

知

首发于  
量化交易&宽客

写文章

登录

# 【量化课堂】教你用机器学习预测涨跌



高斯蒙 · 10 个月前

JoinQuant聚宽量化平台，免费提供数据获取、研究环境、策略回测、实盘连接、发送交易信号等功能，欢迎体验 >>> [JoinQuant聚宽](#)

**导语：**在我们的数学课堂中，我们给大家简单介绍了几种机器学习方法的算法原理（SVM，朴素贝叶斯，随机森林等等），在每篇文章的最后，我们都放了一个非常小的例子来帮助大家使用这些算法。这一篇就给大家展示一个更贴近实际的例子，来帮助大家更好的使用这些算法。

本文中，为了降低提取的特征中的噪声，我们将常见技术分析指标进行离散化，作为特征值，如RSI，MACD，等等。然后应用SVM，朴素贝叶斯和随机森林算法来预测下一个交易日的涨跌情况。

为了和实际应用场景更接近，我们每天都会向训练数据中添加今天的数据，因为训练数据集发生了改变，我们每天都会训练一个新的模型。下面我们正式开工。

本文由JoinQuant量化课堂推出。难度标签为进阶上，理解深度标签：level-0

作者：yongpeng.r

编辑：宏观经济算命师

更好的阅读体验以及查看完整代码请点连接君：[👉 ^ 🙌](#) 连接在此,谁敢点我!

**阅读前需要了解：**

文中涉及到SVM，随机森林，朴素贝叶斯算法，可以参照量化课堂中的[支持向量机入门](#)，[随机森林入门](#)，[朴素贝叶斯入门](#)。

**训练预测代码介绍：**

本文是之前的支持向量机，随机森林，朴素贝叶斯等机器学习算法的应用篇，具体的代码可以参照文中后面的研究模块，下面先将一些重要的代码拿出来说说，以方便大家理解代码的逻辑。

**特征计算**

我们先通过get\_price方法得到该支股票的前35个交易日的开高低收和交易量数据，然后利用talib

包提供的特征计算API计算相应的特征。在这次实验中，我们选取了SMA，WMA等十多个技术指标作为训练模型的特征。如计算SMA指标的代码如下：

```
trading_days = get_all_trade_days()
start_day = trading_days[index - 35]
end_day = trading_days[index]
stock_data = get_price(test_stock, start_date=start_day, end_date=end_day, \
                        frequency='daily', fields=['close', 'high', 'low', 'volume'])
close_prices = stock_data['close'].values
sma_data = talib.SMA(close_prices)[-1]
```

## 分类标签确定

如果下一个交易日的收盘价比当前交易日的收盘价高，label的值为True，否则为False。相关代码如下：

```
start_day = trading_days[index]
end_day = trading_days[index + 1]
stock_data = get_price(test_stock, start_date=start_day, end_date=end_day, \
                        frequency='daily', fields=['close', 'high', 'low', 'volume'])
close_prices = stock_data['close'].values
label = False
if close_prices[-1] > close_prices[-2]:
    label = True
```

## 特征离散化

通过talib库计算出来的特征都是连续值，由于股市中的噪声非常多，如果直接将连续值特征放到机器学习算法中训练，得到的结果很可能过拟合的（回测结果好，实盘结果差）。为了避免过拟合，本文中结合各个特征的含义，将连续值特征离散化为二值特征。比如对于SMA特征，离散化的方法就是和当日收盘价相比较，若SMA小于当前收盘价，离散之后的特征值是1，若SMA大于等于收盘价，离散之后的特征值是-1。SMA离散化的代码如下，x\_all是存放所有特征值的二维数组，x\_all中的第1列存放的就是SMA特征数据。

```
# SMA
if x_all[index][0] < x_all[index][-1]:
    x_all[index][0] = 1
else:
    x_all[index][0] = -1
```

## 应用机器学习算法

在准备好数据之后，就需要利用机器学习算法来训练模型并且利用训练好的模型预测下一个交易

日的涨跌情况了。下面是通过SVM算法预测的代码：

```
clf = svm.SVC()  
clf.fit(x_train, y_train)  
prediction = clf.predict(x_test)
```

## 预测结果

随着时间增加，训练数据随之增多，每天对下一天进行，整体成功率变化如下：

最后成功率大概是58%。

本文由JoinQuant量化课堂推出，版权归JoinQuant所有，商业转载请联系我们获得授权，非商业转载请注明出处。



更好的阅读体验以及查看完整代码请点连接君：   连接在此,谁敢点我!

**JoinQuant聚宽**，免费提供金融数据获取,分析,投资策略编写,回测,实盘连接,发送交易信号等功能服务。欢迎体验：)

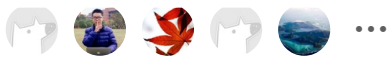
机器学习

股票

编程

☆ 收藏    分享    举报

 249



## 20 条评论

写下你的评论



**豪哥**

我直接用软件自带的M A C D判断方向的成功率都超过60%了，我还去算什么。虽然古话说，艺多不压身，但是在交易市场，方法越简单越好。。

10 个月前

9 赞



**真糊涂**

怎么收藏啊。这是用什么语言编程的。,

10 个月前



**Wells**

58%不是和抓阄差不多，搞不好直接使用趋势交易方法的成功率都会高于这个.....

10 个月前

4 赞



**范闲**

对于聚宽来说，面向的人员应该是对量化交易有兴趣的程序员或者掌握基本编程技能的人，但是对市场的理解是需要大量时间去实践或者培养，所以我觉得量化交易平台应该是个半衰斯长而收益高的事情。

10 个月前

1 赞



**Dirac**

预测的效率太低，可能还不及有经验的人

10 个月前



**拉面max**

mark

10 个月前



**你孩子他爹**



10 个月前



龙潇

也太粗糙了吧,

10 个月前

1 赞



巴聪

58% is good enough. In terms of daily trading, if you can achieve 58% in a long term, you will be millionaire in a few years.

10 个月前

3 赞



孙逸豪

.....只能说用机器学习在股票大概是不合适的.....

10 个月前

下一页

## 文章被以下专栏收录



量化交易&宽客

发布JoinQuant 社区的好文章, Welcome to JoinQuant!

[进入专栏](#)

## 推荐阅读

### 【量化课堂】股票手续费，一文全学会！！

JoinQuant聚宽量化平台，免费提供数据获取、研究环境、策略回测、实盘连接、发送交易信号、学习交流等功能，欢迎体验 &gt;&gt;&gt; JoinQuant聚宽导语：在股票交易中，手续费往往是最容易被... [查看全文](#) >

高斯蒙 · 9 个月前 · 发表于 量化交易&宽客

## 【量化课堂】多头趋势回踩策略

导语：雪球文章《多头趋势回撤点：一个好懂又好用的均线策略》介绍了一个择时选股的辅助判断方法。这篇文章将根据这个思路构建一个交易策略，并使用历史回测来检验它的效果。感谢佰股精 。... [查看全文](#) >

高斯蒙 · 10 个月前 · 发表于 量化交易&宽客



## 怎样在陌生的城市挑餐厅

发于公众号LeanTravel难得出篇「方法论」，分享个人在旅途中觅食的所有心得。为了避免看着太... [查看全文](#) >

鱼丸 · 1 个月前 · 编辑精选 · 发表于 极简旅行



## 115年历史首战英超！“海鸥”布莱顿前来报道！

作为球迷，如果你偶尔足彩怡情的话，那你肯定会知道，本赛季的英冠存在着一支“宇宙级稳胆”... [查看全文](#) >

红白塞思克 · 1 个月前 · 编辑精选 · 发表于 红白塞思克和你谈谈足球