

潜力股挖掘器详细设计

一 项目背景

本程序实现了一个股票预测系统，它可以实现以下三个主要功能：

- (1) 预测明日股价
- (2) 挖掘潜力股
- (3) 显示股票的价格波动

用户可以通过使用本系统预测某支股票的明日股价，也可以使用本系统发掘出将有涨幅的潜力股。

二 预测算法原理：梯度提升树 GBDT

GBDT 作为集成学习的一个重要算法，在被提出之初就和 SVM 一起被认为是泛化能力较强的算法。具体而言，GBDT 是一种迭代的决策树算法，它基于集成学习中的 boosting 思想，每次迭代都在减少残差的梯度方向新建立一颗决策树，迭代多少次就会生成多少颗决策树。其算法思想使其可以发现数据中有区分性的特征以及特征组合。

GBDT 主要结合回归树 (Regression Decision Tree) 和提升树 (Boosting Decision Tree) 的思想，并提出利用残差梯度来优化回归树的集成过程。

三 tushare 模块

Tushare 是一个免费、开源的 python 财经数据接口包。主要实现对股票等金融数据从数据采集、清洗加工 到 数据存储的过程，能够为金融分析人员提供快速、整洁、和多样的便于分析的数据，为他们在数据获取方面极大地减轻工作量，使他们更加专注于策略和模型的研究与实现上。考虑到 Python pandas 包在金融量化分析中体现出的优势，Tushare 返回的绝大部分的数据格式都是 pandas DataFrame 类型，非常便于用 pandas/NumPy/Matplotlib 进行数据分析和可视化。当然，如果您习惯了用 Excel 或者关系型数据库做分析，您也可以通过 Tushare 的数据存储功能，将数据全部保存到本地后进行分析。应一些用户的请求，从 0.2.5 版本开始，Tushare 同时兼容 Python 2.x 和 Python 3.x，对部分代码进行了重构，并优化了一些算法，确保数据获取的高效和稳定。

四 历史数据的获取

我们用的是 get_hist_data 函数 ,获取个股历史交易数据（包括均线数据），可以通过参数设置获取日 k 线、周 k 线、月 k 线，以及 5 分钟、15 分钟、30 分钟和 60 分钟 k 线数据。本接口只能获取近 3 年的日线数据，适合搭配均线数据进行选股和分析，如果需要全部历史数据，请调用下一个接口 get_h_data()。

参数说明：

- code: 股票代码，即 6 位数字代码，或者指数代码（sh=上证指数 sz=深圳成指 hs300=沪深 300 指数 sz50=上证 50 zxb=中小板 cyb=创业板）
- start: 开始日期，格式 YYYY-MM-DD
- end: 结束日期，格式 YYYY-MM-DD
- ktype: 数据类型，D=日 k 线 W=周 M=月 5=5 分钟 15=15 分钟 30=30 分钟 60=60 分钟
- retry_count: 当网络异常后重试次数，默认为 3
- pause:重试时停顿秒数，默认为 0

返回值说明:

- date: 日期
- open: 开盘价
- high: 最高价
- close: 收盘价
- low: 最低价
- volume: 成交量
- price_change: 价格变动
- p_change: 涨跌幅
- ma5: 5 日均价
- ma10: 10 日均价
- ma20:20 日均价
- v_ma5:5 日均量
- v_ma10:10 日均量
- v_ma20:20 日均量
- turnover:换手率[注: 指数无此项]

调用方法:

```
import tushare as ts
```

```
ts.get_hist_data('600848') #一次性获取全部日 k 线数据
```

五 股票预测里梯度提升树算法的训练

训练样本的特征是 (12 维 注意没有 close 收盘价)

- open: 开盘价
- high: 最高价
- low: 最低价
- volume: 成交量
- price_change: 价格变动
- p_change: 涨跌幅
- ma5: 5 日均价
- ma10: 10 日均价
- ma20:20 日均价
- v_ma5:5 日均量
- v_ma10:10 日均量
- v_ma20:20 日均量

标签是第二天的收盘价(close)

根据前一天的开盘价,成交量等特征预测第二天的收盘价

六 模块划分

(1) 显示价格波动

show_history_price 从服务器获取不同粒度的历史数据 (月周日等) 用 matplotlib 显示到屏幕上

（2）推荐潜力股

`recommand_potential_stock` 我们如果预测某支股票的股价第二天涨了，这支股票就成为潜力股
我们对 600000 到 601000 的股票进行预测 把第二天会涨价的输出到界面上

（3）预测明日股价

`predict_tomorrow_price` 利用梯度提升树进行回归分析，进行明日股价的预测

输入输出

显示价格波动

输入：股票代码，输出：股票的价格波动图

推荐潜力股

输入：无 ，输出：股票代码和预测涨幅

预测明日股价

输入：股票代码 ，输出：预测的明日股票的价格

除了这三个模块，还有一个模块是图形化界面绘制，为这三个模块服务

七 调用的包

图形化使用的是 `python` 自带的 `tkinter` 库

八 功能要求

用户在文本框中输入股票的代码，点击“预测明日股价”按钮，窗口里即可以显示预测得到的明日股价。

用户在文本框中输入股票的代码，点击选择“日”，“周”，“月”，“时”，屏幕上即可按照不同的指定粒度显示股票的价格波动。

用户点击按钮“推荐潜力股票”，屏幕上即可以显示出预测有涨幅的股票代码和预测的涨幅。

九 界面设计



控件功能：

文本框：输入股票的六位代码

显示价格波动按钮：点击调用显示股票价格波动模块

推荐潜力股票按钮：点击调用推荐潜力股票模块

预测明日股价按钮：点击调用预测明日股价模块

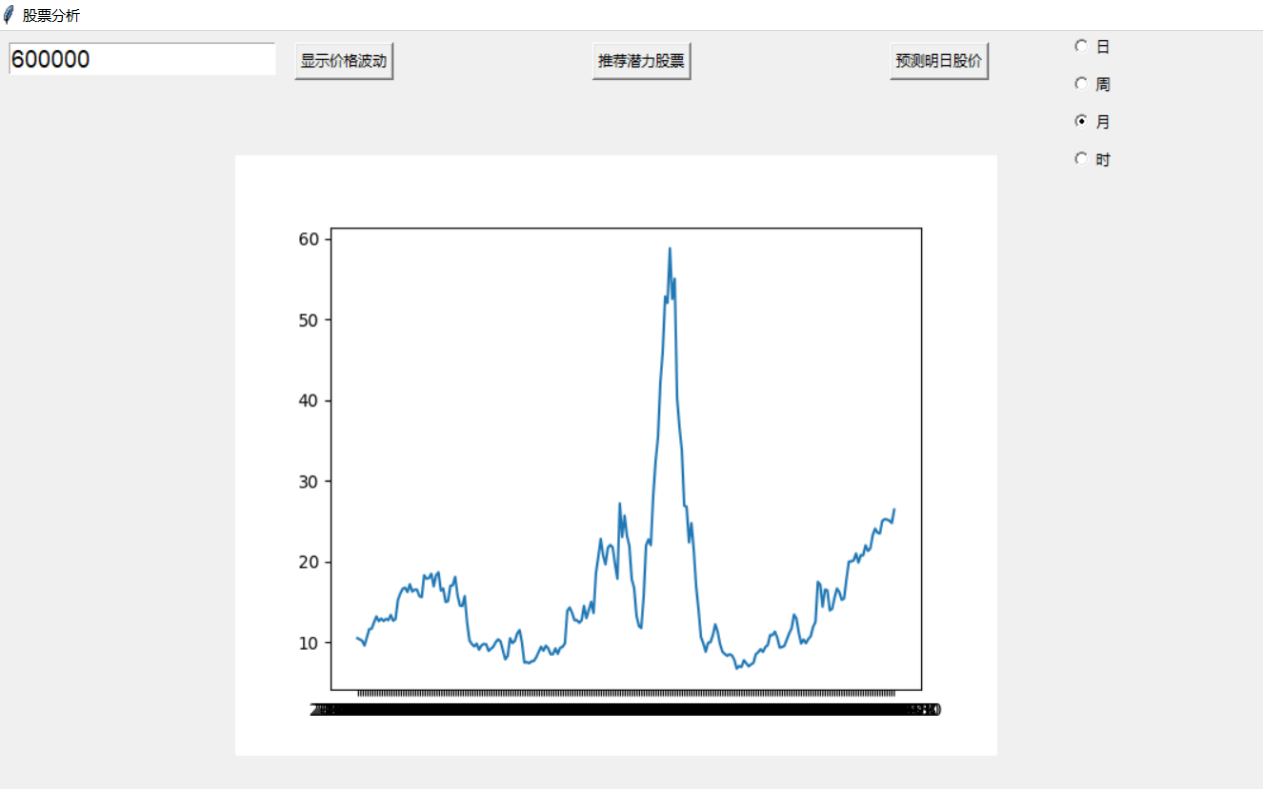
日，周，月，时按钮：点击选择不同的股票价格变化显示粒度

运行结果：

1. 显示价格波动

显示某只股票的价格波动

输入股票代码：600000 选择变化梯度：月 结果截图如下：



2.推荐潜力股票

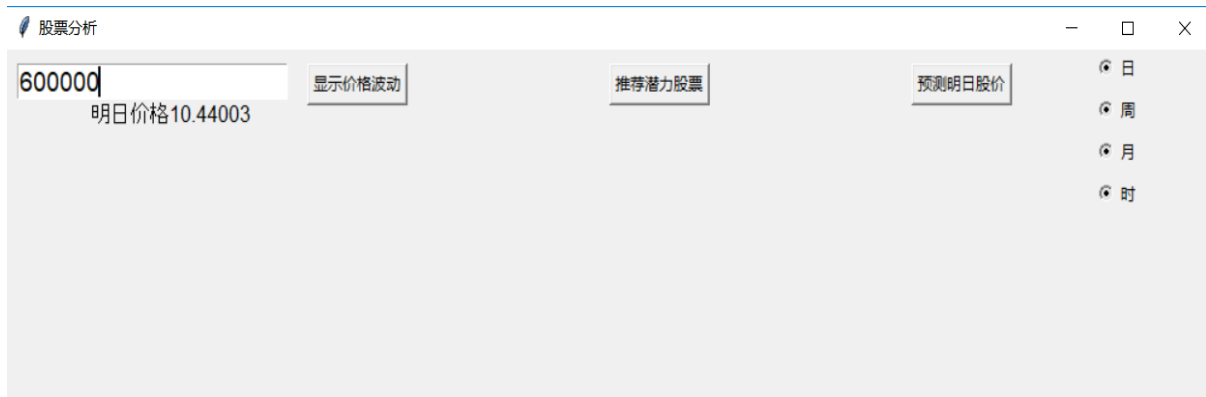
如果某只股票第二天的收盘价比第一天的价格要高，则这只股票成为潜力股



3.预测明日股价

输入股票的代码，显示股票明天的收盘价

输入股票代码：600000，结果截图如下：



4.模型对比

我们使用了四种模型进行对比.

分别是 Gradient Boosting Regressor, Support Vector Regressor, Linear Regression 和 SGDRegressor. 评测指标是 R2 指数.

Gradient Boosting Regressor:0.9955

Support Vector Regressor score:-0.049

Linear Regression:0.9968

SGD Regressor score:-2.06

这四种模型中 Linear Regressor 的效果最好.