

迭代一问题

黄蕾

南京大学软件学院

2017-03-16



GitLab相关

- Q：注册方面的问题

- A：

- ①没按要求注册的同学，如果邮箱中有学号信息，助教已经把username和name都改为学号，使用学号和密码进行登录。
- ②注册时，name、username、email和学号完全无关的账号，都被标记为无法识别身份，不予通过验证。
- ③依照3月初的花名册，还有三位13级同学，两位15级同学没有注册gitlab，相关学号已在moodle讨论区公布。

GitLab相关

- Q : 如何修改Gitlab密码 ?
- A :
 - ①在左侧导航栏点击自己的用户名
 - ②Profile Settings -> Password

GitLab相关

- **Q：项目文档提交到Gitlab中还是moodle上？**
- **A：**
 - **文档也都提交到Gitlab项目中，需要提交哪些文档与相关课题助教确认。**
 - **文档在项目中的位置没有硬性规定，但需要在README.MD中声明，方便助教检查。**

GitLab相关

- **Q : commit之后个人contributions没有更新相关信息，对成绩是否有影响？**
- **A :**
 - ①检查是否缺失push操作。
 - ②在Gitlab的项目commits中查看自己的commits数目是否正确。
 - ③本地查看git log是否存在该次commit信息。
 - 以上各项检查无误后，不会影响到成绩。

GitLab相关

- 由于系统本身的原因，同学们在fork团队项目到自己个人项目时可能出错
- 部分小组创建了多个团队仓库，为便于助教分析各阶段统计结果以及为作业打分，请只保留一个团队仓库，多余的仓库删除掉
 - 助教评分时对一个小组只批改其一个仓库

GitLab相关

- 使用本地自定义user.name的人数众多.....
- 部分数据截图如下

user.name	Project	提交数
Jiebaoda-Wang	151250125_cseiii_Quantour / QuantourExploration	9
KeekyG	151250125_cseiii_Quantour / QuantourExploration	4
Luo	151250058_cseiii_Quantour / Quantour	62
guyifei	151250058_cseiii_Quantour / Quantour	23
Example Surname	151250058_cseiii_Quantour / Quantour	12
HikariChao	151250018_cseiii_Quantour / QuantHub	1
foxwel88	151250018_cseiii_Quantour / QuantHub	1
IDontKonwDontCallMe	151250129_cseiii_Quantour / Quantour	16
JoeNJU	151250129_cseiii_Quantour / Quantour	3
Fullcreammilk	151250119_cseiii_MovieReview / MRAS	1
151250195@smail.nju.edu.cn	151250195_cseiii_MovieReview / DHMovie	13
xyj15	151250195_cseiii_MovieReview / DHMovie	4
zz	151250195_cseiii_MovieReview / DHMovie	2
DemonKatana	151250070_cesiii_Quantour / Stock	4
luluchencheng	151250114_cseiii_Quantour / RichMan	5
insomniaLee	151250114_cseiii_Quantour / RichMan	3
ParanoiaSun	151250208_cseiii_MovieReview / RottenPotatoes	28

GitLab相关

- 从**3月16日晚12点**开始，仍然不按照要求配置git，继续使用自定义user.name的同学将**没有个人成绩**
 - 直接把committer name修改为自己的学号
 - 助教不再处理user.name和学号之间的对应关系，无对应关系的commits等信息将被忽略，不计入同学的个人贡献中
- 关乎同学们的个人成绩，请相互转告

CASE1相关

- **Q:电影数据源有接近八百万条，是否考虑筛减一些数据减少工足量？**
- **A:考虑到数据量太大，故进行数据筛选，挑选出了电影评论数top1000的电影作为筛选数据集。**

CASE1相关

- Q:电影数据中有很多数据是重复的，即除了productId和userId不同，其余信息包括评论内容、评论时间等等全部一样？
- A:对于相同内容的评论，对于不同的productId，如果进网页观察的话，发现其实这些productId对应的是同一部电影的不同版本，所以需要同学们自行剔除相同的电影评论，剔除后对后续的预测分析也没有影响。

CASE1相关

- Q:对一个用户如果只有一条评论，对该用户做相关评论统计，统计需求功能无法实现？
- A:对只有一条评论的用户，因为数据量太少不适合做统计，统计也没有效果，所以查询的时候可以对有较多评论的用户进行统计展示。

CASE1相关

- **Q:电影统计信息展示需求中，统计该电影评论在时间上面的分布具体指的是什么？**
- **A:指的是统计该电影的评论数量在时间上面增长（年，月，日）**

CASE1相关

- **Q:用户统计信息中文字描述长度的分布需求具体指的是什么？**
- **A:描述文字长度分布需求，指用户评论电影字数0~200+，每20一个step，200+作为一个区间，展示评论字数在这些区间上的分布，图形不做要求。**

CASE1相关

- **Q:对两个类做分词和词频以及关键字显示具体指的是什么？**
- **A: 对电影类是电影A的所有评论中使用最多的5个词；对用户类是用户B所有电影评论中使用最多的5个词。**

CASE2相关：图标绘制

- **Q:能否使用JS绘图？**
- **A:不能。**
- **Q:迭代一的界面可以用fxml和css吗？**
- **A:可以使用，但不能使用Javascript,主要是不能通过JavaFX的webview直接加载网页。**

CASE2相关：名词解释

- Q:收盘价和复权收盘价的差异，应该在什么情况下使用？复权价是将价格调整为相同股本口径的价格。简单说，这个价格更能体现股票的真实涨跌情况，但不是股票的真实交易价格。
- A:
 - 对数收益率和涨跌幅使用复权收盘价（Adj Close）计算,别的地方使用收盘价
 - 参考Quantour名词补充解释文档

CASE2相关：异常数据处理

- Q:缺少前一日的数据怎么处理？
- A:
 - 缺失数据一个是因为有非交易日(周末和节假日)的存在，这个对所有股票都是一样的，计算涨跌幅是按照上一交易日计算的；还有个可能是股票停牌，这种情况复盘后的涨跌幅也是根据停牌前上个交易日计算。
 - 绘制k线图时均线图的横坐标是交易日，也就是说剔除了非交易日，也剔除了由于停牌数据缺失的日期。

CASE2相关：异常数据处理

- **Q：CASEII数据中英文字符有一部分是全角的，一部分是半角的，同一字符的全角半角的编码并不相同。这次项目怎么处理？**
- **A：本项目中全角半角不加以区分，请同学们处理数据时完成全角半角的统一。**

CASE2相关：异常数据处理

- Q 在提供的数据文件中，有的数据是成交量是0，但是还是出现在了文件里，这些记录是否意味着当日为非交易日？这些数据是否应该参与涨跌幅等的计算呢？
- A：这些数据属于无效数据，直接舍弃。

CASE2相关：异常数据处理

- Q：有些时候股票的涨幅和跌幅绝对值特别大，达到了70%甚至120%，请问这些数据算作脏数据，还是由于某种原因造成的属于正常数据，算作涨停 / 跌停呢？
- A：可以通过其他网站核对一下数据，数据有误的话算脏数据舍弃，当然也有可能因为某些特殊情况导致涨跌幅暂时不受10%限制。
- PS：上面两种情况为“脏数据”的示例

CASE2相关：软件设计

- Q:准备设计一个股票类来存储数据，但这么做的话对象的数量会很多（800个以上），是否可行？
- A:生成一个类的多个实例是可行的。

CASE2相关：软件设计

- Q:是否要为之后的可扩展性进行预先设计
- A:
 - 每个迭代是基于上一个迭代的产品去开发而不是重新写，在设计的时候考虑可扩展性是必要的
 - 但也不用过度设计

CASE2相关：需求

• Q:

- 关于股票间的比较，最低值和最高值是每天的最高值和最低值还是这段时间的最高值和最低值？
- 同样的涨幅和跌幅是每天都要比较，还是比较这段时间的最值？
- 还有对数收益率？

• A:

- 最低值和最高值、涨幅和跌幅都是指这段时间内两只股票的最低值最高值、涨幅跌幅。
- 计算涨幅、跌幅时使用最后一日收盘价和第一日的前一日的收盘价计算涨跌幅。
- 对数收益率则要计算这两只股票每天的对数收益率，并计算这段时间内两只股票的对数收益率方差。第一日的对数收益率也要计算，使用前一日的复权收盘价。

CASE2相关：需求

- Q:绘制均线图时，前五日的数值怎么计算

举例说明：某股连续十个交易日收盘价分别为：（单位：元）

8.15、 8.07、 8.84、 8.10、 8.40、 9.10、 9.20、 9.10、 8.95、 8.70

以五天短期均线为例：

第五天均值= $(8.15+8.07+8.84+8.10+8.40)/5=8.31$

第六天均值= $(8.07+8.84+8.10+8.40+9.10)/5=8.50$

第七天均值= $(8.84+8.10+8.40+9.10+9.20)/5=8.73$

第八天均值= $(8.10+8.40+9.10+9.20+9.10)/5=8.78$

- A:

- 使用查询日期前几日的数据进行计算，如果股票刚发行，前N日的N均线图不用画
- 比如搜索2010.3.1-2010.3.10，数据源里是有2月25~3月1号的数据的，就可以用这些数据画均线图

CASE2相关：需求

- **Q:需求描述中所说的“用户输入股票编号（或股票名称）”这句话是我们自由选择股票的输入形式的一种来进行实现 还是我们写的功能能够使得用户不论输入股票编号还是输入股票名称都可以进行相关信息的查询？**
- **A:这两个功能建议都实现，可以提供选项，让用户选择输入股票编号查看还是输入股票名称查看。**

CASE2相关：需求

- **Q:对于数据的处理，是否可以提前将原始文件的数据拆分处理写入新的文件，而不使用提供的原始文件？**
- **A:允许先预处理数据，使用处理好的数据作数据源。**

CASE2相关：需求

- **Q:是否要求要使用客户端服务器结构？迭代一可否只使用本地化的程序**
- **A:迭代一可以使用纯本地程序，但请做好分层，以便之后的迭代开发。**

CASE2相关：提交

- Q:请问迭代一上交产物中文档包括哪些呢？迭代二的计划文档包含在内吗？
- A:组内完成并且认为需要提交的产物就提交，迭代二计划可以在周四上完课再完成。

谢 谢

leihuang@nju.edu.cn