

将预训练模型提示纳入多模式 库存量变动预测

陈睿波*
ruibo chen@pku.edu.cn
北京大学,中国

张志远*
zzy1210@pku.edu.cn
北京大学,中国

刘毅
yliu.pku@outlook.com
北京大学,中国

鲍瑞涵
ruihan.bao@mizuho-sc.com
日本瑞穗证券有限公司

张本惠子
keiko.harimoto@mizuho-sc.com
日本瑞穗证券有限公司

孙旭
xusun@pku.edu.cn
北京大学,中国

抽象的

利用股票相关新闻进行多模式股票交易量变动预测是金融领域的基本问题之一。现有的从头开始训练模型的多模态工作在对财经新闻建模时面临着缺乏通用知识的问题。此外,由于数据集中的数据不足,模型的能力可能会因缺乏领域相关知识而受到限制。为了解决这个问题,我们提出了基于提示的多模态股票交易量预测模型 (ProMUSE)来处理文本和时间序列模态。我们使用预先训练的语言模型来更好地理解财经新闻,并采用即时学习方法来利用其普遍知识的能力来对文本信息进行建模。此外,简单地融合两种模式可能会对单峰表示造成损害。因此,我们提出了一种新颖的跨模态对比对齐,同时保留融合头旁边的单模态头来缓解这个问题。

大量实验表明,我们提出的 ProMUSE 优于现有基线。与潜在变体和学习机制相比,综合分析进一步验证了我们的架构的有效性。我们的代码将在 <https://github.com/RayRuiboChen/ProMUSE> 中提供。

CCS 概念

·计算方法→神经网络。

关键词

多模态学习、股票走势预测、即时学习

ACM 参考格式: Ruibo Chen*, Zhiyuan 张*, Yi Liu, Ruihan Bao, Keiko Harimoto 和 Xu Sun. 2023. 将预训练模型提示纳入多模式库存量变动预测。 KDD '23 金融机器学习研讨会, 2023 年 8 月, 研讨会。 ACM, 美国纽约州纽约市, 9 页。 <https://doi.org/xxxx/xxxx>

允许免费制作本作品全部或部分内容的数字或硬拷贝以供个人或课堂使用,前提是制作或分发副本不是为了盈利或商业利益,并且副本在页面上附有此通知和完整引用。必须尊重 ACM 以外的其他人拥有的本作品组件的版权。允许以信用方式提取。要以其他方式复制、重新发布、发布到服务器上或重新分发到列表,需要事先获得特定许可和/或付费。从 permissions@acm.org 请求权限。

KDD '23 研讨会, 2023 年 8 月, 研讨会 ©
2023 计算机协会。
ACM ISBN xxx-x-xxxx-xxxx-X/xx/xx。 。 。 \$xx
<https://doi.org/xxxx/xxxx>

1 简介

股票交易量变动预测是金融领域备受关注的基础任务之一[1,4,26],它具有各种重要的下游应用,例如算法交易[13,39]和股票交易异常检测[32]。

传统研究仅使用历史交易数据并严重依赖特征工程。他们使用统计模型来预测时间序列,例如自回归综合移动平均模型 (ARIMA) [2, 32]。随着深度学习技术的发展,Garehchopogh 等人。 [12]使用线性回归进行股市交易量预测。利用 LSTM 和 CNN 的更复杂的模型也应用于该领域[6,30,38]。然而,这些方法仅涉及历史交易数据,缺乏基本的股票信息。人类交易员

他们根据多种因素做出决定,包括股票相关新闻。因此,仅使用历史交易数据作为输入的单峰模型可能会做出错误的预测。因此,研究人员开始引入新闻和推文等文本信息,以更好地

对股票走势进行建模。通常,情感分析模块用于反映市场情绪 [22,31,33,35]。这些方法通常采用管道架构,两种模式没有正确集成。情绪预测可能会涉及其他错误。最近的作品设计了多模态模型来联合处理文本和时间序列数据,以获得更好的理解。例如,李等人。 [20]建议使用事件驱动的 LSTM 模型来利用新闻数据。 Zou 和 Herremans [47]使用 CNN 和 SVM 作为骨干设计了一种混合多模态模型。

先前作品的一个显着弱点是它们倾向于构建某种架构并从头开始训练。目前的高质量财经新闻数据集的规模往往比从互联网收集的大规模未标记语料库要小得多,因为编写、收集和过滤获取相关新闻的成本很高。因此,基于它们训练鲁棒的大型多模态模型非常困难,与 Fin-BERT [43]和 ChatGPT1 等预训练语言模型相比,其能力较差。

此外,由于财经新闻的主题和内容可能很广泛,文本学习过程需要领域知识和通用知识,因此在这项任务中需要结合预先训练的语言模型。

*平等贡献1 <https://openai.com/product/chatgpt>

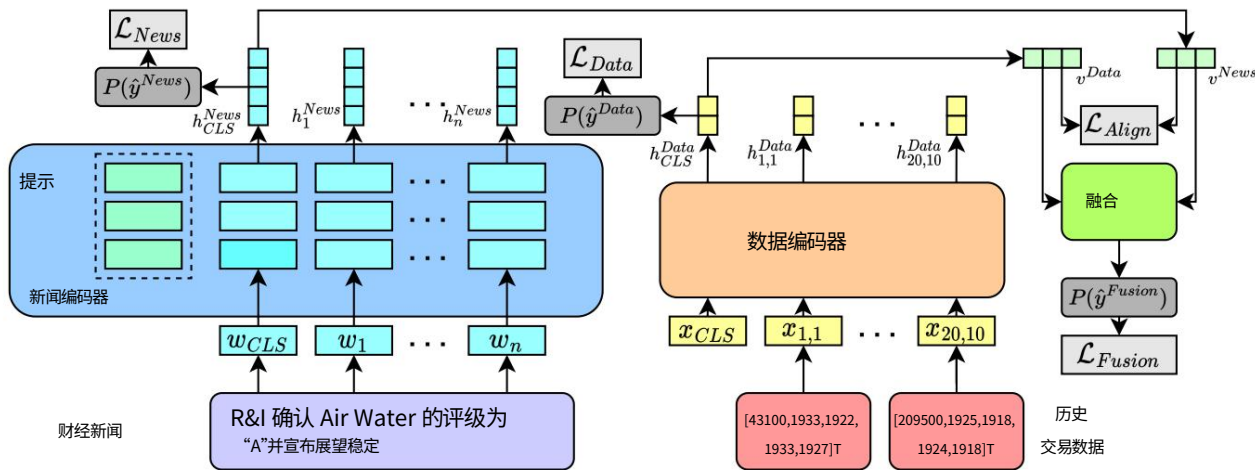


图 1: 我们的模型概述。财经新闻和历史交易数据分别编码。新闻编码器采用冻结的 Financial-RoBERTa 作为骨干, 只有所有层中的连续提示是可调的。数据编码器采用预训练的 6 层变压器, 该变压器将在训练期间进行微调。融合模块利用新闻和交易数据表示来获得综合预测。对齐损失是为跨模态对比对齐而设计的, 并防止训练期间对单模态表示的损坏。

为了解决缺乏必要知识的问题, 我们提出了基于提示的多模态股票成交量变动预测模型 (ProMUSE), 其中包括新闻编码器、数据编码器和融合模块。如图 1 所示, 新闻编码器使用预训练的语言模型作为骨干, 我们使用即时学习方法来有效地利用预训练模型中的知识。然而, 两个编码器的直接融合将严重损害两种模态的表示学习。因此, 我们建议为每个编码器保留一个预测头来接收单模态监督。特别地, 我们添加了跨模态对比对齐损失来对齐嵌入空间, 减轻联合训练期间对其表示的损害。为了进行推理, 我们使用两个单模态头部预测的算法平均值和通过融合模块的多模态融合结果。这使得我们的模型甚至可以使用单峰输入稳健地生成输出。

我们在 TOPIX500 交易数据集和路透社的隔夜财经新闻[21]上验证了我们提出的 ProMUSE 模型。我们广泛的实验表明, 我们提出的 Pro- MUSE 优于单峰方法和多峰基线, 但存在显著差距。我们还进行了一系列消融研究来分析模型中的不同模块。我们表明, 仅融合的方法表现更差, 因为它们缺乏单峰监督, 因此会对单峰表示造成损害。另一方面, 仅集成的方法失去了跨模态对比对齐。我们的方法实现了多模态融合和对齐, 保留单模态预测头, 以减轻对表示学习的损害, 从而达到最佳结果。此外, 我们提出的 ProMUSE 可以在不同的数据大小下一致地达到比从头开始训练的方法更高的性能。

我们的主要贡献可概括如下:

- 我们将预训练的语言模型提示引入多模式库存量变动预测任务中, 以带来

数据集有限的领域知识和普遍知识。

我们提出了一种多模式方法 ProMUSE, 以整合来自文本和历史交易数据的知识。单模态和多模态监督训练与跨模态对比对齐一起使用, 以减轻多模态学习期间对表示学习的损害。

实验结果表明, 我们的方法明显优于竞争基线, 并且进一步的分析验证了我们在 ProMUSE 中提出的模块的有效性。

2 方法论

在本节中, 我们首先介绍多模式库存量变动预测的任务制定。然后我们详细说明了多模态股票成交量变动预测模型的架构、跨模态对齐的对比损失, 并最后提出了我们的训练目标和推理算法。

2.1 任务制定

隔夜股票成交量变动预测任务是一个二元分类问题: 标签 1/0 表示下一个交易日成交量上涨/下跌。该模型的分类决策基于隔夜新闻和过去 20 天的股票历史交易数据, 这两者都作为模型的输入。

隔夜新闻。按照 [21, 46], 我们只采用新闻标题来完成我们的任务, 因为它们的信息量更大, 并且具有适合处理的长度。隔夜新闻可以被建模为带有 tokens = {1, 2, ...} 的输入句子。隔夜消息发生在 20 日交易市场收盘后、次日开盘前。

股票历史交易数据。整体股票历史交易情况数据包括过去20天的交易量和价格每天10个时段,粒度为30分钟。特定时段的交易数据包括成交量、最高价、最低价、开盘价、收盘价。这,整体股票历史交易数据可表述为 $\{|\in[1,20]\cap\mathbb{Z},\epsilon\in[1,10]\cap\mathbb{Z}\}$,其中表示第-天第-个时段的交易数据,即 $[v_{t,i}^m]_{t,i}\in\mathbb{R}^5$ 。

我们的目标是预测第一个时段的交易量变动第21天,即开市后30分钟21,1,市场。继赵等人之后。[46],我们定义运动这里作为与平均体积之间的比较 $\hat{v}_{21,1}$ 过去20天的同一时间段,因此最终的预测目标标签可以表示为:

$$= \text{我} \quad \hat{v}_{21,1} > - = \begin{matrix} 1, & \text{如果} \\ 0, & \text{如果} \end{matrix} \quad \hat{v}_{21,1} \leq , \quad (1)$$

其中表示过去 20 个时段中第一个时段的平均交易量天,定义为: $= \frac{1}{20} \sum_{t=1}^{20} v_{t,1}$ 。那么我们的数据集可以表示为 $D = \{((\text{ } , \text{ }), (\text{ }))\} = 1$ 。我们采用准确度 (ACC)指标来评估性能:

$$\text{ACC} = 100\% \times \sum_{i=1}^n \text{我arg最大值} (|\text{ }^{(i)} , \text{ }^{(i)}) = \text{ }^{(i)} , \quad (2)$$

其中表示输入历史交易数据,表示存相关新闻,表示目标标签, $(|\text{ }^{(i)})$ 表示预测概率, $\arg \max (|\text{ }^{(i)} , \text{ }^{(i)})$ 去库预测标签。

2.2 模型架构概述

我们提出的ProMUSE主要包括两个大规模预训练处理输入的模块:用于文本模态的新闻编码器和用于时间序列模态的数据编码器。他们俩编码器可以有效地捕获输入的特征并将它们转化为相应的向量。他们还用他们的预测头产生单峰损失,预测接受单峰监督,这将在2.3节中介绍

和第 2.4 节。在2.5节中,我们进一步利用融合模块用于产生多模态损失和输出。跨模态2.6 节中描述的对比对齐旨在帮助改善多模态学习期间的单模态表征。最后,我们使用两个单峰损失、多峰损失和对齐损失的加权和作为第2.7 节中的训练目标。在2.8节的推理过程中,单峰和多峰

预测被结合起来。模型架构概述

如图1所示。

2.3 新闻编码器

新闻编码器是 Financial-RoBERTa2, 24 层 RoBERTa [27]对财务文本 (例如财务报表)进行预训练的模型,新闻和收益公告。为了避免过度拟合2https://huggingface.co/soleimanian/financial-roberta-large-sentiment

有限的新闻数据以及加快训练过程,我们采用即时学习方法并选择P-Tuning v2 [25]作为它。在我们的初步实验中达到了最佳性能。这种方法在基于 Transformer 的每一层中插入软提示。我们将提示长度设置为 20 并启用重新参数化。在我们的实验中,只有提示是可调的,而语言模型中的所有其他参数都被冻结。

给定包含令牌的输入新闻,我们插入一个 CLS 句子开头的标记: $\{\text{CLS}, 1, 2, \dots\}$ 。然后我们利用 Financial-RoBERTa 和 P-Tuning v2 方法将其转换为一系列向量 $[\text{消息}_{\text{CLS}}, \text{新闻}_1, \text{新闻}_2, \dots, \text{新闻}]$:

$$\{\text{消息}_{\text{CLS}}, \dots, \text{新闻}\} = \text{新闻编码器} (\{\text{CLS}, \dots, \text{新闻}\}), (3)$$

哪里消息 $\in \mathbb{R}^{\text{新闻}}$,新闻是语言的隐藏大小模型,并且在我们的设置中News = 1024。我们进一步获得单峰预测^新闻消息符合-正在消息CLS通过线性分类头Linear-HeadNews来预听写每个类别的分数,Linear-HeadNews消息CLS $\in \mathbb{R}^2$:

$$\text{消息}_{\text{CLS}} = \text{Softmax Linear-HeadNews}_{\text{消息}_{\text{CLS}}}。 (4)$$

新闻编码器头的单峰损失是熵损失:

$$\text{LNews} = - \log \text{新闻}_{\text{消息}_{\text{CLS}}}, (5)$$

2.4 数据编码器

为了获得历史交易数据的高质量表示,我们预训练了 6 层 Transformer [40]模型。回想起那个历史交易数据= $\{|\in[1,20]\cap\mathbb{Z},\epsilon\in[1,10]\cap\mathbb{Z}\}$, $\epsilon\in\mathbb{R}^5$, 因此,输入时间序列可以表示为 $[\text{CLS}; 1,1; 1,2; 1,3; \dots; 2,1; \dots; 20,10] \in \mathbb{R}^{201 \times 5}$ 在时间或-der,其中时隙号为 $1 + 20 \times 10 = 201$ 。数据编码器将输入序列转换为隐藏向量序列 $[\text{数据}_{\text{CLS}}; \text{数据}_{1,1}; \text{数据}_{1,2}; \text{数据}_{1,3}; \dots; \text{数据}_{2,1}; \dots; \text{数据}_{20,10}] \in \mathbb{R}^{201 \times \text{数据}}$:

$$[\text{数据}_{\text{CLS}}; \dots; 20,10] = \text{数据编码器} [\text{CLS}; \dots; 20,10], (6)$$

哪里数据 $\in \mathbb{R}^{\text{Data}}$ 。这里Data是 Trans- $\epsilon \mathbb{R}^{\text{Data}}$ 的隐藏大小。前者,我们设置Data = 200。在预训练阶段,我们在之后使用线性头直接预测日子。我们采 $\hat{v}_{21,1}$, 21号第一个时段的成交量用均方误差 (MSE)损失进行优化:

$$\text{LMSE} = (\text{线性预}(\text{数据}_{\text{CLS}}) - \hat{v}_{21,1})^2。 (7)$$

对于多模式库存量变动预测,类似于新闻编码器,我们得到单峰预测^数据消息CLS和单峰损失LData如下:

$$\text{数据}_{\text{CLS}} = \text{Softmax 线性头数据}_{\text{数据}_{\text{CLS}}}, (8)$$

$$\text{LData} = - \log \text{数据}_{\text{数据}_{\text{CLS}}}, (9)$$

其中Linear-HeadData是 Data 的线性分类头利用编码器和交叉熵损失。我们在期间继续微调所有 Transformer 参数多模式培训。News Encoder采用的提示方式

这里不保留,因为 6 层 Transformer (1.7M 参数)的尺寸比 RoBERTa 模型 (335M 参数)。

2.5 新闻编码器和数据编码器的融合

News Encoder 和 Data Encoder 为我们提供了文本特征消息和时序特征。我们建立我们的在这些功能之上的多模态融合块。我们首先线性项目消息,数据进入消息,数据位于一个共同的embedding空间 R 对齐:

新闻,数据 = LinearNews(新闻 CLS),线性数据(数据 CLS)。 (10)

消息,数据接收来自对比的监督信号对齐,这将在下一节中讨论,以学习多模态表示。

融合预测*融合和融合损失LFusion是一个通用的融合过程Fusion(News, Data)。吃过消息,数据同样,也采用交叉熵损失:

融合新闻,数据 = Softmax Fusion (新闻,数据), (11)

LFusion = - log 融合新闻,数据。 (12)

我们在实验中实现了几种融合方法,我们发现线性融合函数达到了最好的效果结果,如我们的分析所示。线性融合函数设计为:

FusionLinear (新闻,数据) = 新闻动态+ 数据 数据+。 (13)

在哪里消息,数据 ∈ R 2×对齐, ∈ R 2。

2.6 跨模态对比对齐

对比学习是多模态中广泛使用的技术区域。CLIP [36]和 ALPRO [18]等方法使用跨模态对比损失以对齐不同嵌入中的表示空间。在我们的模型中,我们实现了跨模态对比隔夜新闻和历史交易数据之间的一致性对准损失。

只有批次中匹配的对才被视为正对。给定一批对消息 = [消息1, ..., 消息n], 数据 = [数据1, ..., 数据n] ∈ R

我们使用以下[15,18,36]的点积来定义相似度:

西姆(,) = 消息 · 数据, (14)

以及两个对称的 News-to-Data (N2D)、Data-to-News (D2N) 对准损耗随后定义为:

L N2D 对齐 = - 1 / Σ 日志 (exp Sim(,)/exp Sim(,)/=1, (15)

L 夜到夜 对齐 = - 1 / Σ 日志 (exp Sim(,)/exp Sim(,)/=1, (16)

这里是温度参数。总跨模态对比对齐损失为:

L 对齐 = L N2D 对齐 + LD2N (17)

请注意,以前的作品使用跨模态对比损失提高图像分类等检索任务的性能,有些甚至报告对其他任务造成损害[18]。

然而,在我们的模型中,LAlign旨在改进不同模态的表示学习,以及定义的相似度Sim 不用于推理。

此外,文本和历史数据的嵌入空间可以是显着不同,因为它们之间的联系不那么紧密作为图像文本或视频文本场景中的典型设置。文本数据模式之间的深层联系可以导致模型学习

不正确的关系并陷入如图所示的过度拟合陷阱在我们的分析实验中。我们发现,作为相对简单的跨模态对比对齐损失应用于输出编码器,它鼓励多模态融合和对齐且不损害大型模型的结构。

2.7 培训目标

我们将上述四种不同的损失结合起来作为我们的最终结果培训目标。两个损失LNews,LDData分别来自uni-。而且,仅基于的模态编码器消息,数据 CLS,CLS我们利用融合来构建目标标签分布的多模态预测。最终训练目标

是加权和:

L = NLNews + DLData + FLFusion + ALAlign。 (18)

2.8 推论

我们的预测结果也包含上述三个要素前。我们使用等加权平均值来生成集成预测概率:

() = (消息_ + 数据_ + ^融合) / 3, (19)

最终的预测标签由以下方式生成:

^ = arg 最大值(|,), (20)

其中 表示输入历史交易数据,表示与股票相关的财经新闻。

请注意,如果两者之一,我们的方法仍然可以发挥作用缺少模态输入。这可以通过简单地禁用相应的编码器,使我们的模型更加鲁棒不同的输入情况。

3 实验

在本节中,我们首先介绍数据集,然后描述基线算法、详细设置和实验结果。

3.1 数据集和数据处理

历史交易数据提取自TOPIX500,即由流动性最强、市值最高的 500 家公司组成东京股票。我们按时间顺序分割数据集以避免信息泄露。2013年1月1日至2017年12月31日数据

将预训练模型提示纳入多模式库存量变动预测

KDD 23 研讨会,2023 年 8 月,研讨会

表 1:股票走势预测的数据集统计。

分裂	训练开发测试
样品8,483	687 938

表 2:数据编码器预训练的数据集统计数据。

分裂	训练开发测试
样品74,950 2,214 4,072	

以2018年1月1日至2018年4月30日为训练集,2018年5月1日至2018年9月30日为测试集。在数据处理过程中,我们删除缺少条目的数据点。隔夜新闻摘自路透社财经新闻3。继[5,21,46]之后,数据用路透社提供的RIC标签进行过滤,这些标签是可能受到该消息影响的股票。

在本文中,我们仅选择隔夜新闻和历史交易数据均可用于股票走势预测的数据。此外,我们过滤掉成交量变动不够显著的数据[46],减轻随机性和次要的、不相关的新闻的影响:

$$= \frac{1}{20} \sum_{i=1}^{20} \left(\frac{1}{\sigma_i} \right)^2, \quad \sigma_i = \frac{21,1}{\sigma_i}.$$

如果 $\left| \frac{1}{\sigma_i} \right| \leq 0.5$,我们认为其体积变化不显着,并将其从数据集中删除。决赛统计信息

数据集见表1。

为了预训练我们的数据编码器,我们使用与股票走势预测数据集相同的分割。数据集统计数据如表 2 所示。

3.2 基线

在本节中,我们介绍基线模型的算法,包括传统统计方法、单峰模型、融合方法和集成方法。

3.2.1 统计方法。

随机:随机预测标签 $(1, 1/2)$ 。指数移动均线(EMA):在此任务中,EMA 系列定义为:

$$\text{指数移动均线} = \frac{1}{21} (2 \text{EMA}_{-1} + 19 \text{EMA}_0),$$

其中EMA1被初始化为21,1的预测。预测 \hat{y}_t 我们使用EMA20作为1,1, $\hat{y}_t = 1$ 是 $(\text{EMA20} > \hat{y}_t)$ 。

3.2.2 从头开始进行模型训练。我们尝试从头开始训练一个六层Transformer,隐藏大小设置为 200,以替换News Encoder 中预先训练的 Financial-RoBERTa。关于数据编码器,我们尝试了多种结构,包括线性、一层 LSTM [14]和六层 Transformer。详细结构与Zhang等人类似。[45]。我们还从头开始探索用于模型训练的不同融合或集成范例。

3<https://github.com/liweitj47/overnight-stock-movement-prediction>

3.2.3 单峰模型。对于单峰模型,我们仅使用其中一个编码器。结果,最终的训练目标和预测概率分布退化为只有LNews,

消息 \mathbf{x}_{news} 或者 L数据, \mathbf{x}_{data} 。

3.2.4 融合方法。

·注意:使用注意机制产生融合的权重。 $1, 2 \in \mathbb{R}$ Align由Data线性导出: \mathbf{w}_{news} 新闻和

$$1, 2 = \text{Softmax}(\text{Linear}([\text{新闻}, \text{数据}])),$$

融合函数为:

$$\text{FusionAttention} = \text{线性1} \odot \text{新闻} + 2 \odot \text{数据}.$$

这里 \odot 代表逐元素乘法。

·Transformer:将多模态 Transformer 编码器堆叠在两个编码器的顶部。变换到相同维度后,将它们的最外层隐藏状态作为输入,并使用线性分类头进行预测。

3.2.5 集成方法。在本节中,我们介绍基于单峰模型预测 (新闻)和数据的集成方法。

·可学习权重:我们将 $\epsilon \in \mathbb{R}$ 设置为可学习参数,并将其初始化为1。公式为(Learnable) $\hat{y}_t = \text{Sigmoid}(\epsilon) (\text{News}) + (1 - \text{Sigmoid}(\epsilon)) (\text{Data})$ 。·预测权重:使用线性头来预测两种模式之间的权重:

$$1, 2 = \text{Softmax}(\text{线性}([\text{消息}_{\text{CLS}}, \text{数据}_{\text{CLS}}])),$$

然后 $(\hat{y}_t^{\text{预测}}) = \frac{1}{\text{Norm}} (\hat{y}_t^{\text{新闻}}) + 2 (\hat{y}_t^{\text{数据}})$ 。·归一化:将单峰结果归一化为标准化高斯分布: $(\hat{y}_t^{\text{Norm}}) = \text{Norm}(\frac{1}{\text{Norm}} (\hat{y}_t^{\text{News}})) + \text{Norm}(\frac{1}{\text{Norm}} (\hat{y}_t^{\text{Data}})) + 0.5$,其中 $\text{Norm}() = (\cdot - \mu) / \sigma$ 和超参数 μ 和在开发集上计算相应的概率 $(\hat{y}_t^{\text{新闻}})$ 或 $(\hat{y}_t^{\text{数据}})$ 。

3.3 设置和超参数我们对实验中的每个模型进

行 40 个 epoch 的训练,并报告测试精度作为结果。为报告选择开发集上精度最高的检查点。我们将每个实验重复 4 次并报告平均结果。

在我们的主要实验中,我们采用 AdamW 优化器[28],使用学习率为 $1e-5$,权重衰减因子为 $1e-3$ 。批量大小选择为 32。每个损失的权重设置为 $\omega = 1$ 并设置为 200。我们使用Align 作为 $\omega = 0.1$ 。对齐股票历史交易数据的日志值作为数据编码器的输入。

对于数据编码器的预训练,我们运行 100 个 epoch ,并为主要实验保存开发损失最低的检查点。我们还使用 AdamW 优化器,学习率为 $1e-5$,权重衰减因子为 $1e-3$ 。使用系数为 0.95 的附加指数衰减。预训练批量大小设置为 64。

表 3:主要结果。现有方法主要可分为统计方法、单峰方法、仅融合方法和仅集成方法方法,而我们提出的 ProMUSE 实现了最先进的性能。

方法		新闻编码器	数据编码器	ACC
统计	随机的	-	-	50.00
	20 天均线	-	-	59.38
单峰 (消息)	训练 从头开始	变压器	-	58.26
	零射击	金融-RoBERTa -		48.82
	微调	金融-RoBERTa -		62.19
	迅速的	金融-RoBERTa -		63.75
单峰 (数据)	训练	-	线性	58.93
	从头开始	-	长短期记忆网络	54.78
		-	变压器	66.07
	零射击	-	预训练变压器	64.71
	微调	-	预训练变压器	68.44
仅融合	训练	变压器	线性	61.38
	从头开始	变压器	长短期记忆网络	57.70
		变压器	变压器	60.16
	迅速的	金融-RoBERTa预训练变压器		72.47
仅合奏	训练	变压器	线性	58.48
	从头开始	变压器	长短期记忆网络	57.92
		变压器	变压器	60.60
		变压器	线性+LSTM+变压器	61.94
	提示 (平均)金融-RoBERTa 预训练 Transformer			72.28
普罗缪斯	迅速的	金融-RoBERTa预训练变压器		73.56★

3.4 实验结果

我们的主要实验结果见表3。

3.4.1 从头开始训练。我们采用六层 Transformer 作为新闻编码器中 Financial-RoBERTa 模型的替代方案。

就数据编码器而言,我们探索线性、LSTM 和 Transformer模型作为预训练 Transformer 的替代品。这结果表明,可用数据量不足以

从头开始训练新闻编码器,而我们的方法成功地利用了预先训练的金融 RoBERTa 的知识。此外,与 Transformer 相比,线性和 LSTM 结构的性能较差。雇用的好处

预训练的 Transformer 模型也很明显。

3.4.2 单峰模型。单峰模型表现显着

比多模态方法更糟糕,因为只有单个信息模态被提供给他们,两种模态在这方面都很重要任务。此外,我们发现提示 Financial-RoBERTa 可以更有效地利用其知识并表现得更好微调。此外,对历史交易数据进行建模更容易比财经新闻,因此纯数据方法可以取得更好的效果性能优于新型号。

3.4.3 仅融合和仅集成方法。直接多式联运

与单峰方法相比,融合或集成可以取得改进。然而,没有单峰的仅融合方法

监督可能会对单一模式表示造成损害,并且仅集成方法缺乏多模态连接。我们的方法通过构建融合和跨模态来解决这些问题对比对齐以及保留单峰预测

帮助表征学习。

3.4.4 ProMUSE 的有效性。如表 3 所示,ProMUSE 可以达到最佳性能。我们成功超越了金融领域广泛使用的传统 EMA 基线,并展示了我们的

相对于从头开始训练模型的优势,单峰模型、仅融合方法和仅集成方法。

4 分析

在本节中,我们首先进行消融研究以证明我们模型中每个模块的有效性。然后讨论更多

详细部分,对调优范式、融合模型、集成算法和提示学习方法的探索。最后,我们

表明我们的方法有助于表征学习过程。

4.1 消融研究

4.1.1 不同损失和预测头的有效性。我们进行各种消融实验以测试其有效性

我们提出的 ProMUSE 模型中的每个组件如表 4 所示。结果证明融合、集成和跨模态对比对齐达到最佳性能。

表 4:消融研究结果。

方法	LNews LData LFusion LAlign (新闻) (数据) (融合)	ACC
仅融合	✓✓✓ × × ×	72.47
仅合奏	✓ × × × ✓ ×	72.28
不含新闻头	× ✓ ✓ × × ✓	72.15
无数据头	× × ✓ ✓ × × ✓	69.91
不含新闻/数据	× × ✓ × × ×	70.55
无融合	✓ × × ✓ × ×	67.80
无对齐	✓ ✓ × × ✓ ×	70.52
不含合奏	✓ ✓ × × × ×	73.35
ProMUSE	✓ ✓ ✓ ✓ ✓ ✓	73.56

表 5:不同融合模型的结果。以下型号

是没有单峰监督的仅融合模型。这
线性融合优于其他复杂的融合。

融合模型	ACC
注意力	67.00
一层 Transformer 六	65.43
层 Transformer 将数	69.11
据融合到 News Encoder 63.81 的最后一层	
将数据融合到 News Encoder 63.86 中的每一层	
线性	72.47

仅集成方法可能会导致精度下降
由于跨模态交互丢失,从 73.56 下降到 72.28。这
仅融合设置也会导致精度下降至 72.47,因为
没有LAlign对单峰表示的影响更大。
单峰头掉落会严重损害性能
因为这两种方式缺一不可。去除融合或
对齐过程也打破了模式之间的联系。
请注意,在推理阶段,预测(Data)和
(News)不需要太多额外的计算
(Fusion)也需要两个头的计算。预测
他们的合奏可以从 73.35 提高到 73.56。

4.1.2 融合模型的有效性。在本节中,我们进一步
探索表5中不同的融合模型。线性融合方法
我们的提案中采用的方案优于所有其他变体。
这里Attention的描述如公式24,一层/六层
Transformer 堆叠在两个编码器上,如中所述
基线部分。将数据融合到新闻编码器方法中使用
数据 CLS 为新闻编码器生成连续提示。
原因是两个编码器都是预先训练好的,不会
很容易过度拟合小数据集的知识,而
复杂的融合方法,例如引入另一个 Trans-former 模型,是从头开始训练的,这会产生额外的
参数,可能导致不正确的连接和过度拟合。

4.1.3 集成算法的有效性。我们在图 2 中测试了不同集成算法的有效性。在我们对仅集成模型的实验中,我们发现平均新闻
预测和数据预测表现最好。依据

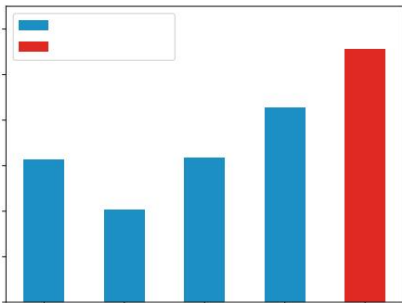


图 2:不同集成算法的结果。两个单峰输出的平均值超过了其他算法。

表 6:不同提示范例的结果。 P-调整 v2

优于硬提示、软提示和微调。

模型	调整方法ACC	
新闻编码器	硬提示	48.82
	软提示	56.08
	微调	62.19
	P-调整 v2	63.75
数据编码器	零射击	64.71
	微调	68.44
	普罗缪斯	66.79
	迅速的	73.56★

我们之前对融合模型的分析,复杂的集成算法也往往会降低性能,因为集成中涉及的新参数是在小数据集上估计的,这

很难概括并可能导致过度拟合。因此我们
选择计算(新闻)、(数据)和的算法平均值
(Fusion)用于我们模型中的推理。

4.2 为什么我们使用 P-Tuning v2 范式

为我们的新闻编码器寻找合适的范例
数据编码器,我们进行表6中的实验。我们在实验中用于硬提示的模板是
News:
{ 1, 2, . . . }。音量将会升高/降低。这里有软提示
表示在嵌入中实现连续提示
层,类似于Prefix-Tuning [23]和P-Tuning [24]。
我们发现 P-Tuning v2 最适合 News Encoder 并且
准确度达到 63.75,插入连续提示
由于有更多的可训练参数和更深的层,因此在每一层中。
微调数据编码器可以得到很大的改进
零样本设置,但微调大型 Financial-RoBERTa
数据有限的模型很容易陷入过拟合问题
并导致次优,准确率下降至 66.79。

4.3 ProMUSE 在较低资源下运行

我们进一步分析基于提示的多模式的效果
通过改变训练数据的大小来建立模型。发展历程

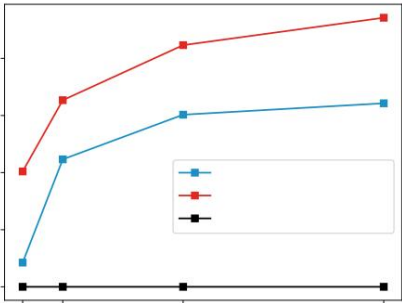


图 3:不同数据大小的结果。 ProMUSE 表现出色
在较低资源下从头开始的方法培训。

表 7:多模态学习对人的损害分析
单峰表示。现有方法会破坏单模态表示,而 ProMUSE 可以减轻多模态学习对单模态表示的损害。

模型	仅限新闻	仅限数据	多模式
新闻编码器	63.75	-	63.75
数据编码器	-	68.44	68.44
不带对齐 不带融合	58.32	61.09	70.52
	62.60	63.65	67.80
普罗缪斯	62.69	67.38	73.56

集和测试集保持不变,我们训练所有模型80 epochs 来补偿数据量的减少。

这里我们将 ProMUSE 与六层的性能进行比较 Transformer 从头开始训练,只接收历史交易数据输入,因为它在我们之前的测试中表现最好

实验。由于数据量持续萎缩,无法获得高质量的编码器,因此我们放弃了 News Encoder

这些情况。结果如图3所示。我们发现 ProMUSE 可以实现比以下培训显著的改进

在所有设置中都从头开始,证明通过即时学习来利用预训练语言模型的通用知识是

在这项任务中至关重要。

4.4 ProMUSE助力表征学习

在本节中,我们讨论为什么我们的提案可以实现改进。如表 7 所示,多模态学习的引入

与单峰模型相比可以获得总体收益。然而,

同时训练两种模式可能会损害各个编码器的表示学习,因为我们看到

仅新闻或仅数据输入场景的准确性。

我们的模型使用跨模态对比对齐和多模态融合,更好地减轻造成的损害

多模态学习过程,提供了强有力的约束和对单峰编码器进行正则化,以最好地避免退化。

5 相关工作

5.1 预训练语言模型

由于 Transformer [40]在自然语言方面取得了巨大成功

在处理 (NLP)领域,提出了各种预训练语言模型,并在众多 NLP 中实现了最先进的性能

任务。和 GPT-3 [3]等模型利用 Transformer 解码器

构造单向往回归语言模型的结构。双向 BERT 类模型[11,16,27]基本上基于

在 Transformer 编码器上。 T5 [37]、 BART [17]和 Flan-T5 [10]

选择采用编码器-解码器框架。他们受过训练

在相应的预训练任务和大型未标记语料库上,

为了追求更好的性能,模型的大小变得越来越大[9, 44] 。最近,人类强化学习

反馈技术应用于 ChatGPT 和 GPT-4 [34] ,

获得优异的业绩。

5.2 多模态学习

现有的多模态学习方法主要集中在图像-文本和视频-文本任务。

UNITER [8]学习联合情境化

通过预训练来表示文本和图像。 ViL- BERT [29]通过共同注意 Transformer 层扩展了 BERT 模型,用于学习与任务无关的图像表示

内容和自然语言。 ALIGN [15]和 CLIP [36]构造

双编码器架构使用图像-文本对比学习来对齐视觉和文本表示,以及跨模态

对比损失已成为许多模型中的重要组成部分。 BLIP-2 [19]使用轻量级

查询 Transformer 来学习

来自冻结图像编码器和大型语言模型。

5.3 库存走势预测

股票走势预测是金融领域的一个重点研究方向。 Xu 和 Cohen [42]提出了一种深度生成模型

共同从推文文本和价格信号中学习,并使用循环潜变量来处理随机性。李等人。 [21]设计一个

用于学习隔夜新闻及其相关性的 LSTM-RGCN 模型

股票之间。陈等人。 [7]建立双进程元学习

挖掘一般模式和股票特定知识的方法。谢等人。 [41]分析ChatGPT在多模态下的零样本能力

股票走势预测。

六,结论

在本文中,我们提出了 ProMUSE,一种基于提示的多模式

股票走势预测模型,充分挖掘财经新闻中的文本信息和预训练的潜力

数据有限的语言模型。我们使用新闻编码器和

数据编码器处理隔夜财经新闻和历史

分别交易数据,并使用融合模型生成多模态输出。单模态监管、多模态监管

跨模态对比对齐用于训练,而

单模态和多模态预测构成最终的推论

结果。大量实验表明我们的方法显着

优于各种基线。综合分析证明

不同模块的有效性。此外,ProMUSE 可以帮助

减轻联合训练期间表征学习的危害

文本和时间序列模式。

参考

[1] Bipin B Ajinkya 和 Prem C Jain. 1989.每日股市交易量的行为.会计与经济学杂志 11, 4 (1989), 331–359.

[2]乔治·EP·博克斯和大卫·A·皮尔斯. 1970.自回归积分移动平均时间序列模型中残差自相关的分布.美国统计协会杂志65, 332 (1970), 1509–1526.

[3] Tom Brown,Benjamin Mann,Nick Ryder,Melanie Subbiah,Jared D Kaplan、Prafulla Dhariwal,Arvind Neelakantan,Pranav Shyam,Girish Sastry,Amanda Askell 等。2020 年.语言模型是小样本学习者.神经信息处理系统的进展33 (2020), 1877–1901. [4]阿尔瓦罗·卡泰亚和塞巴斯蒂安·贾蒙加尔. 2016.以成交量加权平均价格为目标的封闭式执行策略. SIAM 金融数学杂志7, 1 (2016), 760–785.

[5]陈德利,邹艳艳,张本惠子,鲍瑞涵,任宣成,孙旭. 2019.将细粒度事件纳入股票走势预测. arXiv预印本 arXiv:1910.05078 (2019).

[6]陈凯,周毅,戴芳艳. 2015. 基于 LSTM 的股票收益预测方法 :以中国股票市场为例. 2015年 IEEE大数据国际会议 (big data) 。 IEEE,2823–2824.

[7] 陈瑞波,李伟,张志远,鲍瑞涵,张本惠子,孙旭. 2022.利用双进程元学习进行股票交易量预测.在欧洲机器学习和数据库知识发现联合会议中.施普林格,137-153.

[8]陈彦春,李林杰,于立成,Ahmed El Kholy,Faisal Ahmed,甘哲、程宇和刘晶晶. 2019.Uniter: 学习通用图像文本表示. (2019)。

[9] Aakanksha Chowdhery,Sharan Narang,Jacob Devlin,Maarten Bosma,Gaurav Mishra,Adam Roberts,Paul Barham,Hyung Won Chung,Charles Sutton,Sebastian Gehrmann 等。2022. Palm :通过路径扩展语言建模. arXiv 预印本 arXiv:2204.02311 (2022).

[10] Hyung Won Chung,Le Hou,Shayne Longpre,Barret Zoph,Yi Tay,William Fedus,Eric Li,Xuezhi Wang,Mostafa Dehghani,Siddhartha Brahma 等。2022.扩展指令微调语言模型. arXiv 预印本 arXiv:2210.11416 (2022).

[11] Jacob Devlin,Ming-Wei Chang,Kenton Lee 和 Kristina Toutanova. 2018. Bert :用于语言理解的深度双向变压器的预训练. arXiv预印本 arXiv:1810.04805 (2018)。

[12] Farhad Soleimanian Gharehchopogh,Tahmineh Haddadi Bonab 和 Seyyed Reza Khaze. 2013.预测股市交易量的线性回归方法 :案例研究.国际管理价值和供应链杂志4, 3 (2013), 25.

[13]泰伦斯·亨德肖特和瑞安·赖尔丹. 2013.算法交易和流动性市场.金融与定量分析杂志 48, 4 (2013), 1001–1024.

[14]塞普·霍赫赖特和尤尔根·施米德胡贝尔. 1997.长短期记忆.神经计算 9, 8 (1997), 1735–1780.

[15]贾超,杨银飞,夏叶,陈一婷,Zarana Parekh,Hieu Pham,Quoc Le、Yun-Hsuan Sung,Zhen Li 和 Tom Duerig. 2021.通过噪声文本监督扩大视觉和视觉语言表征学习.在国际机器学习会议上. PMLR,4904–4916。

[16]兰振中,陈明达,塞巴斯蒂安·古德曼,凯文·金佩尔,Piyush Sharma 和 Radu Soricut. 2019.Albert :用于语言表征自我监督学习的 lite bert 。 arXiv 预印本 arXiv:1909.11942 (2019)。

[17] Mike Lewis,Yinhan Liu,Naman Goyal,Marjan Ghazvininejad,Abdelrahman Mohamed、Omer Levy,Ves Stoyanov 和 Luke Zettlemoyer. 2019. Bart :用于自然语言生成、翻译和理解的序列到序列去噪预训练. arXiv 预印本 arXiv:1910.13461 (2019)。

[18]李东旭,李俊楠,李洪东,胡安·卡洛斯·尼布尔斯,史蒂文·CH·霍伊. 2022.对齐和提示 :带有实体提示的视频和语言预训练. IEEE/CVF 计算机视觉和模式识别会议论文集. 4953–4963.

[19]李俊楠,李东旭,Silvio Savarese,Steven Hoi. 2023. Blip-2 :使用冻结图像编码器和大型语言模型引导语言图像预训练. arXiv 预印本 arXiv:2301.12597 (2023)。

[20]李青,谭景华,王军,陈新春. 2021.使用在线新闻进行股票预测的多模态事件驱动 LSTM 模型. IEEE知识与数据工程汇刊 33, 10 (2021), 3323–3337. <https://doi.org/10.1109/TKDE.2020.2968894>

[21]李伟,鲍瑞涵,张本惠子,陈德利,徐晶晶,苏琪. 2021 年. 使用图网络对股票关系进行建模,以预测隔夜股票走势.载于第二十九届人工智能国际联合会会议论文集. 4541–4547。

[22]李晓东,谢浩然,陈力,王建平,邓小铁. 2014.通过情绪分析新闻对股价回报的影响.基于知识的系统69 (2014), 14–23.

[23]李翔,梁珀西. 2021.前缀调优 :优化生成的连续提示. arXiv 预印本 arXiv:2101.00190 (2021)。

[24]刘晓,季凯旋,付一成,谭翁,杜正晓,杨志林,唐杰. 2022. P-Tuning :快速调整可以与跨尺度和任务的微调相媲美.计算语言学协会第 60 届年会论文集 (第 2 卷 :短论文) 。计算语言学协会,爱尔兰都柏林,61-68. <https://doi.org/10.18653/v1/2022.acl-short.8> [25]刘晓,纪凯旋,付一成,谭翁林,杜正晓,杨志林,唐杰. 2021. P-tuning v2 :快速调整可以与跨尺度和任务的普遍微调相媲美. arXiv 预印本 arXiv:2110.07602 (2021)。

[26]刘晓涛,赖建强. 2017.使用基于动态 SVM 的方法进行日内交易量百分比预测.系统科学与复杂性杂志30 (2017), 421–433.

[27] Yinhan Liu,Myale Ott,Naman Goyal,Jingfei Du,Mandar Joshi,Danqi Chen,Omer Levy、Mike Lewis,Luke Zettlemoyer 和 Veselin Stoyanov. 2019.Roberta :一种稳健优化的 bert 预训练方法. arXiv 预印本 arXiv:1907.11692 (2019)。

[28]伊利亚·洛什奇洛夫和弗兰克·哈特. 2017.解耦权重衰减正则化. arXiv 预印本 arXiv:1711.05101 (2017)。

[29]卢嘉森,Dhruv Batra,Devi Parikh,Stefan Lee. 2019.Vilbert :针对视觉和语言任务的与任务无关的视觉语言表示的预训练.神经信息处理系统的进展 32 (2019)。

[30]卢文杰,李家政,王景阳,秦乐乐. 2021.一种用于股票价格预测的CNN-BiLSTM-AM方法.神经计算和应用 33 (2021), 4741–4753。

[31]安舒尔·米塔尔和阿尔皮特·戈埃尔. 2012. 使用 Twitter 情绪分析进行股票预测.斯坦福大学,CS229 (2011 <http://cs229.stanford.edu/proj2011/GoeMittal-StockMarketPredictionUsingTwitterSentimentAnalysis.pdf>)15 (2012) ,2352 。

[32]布莱恩·K·纳尔逊. 1998.使用自回归积分移动平均 (ARIMA) 模型进行时间序列分析.学术急诊医学 5, 7 (1998), 739–744。

[33] Thien Hai Nguyen,Kiyooki Shirai 和 Julien Velcin. 2015.用于股票走势预测的社交媒体情绪分析.专家系统与应用42, 24 (2015), 9603–9611。

[34] 开放人工智能. 2023.GPT-4 技术报告. arXiv:2303.08774 [cs.CL]

[35] Venkata Sasank Pagolu,Kamal Nayan Reddy,Ganapati Panda 和 Babita Majhi. 2016. Twitter 数据的情绪分析用于预测股市走势. 2016年信号处理、通信、电源和嵌入式系统国际会议 (SCOPES) 。 IEEE,1345–1350。

[36] Alec Radford,Jong Wook Kim,Chris Hallacy,Aditya Ramesh,Gabriel Goh、Sandhini Agarwal,Girish Sastry,Amanda Askell,Pamela Mishkin,Jack Clark等。2021.从自然语言监督学习中可迁移的视觉模型.在国际机器学习会议上. PMLR,8748–8763。

[37] Colin Raffel,Noam Shazeer,Adam Roberts,Katherine Lee,Sharan Narang、Michael Matena,Yanqi Zhou,Wei Li 和 Peter J Liu. 2020.使用统一的文本到文本转换器探索迁移学习的局限性.机器学习研究杂志21, 1 (2020), 5485–5551。

[38] Sreelekshmy Selvin,R Vinayakumar,EA Gopalakrishnan,Vijay Krishna Menon 和 KP Soman. 2017. 使用 LSTM,RNN 和 CNN 滑动窗口模型进行股价预测. 2017 年计算、通信和信息学进展国际会议(icacci). IEEE,1643–1647。

[39] Philip Treleaven,Michal Galas 和 Vidhi Lalchand. 2013.算法交易审查.交流. ACM 56, 11 (2013), 76–85。

[40] Ashish Vaswani,Noam Shazeer,Niki Parmar,Jakob Uszkoreit,Llion Jones,Aidan N Gomez、ukasz Kaiser 和 Illia Polosukhin. 2017年,你所需要的就是关注.神经信息处理系统的进展 30 (2017)。

[41]谢倩倩,韩伟光,赖艳照,彭敏,黄吉敏. 2023 年. 华尔街新手 :针对多模式股票走势预测挑战的 ChatGPT 零样本分析. arXiv 预印本 arXiv:2304.05351 (2023)。

[42]徐玉墨和谢伊·B·科恩. 2018.根据推文和历史价格预测股票走势.计算语言学协会第 56 届年会记录 (第一卷 :长论文) 。 1970 年至 1979 年。

[43]易阳,马克·克里斯托弗·赛威,艾伦·黄. 2020.Finbert :用于金融通信的预训练语言模型. arXiv 预印本 arXiv:2006.08097 (2020)。

[44] Susan 张,Stephen Roller,Naman Goyal,Mikel Artetxe,Moya Chen,Shuohui Chen、Christopher Dewan,Mona Diab,Xian Li,Xi Victoria Lin,等。2022. Opt :开放预训练的 Transformer 语言模型. arXiv 预印本 arXiv:2205.01068 (2022)。

[45]张志远,李伟,鲍瑞涵,张本惠子,吴云芳,孙旭. 2023. ASAT :时间序列中的自适应规模对抗训练.神经计算522 (2023), 11–23。

[46]赵亮,李伟,鲍瑞涵,张本惠子,孙旭,等。2021.长期、短期和突发事件 :基于图的多视图建模的交易量变动预测. arXiv 预印本 arXiv:2108.11318 (2021)。

[47]邹彦照和多里安·赫尔曼斯. 2022 年.带有 Twitter FinBERT 嵌入的多模态模型,用于预测比特币的极端价格变动. arXiv预印本 arXiv:2206.00648 (2022)。