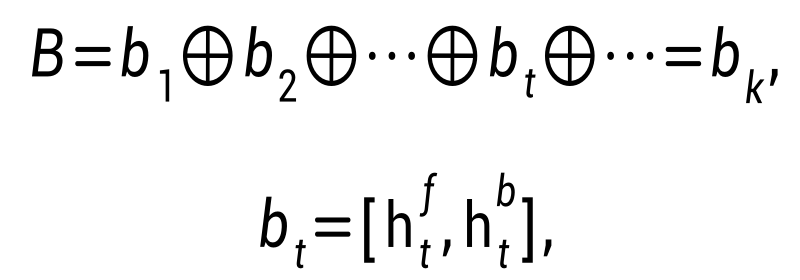
其中latexmath是LSTM的输入门，可以决定通过和更新那些信息。

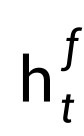
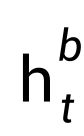




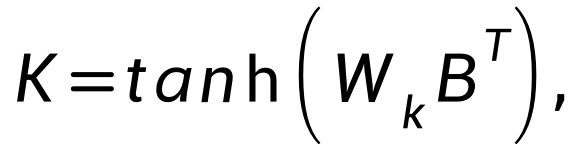
其中是LSTM的输出门，通过和可以得到输出的判别条件，通过输出门和信息单元可以得到LSTM的输出。

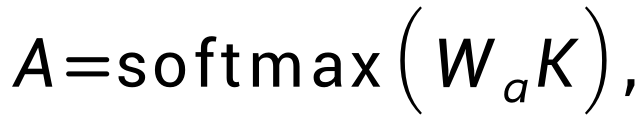
最后，BiLSTM的输出可以表示为：

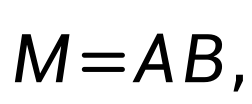


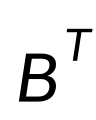
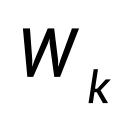
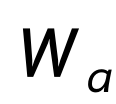
其中为前向LSTM的第t个单元的输出，而为后向LSTM的第t个单元的输出。

此外，我们在 BiLSTM 的输出阶段引入了注意力机制[15]，为特征分配不同权重，使模型更加关注关键特征。每个通道的输出特征均经过注意力机制加权处理，确保信息的有效表达。 注意力机制对Bi LSTM的输出特征赋予不同的权重，并将输出作为单通道编码特征。具体如下：

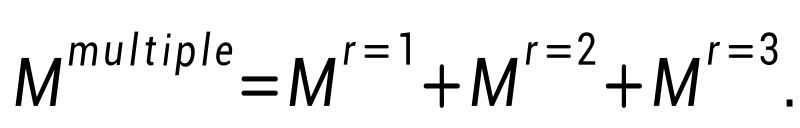


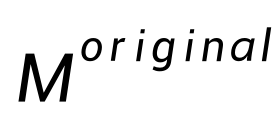


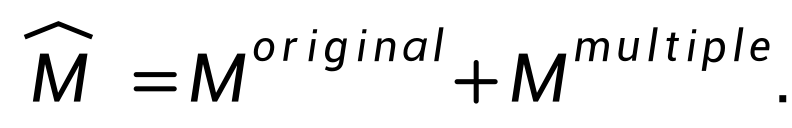


其中是B的转置矩阵，M为注意力层的输出，,为可训练的参数。

以上描述详细介绍了单通道的标准结构。然后采用3种不同的扩张率构建多个通道用于提取不同尺度的情感特征，3个通道的输出可以融合为:



为了进一步降低特征遗漏和偏差的风险，我们特意保留了一条通道用于处理原始图片特征，确保原始信息的完整性。这种多通道架构在结合上下文信息、捕捉多尺度特征方面展现出了卓越的性能。其中没有使用扩张卷积，但其余与标准结构相同。设为原点，表示该信道的输出。最后，可伸缩多通道编码器的输出可以描述为：

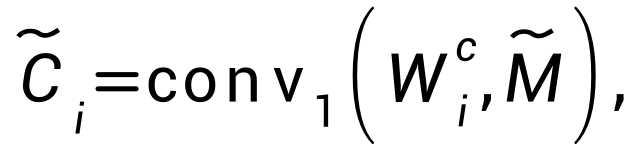


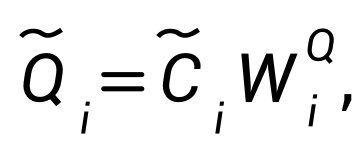
#### 3.3.3 股票特征解码

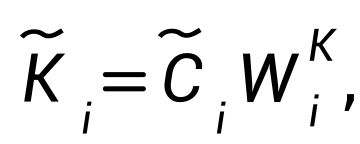
特征解码模块的主要功能是执行全局权重分配，并对多通道编码器生成的股票特征进行解码。该模块由两部分组成：全局注意力模块和MLP模块。

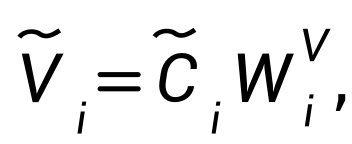
全局注意力模块用于对特征进行加权处理。通过全局注意力机制，模型能够突破局部上下文的限制，为每个特征位置分配不同的权重，从而捕捉特征间的全局依赖性。这种方法尤其适合分析股票细微而复杂的变化趋势，能够精准识别其变化规律。具体而言，全局注意力机制通过筛选关键特征并动态分配权重，使模型具备更强的全局感知能力。

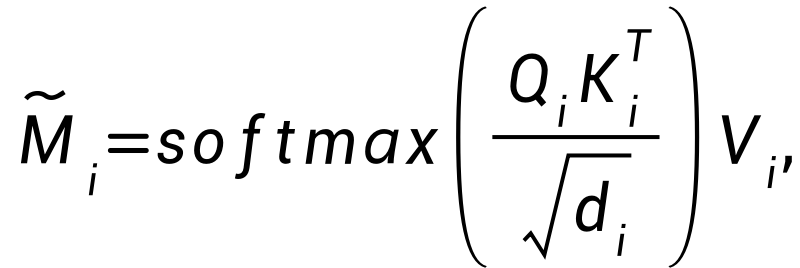
使用全局注意力模块对特征进行不同的关注。具体地，采用核大小为1×1的卷积将特征映射到不同的空间，并利用注意力机制进行特征筛选和权重分配。第i个全局注意力空间的细节可以表示为：

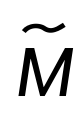
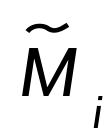
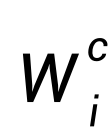
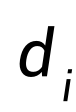
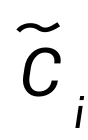




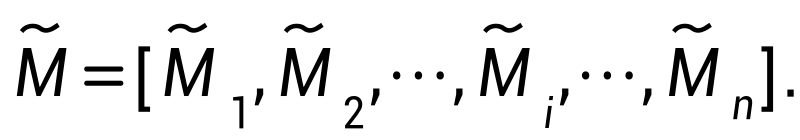




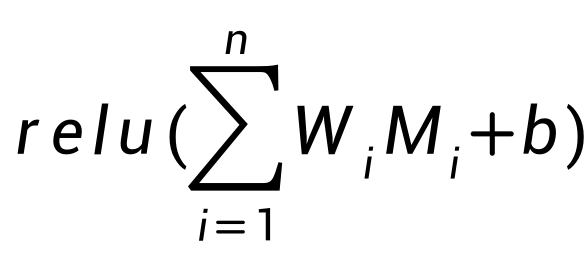


其中为股票特征解码器的输入，表示第i个全局注意力空间的输出，表示卷积的参数，为的维度，其余为注意力机制的参数。

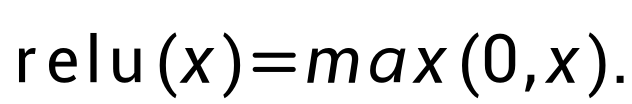
在股票特征解码器中，构建n个全局注意力空间，并使用一个级联层融合所有全局注意力空间的输出。具体公式可表示为：



MLP模块[16]则负责特征的进一步整合。它对编码器输出的特征进行深度处理，加强特征间的相关性与互补性，并将多通道的高维特征压缩为更有利于趋势判断的低维表示。这一过程不仅提升了模型对复杂趋势的表达能力，同时也显著提高了预测的准确性。细节如下：



其中relu的等式如下：



通过全局注意力和 MLP 模块的协同作用，模型能够更全面地捕捉股票趋势的特征，从而在复杂任务中展现出卓越的性能。

## 四、实验

### **4.1 评价指标**

评价模型性能的标准采用排名信息系数（Rank Information Coefficient，Rank IC），该指标用于衡量模型预测分值的排序与未来五天实际收益率排序之间的相关性。

在此评价体系中，我们模型的预测值代表未来一周的平均预期，因此，我们需要生成五份预测值来模拟接下来一周内每天的预测情况。这些预测值的排序将与相应日期的实际收益率排序进行比较，以计算Rank IC，从而评估模型的预测准确性。

### **4.2 数据来源**

为了评估提出方法效能，我们在2024年CSIG-飞桨、工银瑞信金融科技挑战赛提供的数据集上进行了详细的实验。该数据集的训练部分包含了匿名股票的交易特征，具体包括以下字段：股票代码（匿名股票代码）、日期编号（匿名日期代码）、开盘价（开盘价格）、最高价（最高交易价格）、最低价（最低交易价格）、收盘价（收盘价格）、成交量（交易量）以及成交额（交易金额）。我们的预测目标为：在给定日期t，预测未来五天的股票收益率排名得分。

具体而言，t日的未来五天收益率Ret\_t,5的计算公式为：（收盘价\_t+6-收盘价\_t+1）/收盘价\_t+1。这样，我们就可以通过比较预测得分与实际收益率来评估模型的预测能力。

测试集与训练集具有相同的特征结构，确保了实验的一致性和公正性。通过这种方式，我们能够全面检验所提方法在实际金融数据上的表现和可靠性。

### **4.3 训练参数**

算法采用AdamW[17]对网络进行训练，学习率设置为0.00005，按照衰减因子0.01进行更新，训练echo为30。由于dropout[18]可以防止神经网络的共适应，因此将dropout率0.5。为了防止梯度爆炸，采用梯度裁剪[19]，clip\_norm设置为0.5。在本文提出的方法中，CNN的一个目标是降低网络参数和模型复杂度，而Bi LSTM的合并方式是串联，因此CNN滤波器的个数为128，LSTM中隐藏单元的个数为256。此外，训练的主要损失函数是定义的MyMSE。此外采用自定义的均方误差(MSE)损失函数，针对特定任务优化模型的性能。对于每个目标序列，提取其最后一个非NaN的有效值作为目标值，确保计算的有效性和鲁棒性。

### **4.4 实验结果**

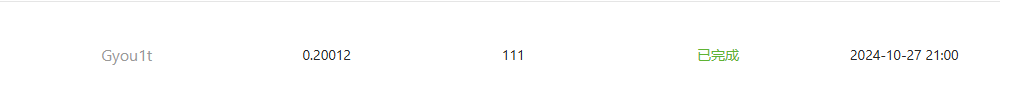
在实验中，我们的模型与XGBoost[20]、MLP[16]、CNN[11]、Simpletransfromer[21]、基于自注意力机制的TCN结合BiLSTM方法[22]在测试集上进行来比较。

**表1 结果比较**

|  |  |
| --- | --- |
| **方法** | **相关系数** |
| XGBoost | 0.0021 |
| CNN | 1. 0892 |
| MLP | 0.1042 |
| Simpletransfromer | 0.1213 |
| TCN+BiLSTM+自注意力机制 | 0.1421 |
| **OURS** | **0.1606** |

从表1中相关系数的实验结果可以看出，我们的方法相较于其他方法表现出显著优势。XGBoost的相关系数仅为0.0021，表明其在捕捉股票趋势特征方面效果极为有限。这主要是因为XGBoost更适合处理结构化数据，对时序特征和多尺度信息的提取能力不足。如果不是基于K线图提取信息，XGBoost的表现或许会有所提升。CNN在提取局部特征方面表现较好，但难以捕捉股票数据中的长程依赖性。我们的模型通过引入Dilated CNN，不仅保留了局部特征提取能力，还通过多尺度扩张卷积克服了传统CNN在不同时间粒度上的局限性，显著提升了对复杂特征的建模能力。相比之下，MLP的相关系数较低，仅为 0.1042，这是因为 MLP 无法有效建模时序数据的时间依赖性。而我们的模型通过BiLSTM和注意力机制，充分利用时序信息，将建模能力提升到了更高的水平。 SimpleTransformer的相关系数为0.1213，略高于MLP，但由于其全局特性捕捉方式可能导致过度平滑或特征稀疏化，限制了对复杂股票趋势的建模能力。相比之下，我们的方法通过局部特征提取（Dilated CNN）与全局依赖建模（BiLSTM 和注意力机制）的有机结合，在处理股票数据的复杂性时表现出更高的鲁棒性。 TCN+BiLSTM+自注意力机制方法的相关系数为 0.1421，与我们的方案更为接近。然而，TCN固定的感受野限制了其特征提取的灵活性，而我们的方法通过多通道的 Dilated CNN 灵活捕捉不同尺度的特征，同时结合注意力机制对关键特征分配更合理的权重，最终使相关系数提升至 0.16。综上所述，我们的方法通过整合 Dilated CNN 的多尺度特征提取能力、BiLSTM 的长程依赖捕捉能力以及注意力机制的关键特征增强能力，在相关性建模上展现了卓越的性能与明显的优势。

我们在初赛训练集上得到了0.20012的分数，如图8所示。



**图8 初赛结果**

### 4.5 模型性能影响因素分析

表1展示了我们模型的卓越性能。为了进一步探讨其性能表现的原因，我们对影响模型性能的关键因素进行了深入分析。具体而言，我们验证了K线图绘制策略的有效性以及不同训练规模对模型性能的影响。实验结果表明，所采用的K线图绘制策略能够显著提升特征提取的质量，同时我们的模型在小样本条件下依然表现出优异的性能，充分证明了其鲁棒性与适应性。

**k线图绘制策略：**相较于baseline中的黑白K线图，我们的K线图不但可以提供更多的信息，如MA5，MA10，MACD，而且还可以通过色彩的不同区分股票的涨、跌，使得模型可以更容易捕捉到关键的市场特征。这些优化使得我们的方法在捕捉市场动态和趋势方面更为精准。如表2所示，我们的方法在相关系数上的提升，证明了其在识别和预测股票市场趋势方面的优越性。我们的K线图绘制策略不仅提供了更丰富的视觉信息，还增强了模型对市场变化的敏感度，为投资者提供了更为可靠的决策支持。

**表2 k线图绘制测量对比**

|  |  |
| --- | --- |
| **k线图绘制策略** | **相关系数** |
| 灰色蜡烛图 | 0.1442 |
| **OURS** | 0.1606 |

**不同训练集规模：**小样本能力也是衡量模型性能的一个方面。我们的模型在数据量很少的情况下，依然能够超过MLP，CNN等传统模型。这使得我们的模型在面对真正的波动市场时，能够仅通过少量的短期数据，做到相对于其他模型更加精准的预测。如表3所示。

**表3 不同训练集规模对比**

|  |  |
| --- | --- |
| **训练集规模** | **相关系数** |
| 5万数据 | 0.1606 |
| 10万数据 | 0.1714 |
| 20万数据 | - |

## 五、总结

文本提出一种基于注意力机制的dilated CNN-BiLSTM多通道联合架构，用于进行股票走势预测。模型通过不同扩张率的dilated CNN提取多尺度的高层特征，利用BiLSTM捕捉特征间的上下文依赖性，并引入局部注意力机制对特征进行编码。为避免股票特征的遗漏，我们保留了一条通道以存储原始图像信息。在特征整合阶段，采用全局注意力机制和MLP进行融合，最终输出一个股票趋势分数。

实验结果表明，所提出模型在数据集上具有较强的预测能力。然而，模型仍存在一些挑战。例如，我们采用RGB格式的图片输入，当图片尺寸为（160，160，3）时，若直接将图片展平处理，可能导致冗余信息浪费颜色信息。而在使用三通道的Conv2D操作时，模型效果也难以达到最佳。如何在图像中平衡特征提取的效率，避免冗余信息，同时有效捕捉特征的时序关系，仍是未来需要进一步研究的方向。

**参考文献**

1. Gençay R, Selçuk F, Whitcher B J. An introduction to wavelets and other filtering methods in finance and economics[M]. Elsevier, 2001.
2. Marković I P, Stojanović M B, Stanković J Z, et al. Stock market trend prediction using support vector machines[J]. Facta Universitatis, Series: Automatic Control and Robotics, 2014, 13(3): 147-158.
3. Zi R, Jun Y, Yicheng Y, et al. Stock price prediction based on optimized random forest model[C]//2022 Asia Conference on Algorithms, Computing and Machine Learning (CACML). IEEE, 2022: 777-783.
4. Mehtab S, Sen J. Stock price prediction using convolutional neural networks on a multivariate timeseries[J]. arxiv preprint arxiv:2001.09769, 2020.
5. Ma D , Yuan D .Enhanced stock price forecasting through a regularized ensemble framework with graph convolutional networks[J].Expert Systems With Applications, 2024, 250.DOI:10.1016/j.eswa.2024.123948.
6. Jahan I, Sajal S. Stock price prediction using recurrent neural network (RNN) algorithm on time-series data[C]//2018 Midwest instruction and computing symposium. Duluth, Minnesota, USA: MSRP, 2018.
7. Wang, G., Wang, Y., & Zhang, Y. (2018). A novel hybrid model based on GRU and IForest for stock price prediction. In Proceedings of the 2018 International Conference on Industrial Economics and Industrial Security (IEIS) (pp. 1-6). IEEE.
8. Teng X , Zhang X , Luo Z .Multi-scale local cues and hierarchical attention-based LSTM for stock price trend prediction[J].Neurocomputing, 2022.DOI:10.1016/j.neucom.2022.07.016.
9. Li Y C, Huang H Y, Yang N P, et al. Stock Market Forecasting Based on Spatiotemporal Deep Learning[J]. Entropy, 2023, 25(9): 1326.
10. Gou C, Zhao R, Guo Y. Stock Price Prediction Based on Non-Stationary Transformers Model[C]//2023 9th International Conference on Computer and Communications (ICCC). IEEE, 2023: 2227-2232.
11. Chua L O. CNN: A vision of complexity[J]. International Journal of Bifurcation and Chaos, 1997, 7(10): 2219-2425.
12. Lei X, Pan H, Huang X. A dilated CNN model for image classification[J]. IEEE Access, 2019, 7: 124087-124095.
13. Daubechies I, DeVore R, Foucart S, et al. Nonlinear approximation and (deep) ReLU networks[J]. Constructive Approximation, 2022, 55(1): 127-172.
14. Mienye I D, Swart T G, Obaido G. Recurrent neural networks: A comprehensive review of architectures, variants, and applications[J]. Information, 2024, 15(9): 517.
15. Guo M H, Xu T X, Liu J J, et al. Attention mechanisms in computer vision: A survey[J]. Computational visual media, 2022, 8(3): 331-368.
16. Taud H, Mas J F. Multilayer perceptron (MLP)[J]. Geomatic approaches for modeling land change scenarios, 2018: 451-455.
17. Zhuang Z, Liu M, Cutkosky A, et al. Understanding adamw through proximal methods and scale-freeness[J]. Transactions on machine learning research, 2022.
18. Tinto V. Dropout from higher education: A theoretical synthesis of recent research[J]. Review of educational research, 1975, 45(1): 89-125.
19. Zhang J, He T, Sra S, et al. Why gradient clipping accelerates training: A theoretical justification for adaptivity[J]. arXiv preprint arXiv:1905.11881, 2019.
20. Chen T, Guestrin C. Xgboost: A scalable tree boosting system[C]//Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. 2016: 785-794.
21. Hou Q, Lu C Z, Cheng M M, et al. Conv2former: A simple transformer-style convnet for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024.
22. Fan J, Zhang K, Huang Y, et al. Parallel spatio-temporal attention-based TCN for multivariate time series prediction[J]. Neural Computing and Applications, 2023, 35(18): 13109-13118.