**机器学习课程设计论文（报告）评分表**

|  |  |  |  |
| --- | --- | --- | --- |
| **课程名称** | **所在年级:** | **姓名** | **学号** |
| **《机器学习课程设计》** | **2021级** | **成颖** | **2126010314** |
| **考核项目** | | **考核分数** | |
| **论文选题的合理性（10分）** | |  | |
| **论文的工作量（10分）** | |  | |
| **论文的文字表达（25分）** | |  | |
| **论文的组织结构（25分）** | |  | |
| **论文的完整性（25分）** | |  | |
| **论文的新颖性（5分）** | |  | |
| **总成绩（总分100分）** | |  | |
| **评语** |  | | |

**教师签名：**

**日 期：2024年 6月 9日**

结合k-means方法的LSTM股票分类预测

姓名：成颖

学号： 2126010314 E-mail：2871128658qq.com

**摘 要 ：**

**目的** 选取沪深京板块为背景，选取部分行业股票进行分析。模仿股票网站的K线，做出箱线图与折线图，方便准确的分析行业的走向。对比分析不同年份的行业数据，调查行业的发展情况，结合实际情况对股票进行分析。根据涨跌幅、振幅等股票特征进行聚类分析，基于已经建立的股票分类，利用长短期记忆网络（Long Short-Term Memory, LSTM）建立分类预测模型，以便帮助投资者更好的理解并投资不同类型的股票。

**方法** 首先在东方财富网挑选合适的股票，对选取的股票数据进行爬取并进行整理。然后进行建模分析，运用k-means算法进行聚类分析，K-means是一种常用的聚类算法，用于将数据集中的观测点分为[不同的](https://so.csdn.net/so/search?q=%E4%B8%8D%E5%90%8C%E7%9A%84&spm=1001.2101.3001.7020" \t "https://blog.csdn.net/m0_56694518/article/details/_blank)群组或簇。聚类是一种无监督学习方法，其目标是发现数据中隐藏的结构，将相似的数据点划分为同一组，同时将不相似的数据点划分为不同的组。其中通过画肘部图确定合适的k值。运用长短期记忆网络，通过选取的聚类股票特征，预处理数据集，划分出测试集与训练集。在使用长短期记忆网络对训练集进行训练后，确定出适合的参数，再对测试集进行测试。

**结果** 实验选取了5个行业板块共计89只股票近年来的数据。通过手肘法，观察手肘图，确定k值为4，故最佳簇数数量为4。大致可以分为4种类型：涨跌幅比较小，振幅比较大；涨跌幅比较大，振幅比较小；涨跌幅与振幅适中；涨跌幅极大，振幅小。使用混淆矩阵对构建的分类模型的性能进行评估。在训练集中，对0类的预测精度为98.6%，对1、2类的预测精度为100%，对第3类的预测精度97%，训练集的预测正确率为99.0132%；在测试集中，0类预测精度为97%，其他类别预测精度为100%，测试集预测正确率为99.2366%。

**结论** 本文所提出的方法与模型，能够对股票数据进行处理，并且对股票进行基础的聚类分析以及预测。

**关键词**：机器学习；K-means聚类；长短期记忆网络；手肘法；股票

**0 引 言**

股票市场是金融市场的重要组成部分。随着全球化和互联网技术的发展，股票市场的规模和影响力越来越大，成为国家经济发展的重要指标。随着国民经济的发展，中国的股票市场的规模也在持续扩大，早已成为了金融投资的重要部分，掌握股票市场的变化规律，对股票能够正确的分类以及预测，无论对股市的监管者还是投资者都具有极其重要的意义。本文旨在将机器学习的相关技术应用到金融数据的预测中，尝试建立能反映股票类别的模型。对股票能够做出较为准确的分类以及分析。

在股票交易中，大多数的参与者不可能注意全部股票变化，因此对股票进行分类预测时十分重要的。股票分类预测能够帮助投资者识别出具有上升趋势、下降趋势或盘整趋势的股票，从而指导其买入、卖出或持有股票的决策。并且能够进行筛选投资机会，帮助投资者在众多股票中找出那些具备上升趋势或者反转趋势的股票，从而找到潜在的投资机会。通过对股票进行分类预测，投资者可以更加精准地设定止损和止盈位，有效地控制投资风险。股票分类预测在股市中具有重要作用，它不仅能够指导投资决策、筛选投资机会、控制风险，还能提高预测准确率并结合基本面分析为投资者提供全面的市场视角。

**1 数据与方法**

* 1. **数据来源**

在东方财富网站上，沪深京板块中，选取半导体、玻璃纤维、电子元件、纺织服装、酿酒行业这5个行业。在这5个行业里各挑选了15-20只股票，收集了股票代码、股票简称、日期、开盘价、收盘价、最高价、最低价、涨跌幅(%)、涨跌额、成交量、成交额、振幅(%)、换手率(%)、行业、地区等信息，作为进行分析的数据。

* 1. **数据处理方法**

1.2.1 手肘法

手肘法是一种确定k值的方法，其核心指标是误差平方根（SSE）

 （1）

式中：*Ci*——第*i*个簇；

*p*是*Ci*中的样本点*；*

*Mi*是*Ci*中所有样本的均值；

SSE是所有样本中的聚类误差，代表了聚类效果的好坏。

随着聚类数k的增大，样本划分会更加精细，每个簇的聚合程度会逐渐提高，那么误差平方和SSE自然会逐渐变小。当k小于真实聚类数时，由于k的增大会大幅增加每个簇的聚合程度，故SSE的下降幅度会很大，而当k到达真实聚类数时，再增加k所得到的聚合程度回报会迅速变小，所以SSE的下降幅度会骤减，然后随着k值的继续增大而趋于平缓，也就是说SSE和k的关系图是一个手肘的形状，而这个肘部对应的k值就是数据的真实聚类数。

1.2.2 K-means聚类方法

K-means聚类方法是划分聚类算法的典型代表之一，其核心思想是基于数据到中心点距离最小化原则无监督地对数据样本进行空间划分。该方法的工作原理是将n个数据对象按照k值进行划分，通过式（1）计算每个聚类平均值作为聚类中心点值p，通过式（2）计算每个数据对象与每个聚类中心点的距离d，依据距离最小值将数据对象划分到相应的聚类中。当所有数据对象聚类结束后，再通过式（1）计算每个新聚类的中心点值P。当P=p时，中心点值收敛，可终止聚类，否则继续聚类。

 (2)

 (3)

式中，xi表示聚类中的数据对象；xi∈{x1,x2,⋯,xn}；dist()表示2范数下的欧几里得距离。

1.2.3 长短期记忆网络 (LSTM)

长短期记忆网络模型设计了三个门，分别是输入门、遗忘门和输出门。通过这三个“门”的控制，长短期记忆网络模型可以很好地保护和控制网络中的细胞状态。长短期记忆网络模型可以通过门控机制，有效地捕捉到序列数据中的长期依赖关系，从而提高模型的准确性，还可以帮助网络有效地控制梯度的流动，从而减轻梯度消失问题的影响。图1展示了长短期记忆网络模型的内部结构，通过设计的独特门控机制，可以灵活地决定哪些信息需要被遗忘、哪些信息需要被保留。

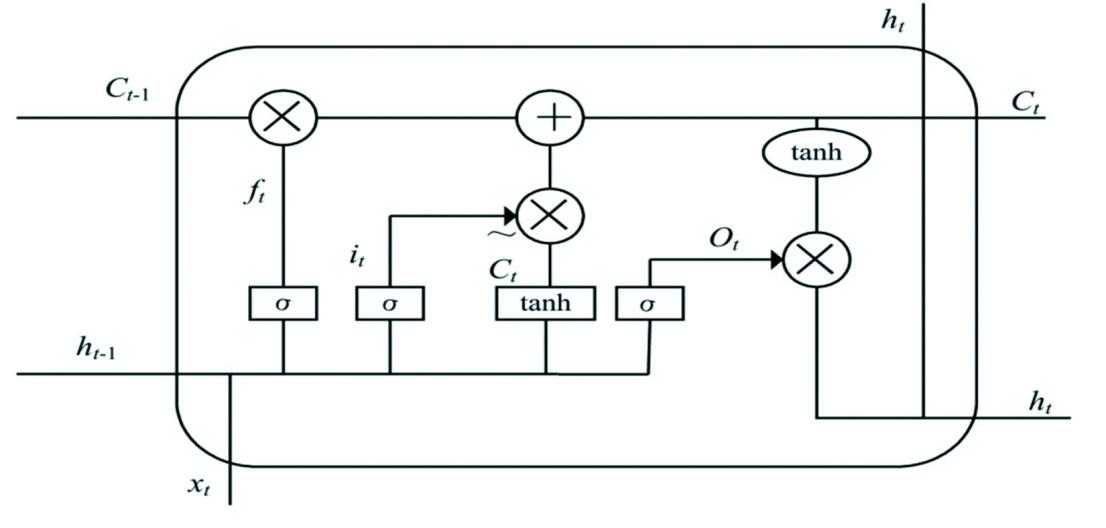


图1 长短期记忆细胞结果图

输入门用来决定细胞中应该加入哪些新状态。此项设计两部分工作：第一部分是通过使用输入门的sigmoid层决定我们要去更新哪些位置上的值，第二部分是通过tanh层生成一个候选值组成可加到细胞状态上的向量。输入门状态更新公式

 (4)

 (5)

式(4)和式(5)中，σ(·)表示sigmoid函数，[ht-1,xt]表示上一时刻长短期记忆网络模型的输出值ht-1与当前时刻长短期记忆网络模型的输入值xt组成的向量，it表示输入门的输出，Wi表示输入门的权重矩阵，bi表示输入门的偏置。

表示当前时刻的单元状态，Wc表示输入单元的权重，bc表示输入单元的偏置。

遗忘门是sigmoid层综合前一时刻的隐藏状态和当前时刻的输入数据做出是否让这些信息通过的决定。其中“0”表示完全忘记，“1”表示完全保留。 遗忘门状态更新公式

 (6)

式(4)中，ft为遗忘门的输出，Wf表示遗忘门的权重矩阵，bf为遗忘门的偏置。

输出门是基于当前的细胞状态，首先通过一个tanh层来把细胞状态映射到区间(-1, 1)内，接着利用输出门的sigmoid层控制输出细胞状态中的信息。输出门状态更新公式

 (7)

式(5)中，ot表示当前输出门的输出，Wo表示输出门的权重矩阵，bo表示输出门的偏置。

更新细胞Ct细胞状态

 (8)

输出

 (9)

**2 结果与分析**

2.1 **2022年以及2023年酿酒行业k线走势**

选取2022年以及2023年酿酒行业的数据作为分析对象，以时间为横坐标，每股单价（每股单价=总成交金额÷总成交股数×100%）作为作为纵坐标，模仿股票网站的k线，做出箱线图（图2与图3）。

从图2中我们可以看出，酿酒行业股票在2022年呈现出明显的波动性和阶段性特征。年初至年中以下跌为主，而下半年则开始复苏并呈现上涨态势。年初至年中，市场整体的调整以及行业内部的竞争压力有关。特别是头部白酒企业如五粮液、泸州老窖、贵州茅台等，其股价在这段时间内也有所回落。下半年中尤其是11月以来，白酒板块涨幅位居A股全行业第3，绝大多数个股上涨超30%。这可能与年底消费旺季的来临以及市场对经济复苏的预期有关。

从图3中我们可以看出，酿酒行业的股价在这2023年中经历了较大的波动。从年初到年末，股价在52500至70000的范围内波动。最高点达到了70000，反映了市场对该行业的乐观预期或某些利好消息的影响。而最低点降至52500，这可能是由于行业整体压力、市场调整或负面消息导致的。其次，从K线的形状和趋势来看，股价在某些时段呈现出上涨趋势，而在其他时段则呈现下跌趋势。这显示了投资者对于酿酒行业的信心和热情有所起伏。

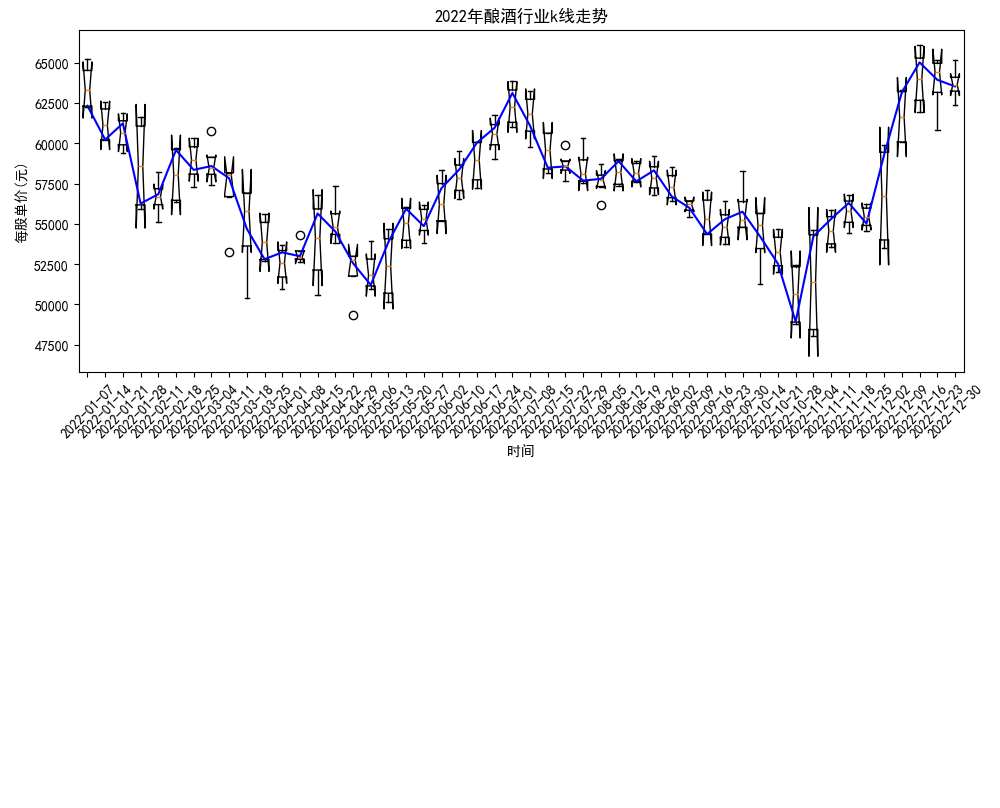


图2 2022年酿酒行业k线走势

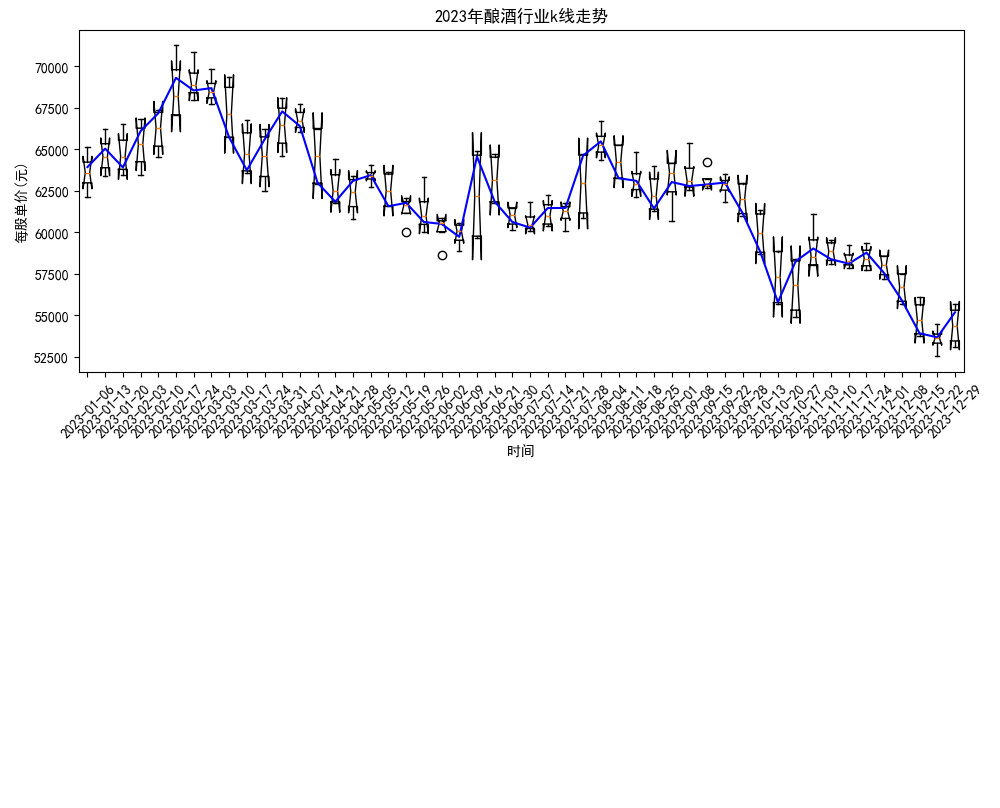


图3 2023年酿酒行业k线走势

**2.2 2022年与2023年各行业折线走势**

选取2022年以及2023年五个行业的数据作为分析对象，以时间为横坐标，每股单价（每股单价=总成交金额÷总成交股数×100%）作为作为纵坐标，做出折线走势图。

图4实线表示2022年各行业的数据，双划线为2023年。从折线图的走势来看，各行业的股价在两年间有不同的变化。

酿酒行业的股价，在2022年的股价波动范围较大，显示出该行业在2022年内的股价有一定的不稳定性，而到了2023年，股价整体相较于2022年是呈现上升趋势的，波动幅度相较2022年较小，但在某个时间段内出现了显著的上涨。

半导体行业的股价在股市里每股单价较低，由于图中的纵轴的刻度较大，图中显示的变化趋势并不太明显。但是根据实际数据，在2022年呈现上升趋势，但在年中时期有所回调。而到了2023年，股价整体上扬，但波动相对较为平缓。

电子元件行业的股价在两年间都呈现了一定的波动，但整体来看其波动幅度相对较小，显示出该行业在全年内可能保持了一定的稳定性。但整体来看，2023年的股价水平要高于2022年。

玻璃纤维行业的股价在2022年呈现出较大的波动，尤其在某个时间段内出现了明显的上涨。而到了2023年，股价波动情况较为稳定，但从整体上看，股票的价格呈现出下跌趋势。

最后，纺织服装行业的股价在两年间都呈现出较为稳定的态势，但2023年的股价整体略高于2022年。

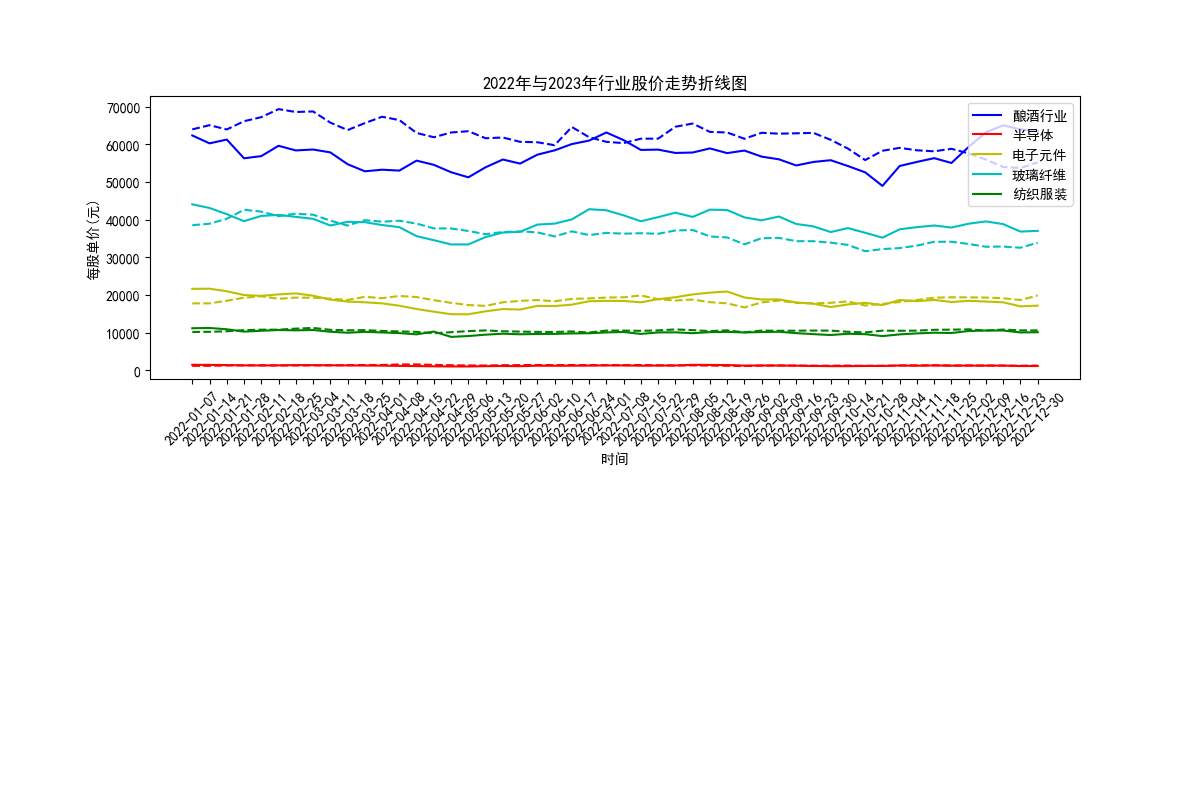


图4 2022年与2023年各行业折线走势

**2.3 2022年与2023年年初各行业股价直方图**

选取2022年以及2023年五个行业的数据作为分析对象，挑选出五个行业年初的数据（表1）进行可视化分析。

从表1以及图5中，我们可以看到2022年和2023年大部分行业的单价是有所下降的，这个跌幅比较大的是电子元件行业和玻璃玻纤行业。而酿酒行业是有所小幅度增长的。

横向对比不同行业，我们可以看到单价最高的是酿酒行业，这个行业的单价远远高于其他行业。其次是玻璃玻纤行业、电子元件和玻璃玻纤行业。而最低的则是纺织服装行业与半导体行业。

对于同一行业的2018年和2022年数据的对比，我们可以看到，除了贸易行业和旅游酒店行业外，其他行业单价都有所增长。其中增幅最大的是酿酒行业、玻璃玻纤行业和化学制药行业。

**表1 2022年与2023年各行业年初数据**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 年份 | 酿酒行业 | 半导体 | 电子元件 | 玻璃纤维 | 纺织服装 |
| 2022 | 62324.69 | 1430.09 | 21596.40 | 44040.11 | 11134.53 |
| 2023 | 63939.67 | 1150.32 | 17747.67 | 38488.96 | 10156.48 |

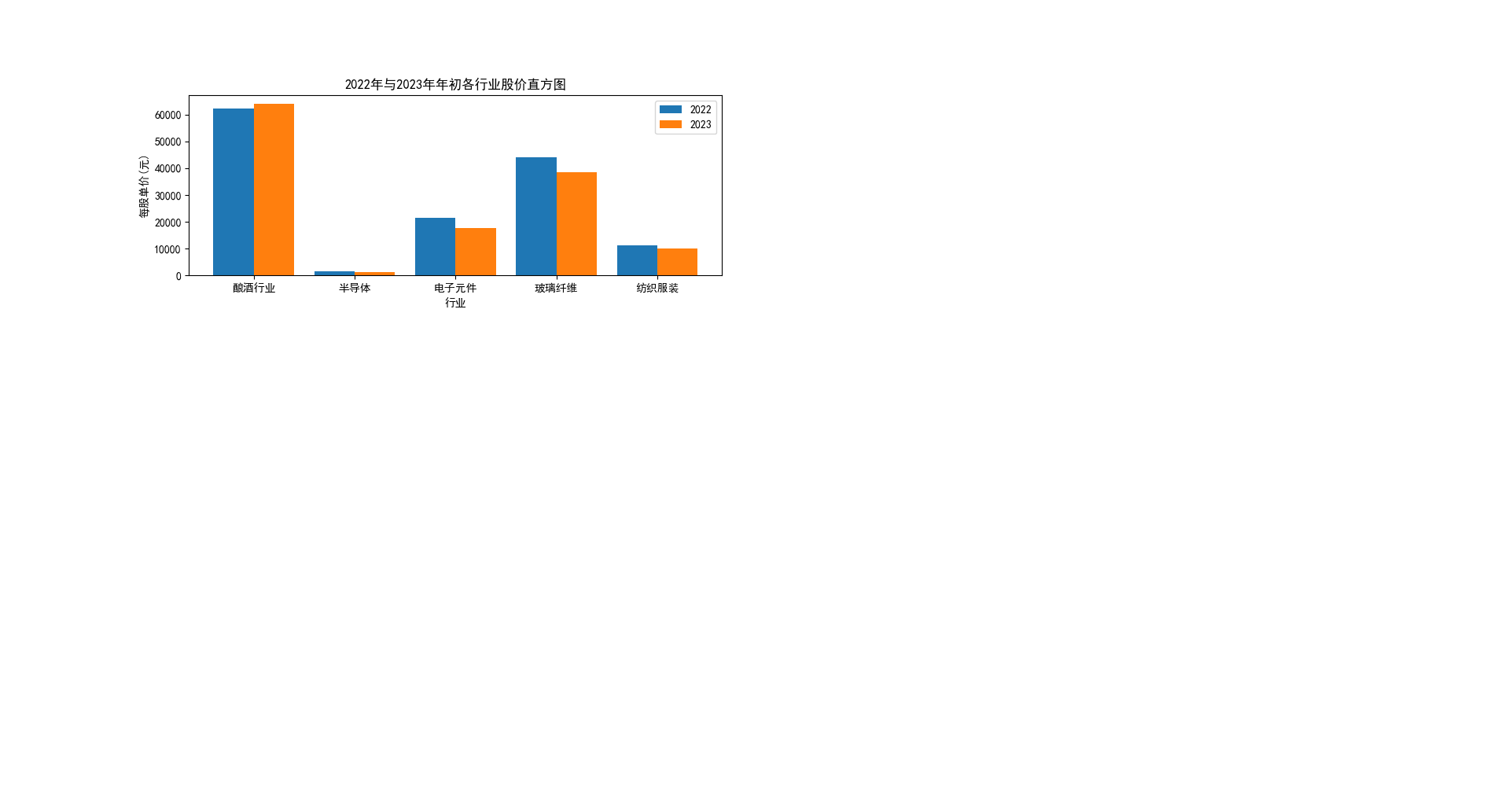
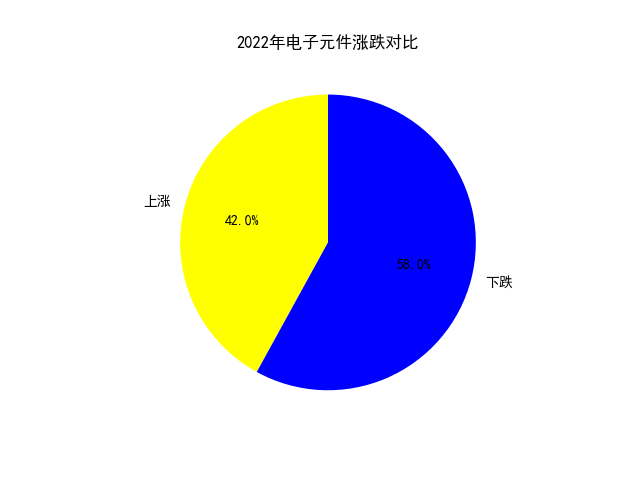
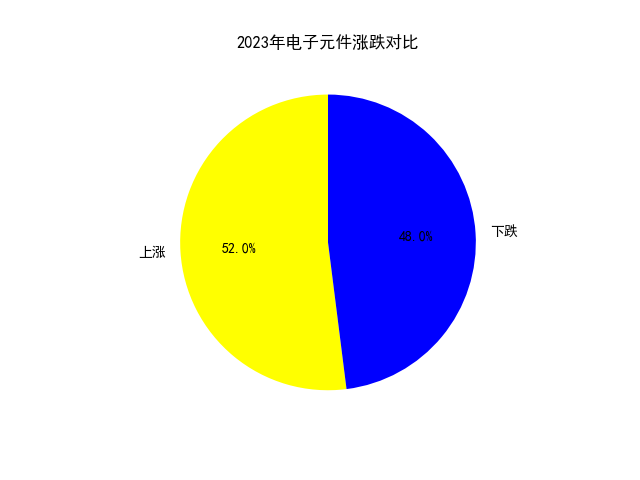


图5 2022年与2023年年初各行业股价直方图

**2.4 电子元件涨跌幅饼状图**

选取2022年以及2023年电子元件行业涨跌幅作为特征指标进行数据分析。首先对电子行业2022年和2023年的涨跌幅数据进行处理，分别统计出涨跌幅为正值和负值，然后与总量作商计算出占比。从图6中可以看出，2022年电子元件行业的涨跌幅相差比较大，下跌幅度略高于上涨幅度，整个行业整体呈现出下跌的趋势。而2023年，电子元件行业整体上涨，且上涨幅度高于下跌幅度，表明电子元件行业发展较好，正在逐渐获得市场的认可和支持。

(a)2022年电子元件涨跌对比图 (b)2023年电子元件涨跌对比图

图6 电子元件涨跌饼状图

**2.5 k-means聚类分析**

2.5.1 手肘图

首先对选取的五个股票行业共计89只股票进行整理，然后选择“涨跌幅”与“振幅”作为进行聚类分析的特征，对整合的数据进行预处理得到表2（部分数据）。

**表2 2020-2023年各股票预处理数据（部分）**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 行业名称 | 股票名称 | 2020 | | 2021 | | 2022 | | 2023 | |
| 酿酒行业 | 泸州老窖 | 117.29 | 1027.35 | 28.91 | 1261.84 | -2.06 | 946.81 | -15.88 | 722.26 |
| 半导体 | 通富微电 | 62.37 | 1342.04 | -16.98 | 901.36 | -6.09 | 954.51 | 44.98 | 1024.69 |
| 电子元件 | 顺络电子 | 18.73 | 975.77 | 54.96 | 1012.28 | -28.8 | 987.26 | 7.97 | 739.78 |
| 玻璃纤维 | 南玻A | 57.39 | 1090.46 | 47.34 | 1177.36 | -29.35 | 926.39 | -14.93 | 384 |
| 纺织服装 | 森马服饰 | 21.44 | 1106.29 | -16.27 | 988.65 | -30.18 | 655.27 | 17.43 | 575.48 |

由于数据之间数值相差较大，所以需要进行标准化处理。对处理后的数据应用手肘法，做出手肘可视图。肘部法则的基本思想是：随着k值的增加，簇内平方和（WCSS）会逐渐减小，因为更多的簇可以更好地拟合数据。然而，当k值增加到某个点时，WCSS的减小速度会突然放缓，形成一个类似“肘部”的形状。这个点通常被认为是最佳的k值，因为它代表了增加簇的边际效益开始降低的位置。从图7中，我们能够看到，随着簇数量的增加，y轴上的度量值先迅速下降，然后下降速度逐渐放缓，形成一个类似“肘部”的曲线形状。根据图中的信息，当k值为4时，下降的速度开始减缓，所以我们可以推断“肘部”大致在簇数量为4，因此，可以初步判断，对于该数据集，选择4个簇可能是一个合适的聚类方案。

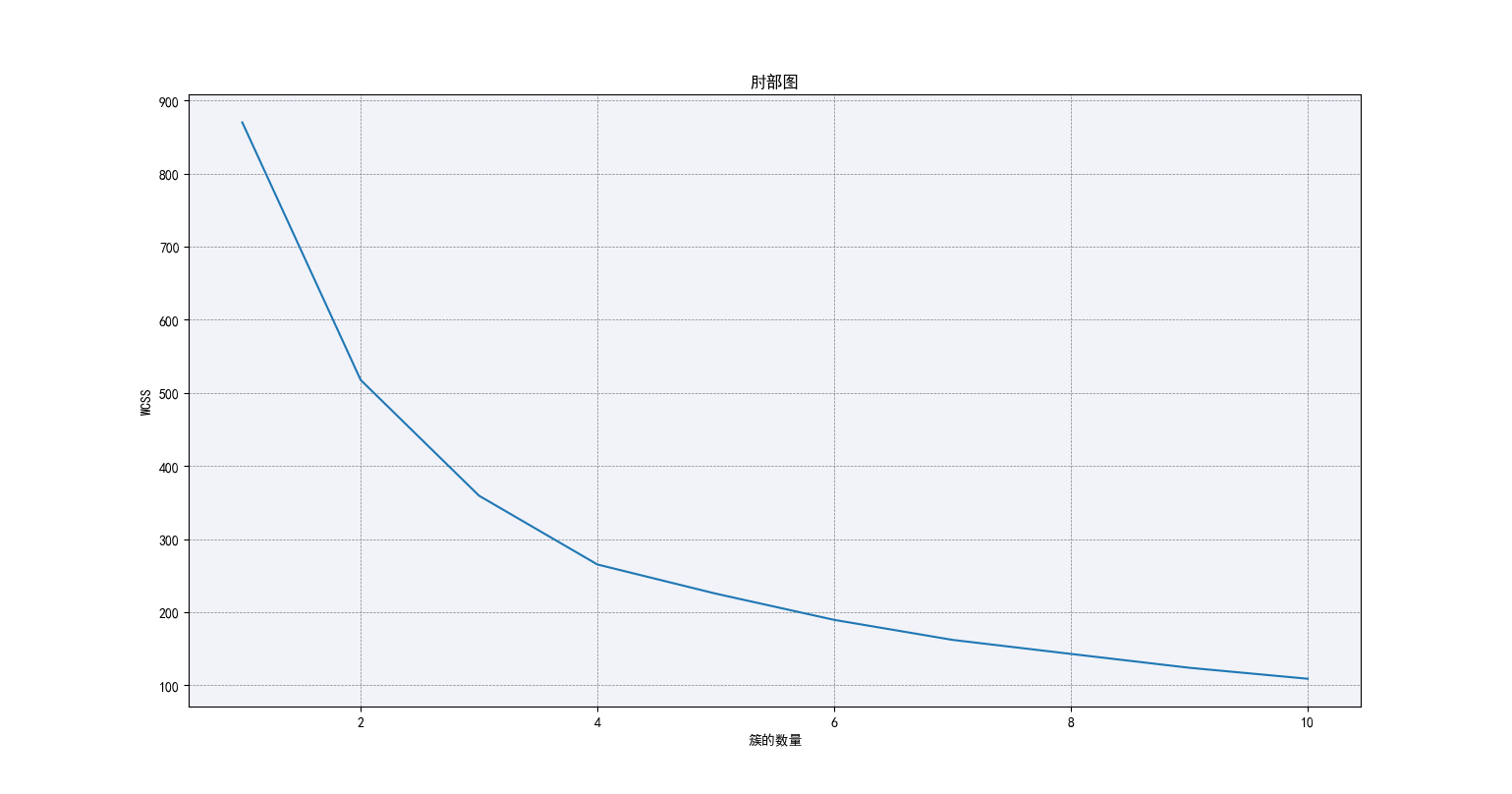


图7 肘部图

2.5.2聚类分析

根据肘部图确定的k值，使用k-means方法对预处理的数据进行聚类分析，得到了聚类后的股票信息表3。并且可视化了聚类分布图。根据聚类数据的统计表4-表7和结果图8，我们可以看出，

簇0:这个聚类包含104个股票，涨跌幅的均值达到了71.99%，说明这个聚类的股票在观察期间平均涨幅较大。振幅的均值非常高，达到了1246.62%，表明这些股票在交易日内价格波动非常剧烈。涨跌幅的标准差为32.95%，意味着这些股票之间的涨跌幅差异较大，可能包含了多种表现不同的股票。最大涨跌幅达到了162.47%，说明这个聚类中有些股票在观察期间表现非常强劲。

簇1:这个聚类包含156个股票，涨跌幅的均值只有4.10%，远低于簇 0，表明这些股票在观察期间的平均涨幅较小。振幅的均值相对较低，为645.74%，但仍然属于较高的水平。这个聚类包含了从大幅下跌（-57.27%）到较大上涨（109.71%）的股票，但大部分股票的涨跌幅都在较低的水平。振幅的标准差较大，说明这些股票在交易日内价格波动较大，但涨跌幅的波动相对较小。

簇2:这个聚类只包含11个股票，但涨跌幅的均值高达235.81%，是所有聚类中最高的，表明这些股票在观察期间表现非常强劲。振幅的均值虽然不高，但考虑到样本量很小，这可能不是一个可靠的指标。这个聚类中股票的涨跌幅标准差很大，表明这些股票的表现差异极大。最小涨跌幅也达到了127.06%，说明这个聚类中的股票在观察期间都取得了显著的上涨。

簇3:这个聚类包含164个股票，涨跌幅的均值为-1.65%，是所有聚类中最低的，表明这些股票在观察期间平均表现较差。然而，这些股票的振幅均值高达995.84%，是所有聚类中最高的，表明这些股票在交易日内价格波动非常剧烈。涨跌幅的标准差为25.62%，相对较小，说明这个聚类中大部分股票的涨跌幅都集中在均值附近。尽管有少数股票表现强劲（如最大涨跌幅达到了54.36%），但整体来说，这个聚类中的股票表现较差。

总结：簇0中的股票在观察期间平均涨幅较大，但价格波动也非常剧烈。说明这些股票可能是质量较高的热门股；簇1中的股票平均涨幅较小，但价格波动仍然较大。说明这些股票可能是行情较好的热门股，但由于市场波动性较大，风险也较高；簇2中的股票表现非常强劲，但样本量较小，需要谨慎解读。这可能反映了一些极端行情的股票，投资者需要谨慎参考；簇3中的股票平均表现较差，但价格波动是所有聚类中最剧烈的；这些股票可能是相对保守的股票。

**表3 簇类样本量**

|  |  |
| --- | --- |
| 簇 | 数量 |
| 0 | 104 |
| 1 | 156 |
| 2 | 11 |
| 3 | 164 |

**表4 簇0数据信息**

|  |  |  |
| --- | --- | --- |
|  | 涨跌幅(%) | 振幅(%) |
| count | 104 | 104 |
| mean | 71.98 | 1246.61 |
| std | 32.95 | 181.150 |
| min | 10.45 | 825.89 |
| max | 162.47 | 1798.92 |

**表5 簇1数据信息**

|  |  |  |
| --- | --- | --- |
|  | 涨跌幅(%) | 振幅(%) |
| count | 156 | 156 |
| mean | 4.10 | 645.74 |
| std | 27.65 | 143.81 |
| min | -57.27 | 26.43 |
| max | 109.71 | 830.22 |

**表6 簇2数据信息**

|  |  |  |
| --- | --- | --- |
|  | 涨跌幅(%) | 振幅(%) |
| count | 11 | 11 |
| mean | 235.81 | 530.96 |
| std | 98.89 | 349.95 |
| min | 127.06 | 27.54 |
| max | 417.04 | 959.75 |

**表7 簇3数据信息**

|  |  |  |
| --- | --- | --- |
|  | 涨跌幅(%) | 振幅(%) |
| count | 164 | 164 |
| mean | -1.64 | 995.83 |
| std | 25.62 | 118.67 |
| min | -80.55 | 830.65 |
| max | 54.36 | 1373.50 |

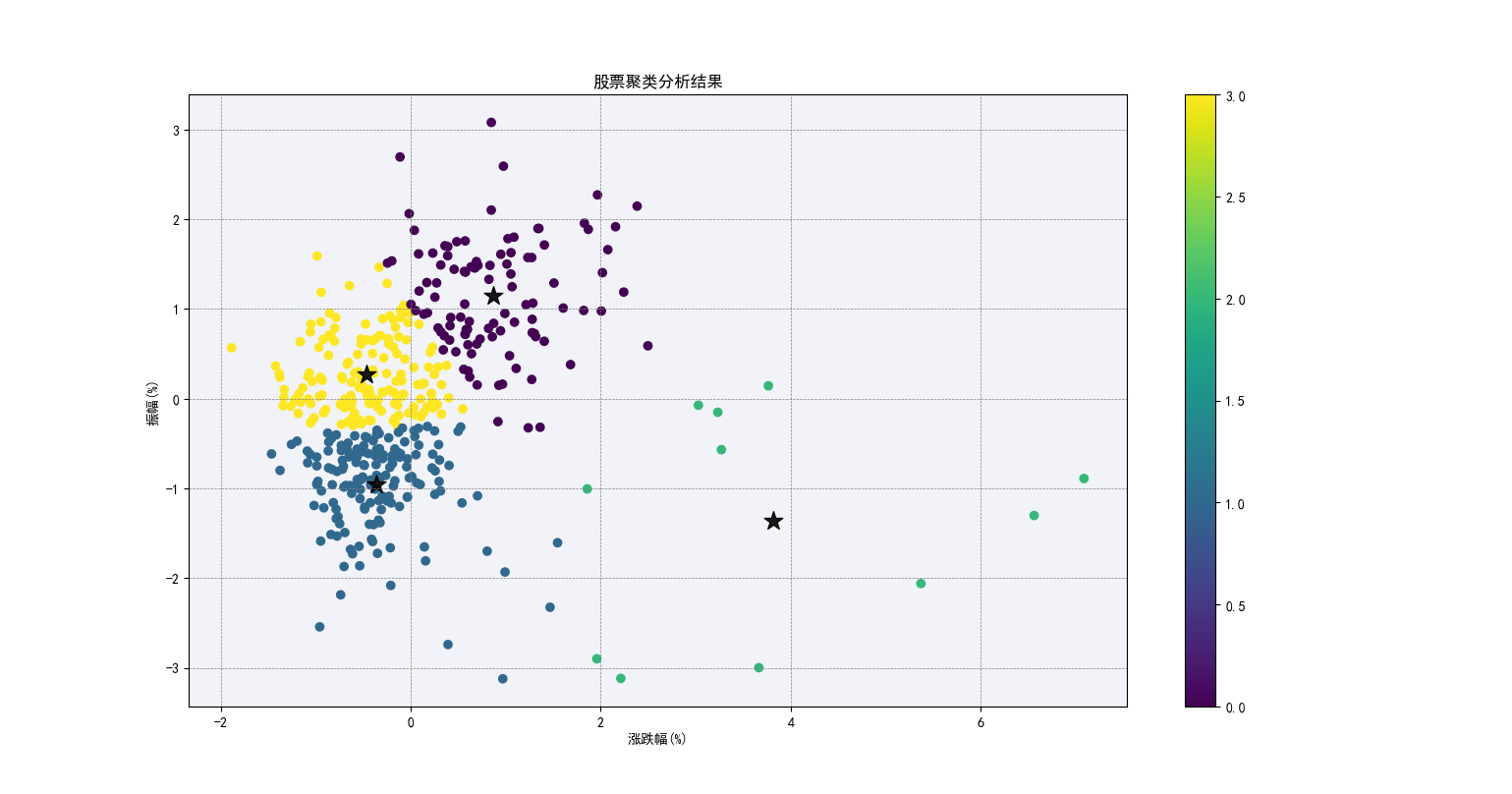


图8 聚类分析结果图

**2.6 长短期记忆网络分类预测**

2.6.1 数据收集与整理

在进行k-means聚类分析时，我们将89只股票共计435条股票信息进行了聚类处理。将处理的数据收集并整理出来，作为长短期记忆网络模型的训练集与测试集，部分数据展示如表8。由于各指标间的水平相差很大，如果直接用原始指标值进行分析，就会突出数值较高的指标在综合分析中的作用，相对削弱数值水平较低指标的作用。因此，为了保证结果的可靠性，需要对原始指标数据进行归一化处理。

**表8 股票聚类信息表（部分展示）**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 股票简称 | 日期 | 聚类结果 | 涨跌幅(%) | 振幅(%) |
| 通富微电 | 2020 | 0 | 62.37 | 1342.04 |
| 通富微电 | 2021 | 3 | -16.98 | 901.36 |
| 通富微电 | 2022 | 3 | -6.09 | 954.51 |
| 中晶科技 | 2020 | 2 | 146.66 | 27.54 |
| 北方华创 | 2020 | 0 | 94.66 | 1368.43 |
| 中晶科技 | 2023 | 1 | -23.84 | 780.77 |
| 易天股份 | 2023 | 0 | 102.02 | 1408.48 |
| 派瑞股份 | 2020 | 2 | 191.95 | 898.02 |

2.6.2 分类预测

在长短期记忆网络模型训练中，选取70%的数据作为训练集、剩下的30%为测试集。在经过1000轮迭代，每轮迭代2次，共计2000次迭代后，训练完成。从长短期记忆网络训练进度图10中，我们发现，在经过100次训练后，准确率陡然上升，在经过大约200次迭代后，模型的准确率近乎达到峰值，同时损失也减小到了一个很低的程度，训练1000次后，损失几乎降到了最低。后续的训练中准确度以及损失并没有什么太大的变化，迭代过程有些过多了，可以适当进行减少。

从训练集与测试集的真实值与预测值对比图中，我们发现预测值与真实值在图像上近乎全部重合，除了极少数点有差距。在训练集中，有301个样本被正确预测，3个样本被错误预测，准确率达到99.0132%;在测试集中，有130个样本被成功预测，1个样本被错误预测，准确率达到了99.2366%。

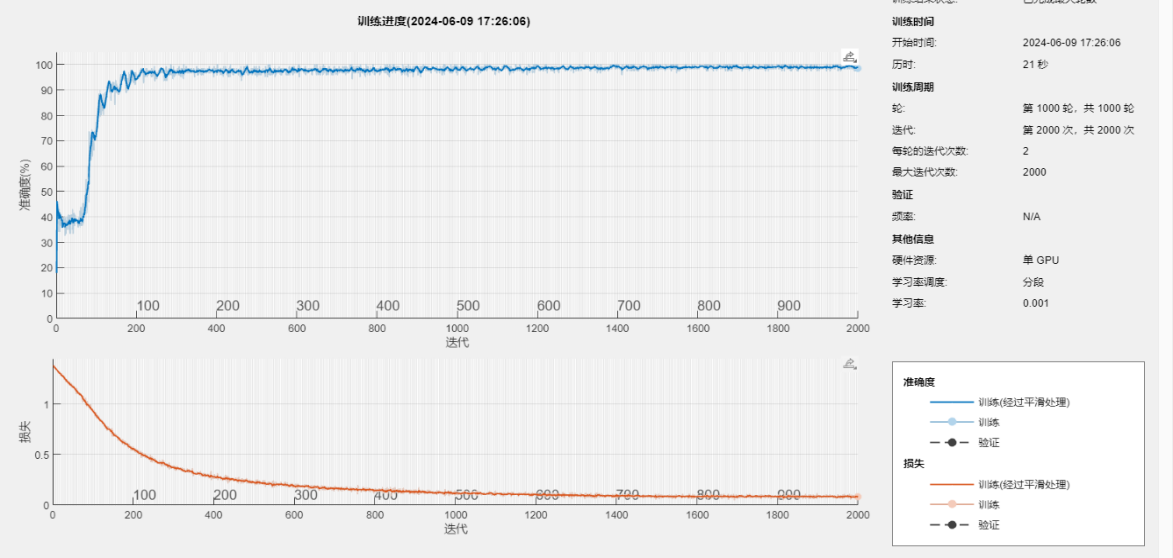
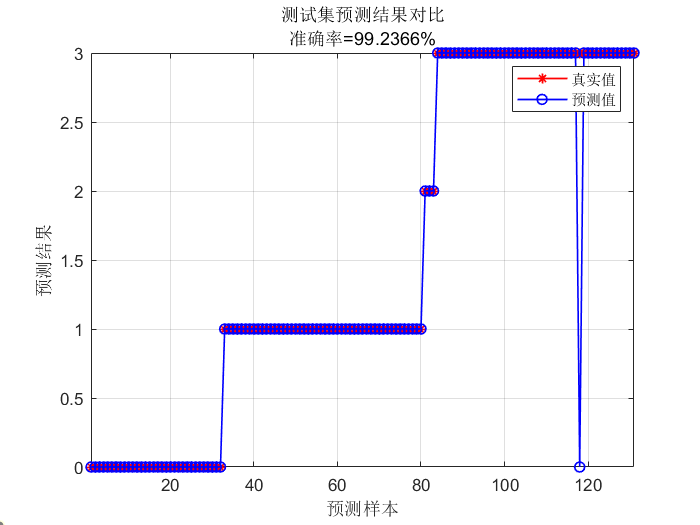
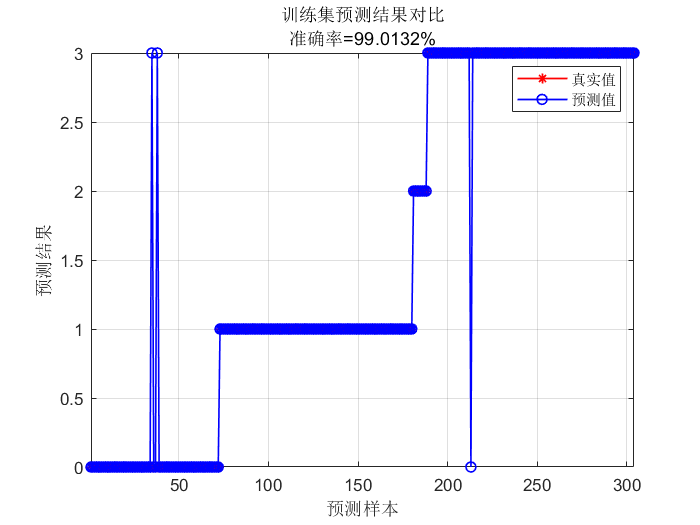


图9 LSTM训练进度图

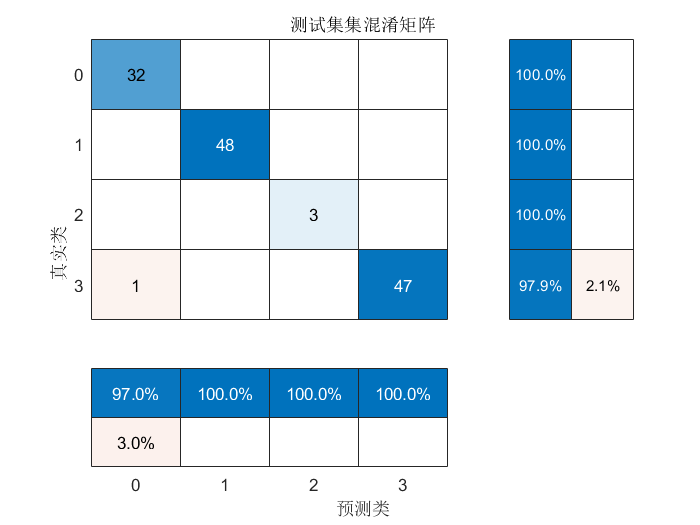
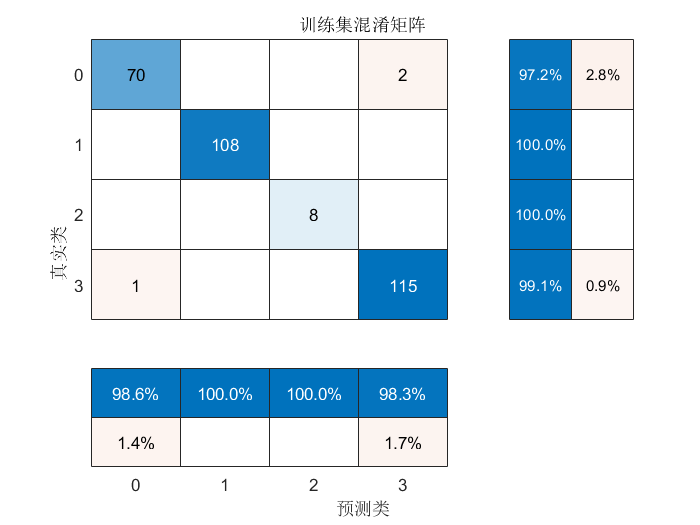


(a)训练集可视化图 (b)测试集可视化图

图10 模型预测整体结果图

2.6.2 模型评估

对于一个模型的好坏，除了通过图像可视化的对比分析，更重要的是进行数据量化评价。我们所研究的是对股票分类的预测，在得到模型训练的结果后，通常我们要对其模型训练的结果进行评估。针对监督学习中的分类问题，我们采取通过混淆矩阵来判断模型的好坏。图11中的数据，右侧行向量表示对每个类别的召回率，底部列向量表示对该类的预测的正确率。从图中我们能看出预测值与真实值之间的偏离程度较小，说明模型的预测能力和预测精度较高。



(a)训练集 (b)测试集

图11 混淆矩阵图

#### 3 结 论

**3.1 小结**

本次课程设计主要是针对股票进行数据分析和预测。实现了对网站数据的爬取，从东方财富网站获取了我们想要的数据，并且通过编程对数据进行了预处理。对股票数据做出了不同的可视化分析。并且提出了一种利用k-means聚类针对涨跌幅和振幅这两个指标对股票进行分类的方法。根据k-means聚类的聚类记过，利用长短期记忆网络对股票的类别进行预测，根据评价指标，显示该模型对股票的分类情况有着较好的预测精度。

**3.2 收获**

通过本次对沪深京板块部分行业股票的分析，我深刻体会到了数据科学在金融市场分析中的强大应用。懂得了数据的获取和预处理的重要性，数据的准确性和完整性是后续分析的基础；学会了聚类分析在股票分析中的应用，对股票进行聚类分析不仅有助于投资者更好地理解不同股票之间的相似性和差异性，还能为投资决策提供有价值的参考；还学会了长短期记忆网络在分类预测中的应用，对股票进行分类预测能够帮助投资者规避风险，做出获取最大收益的决策；明白了模型评估与优化的必要性，对于已经建立的模型，必须有评估指标，来评估模型的性能，通过评估指标的反馈，能够帮助我们清楚地了解构建模型的预测能力，能够帮助我们对模型进行优化和改进；认识到跨学科知识融合的必要性，金融市场的复杂性要求我们不仅具备数据科学的知识和技能，还需要对金融市场有深入的了解和认识。只有将两者结合起来，我们才能更好地利用数据科学来解决金融市场的实际问题。

**3.3 不足与展望**

在本次课程设计中，认为存在的不足之处有：

1、选取的数据还不够全面，呈现的特征不是很具体，如：半导体行业与酿酒行业数据差距太大，导致半导体行业在折线图中表现的走向趋势不够明显；

2、选择的聚类特征少，呈现的聚类效果不太明显；

3、由于选取的指标较少，个别类别的样本量较少，进行分类预测可能存在一定的问题；

4、pycharm中有关LSTM的软件包难以下载，LSTM实现转到matlab。

对未来的展望：

1、增强个人的编程能力，并且对编程软件能够熟练的运用；

2、掌握更多的算法，了解其逻辑，能够进行方法融合使用；

致 谢 此次课程设计地完成，在b站上浏览了大量视频进行学习，在此衷心地感谢b站上发布教学视频的up主们！

参考文献(References)

1. 李浩,赵青,崔辰州,等.基于CNN与LSTM复合深度模型的恒星光谱分类算法[J].光谱学与光谱分析,2024,44(06):1668-1675.
2. 林美,梁艳洁,陆彬.基于K-means与GRNN的高原山区高速公路短时交通流预测[J].交通节能与环保,2024,20(02):67-73.马培梁.
3. 基于K-means算法的数据挖掘与客户细分研究[J].市场研究,2019,(11):66-67.DOI:10.13999/j.cnki.scyj.2019.11.024.刘盼.
4. 考虑动静态影响因素的电芯分选研究[J].中国集体经济,2021,(31):160-162.