



海南大学
HAINAN UNIVERSITY

数据结构课程设计论文

基于最小生成树的股票分层工具 ——以沪深300为例

2015.12.11

卢俊星
海南大学

StoryInStone@outlook.com

<https://github.com/StoryInStone>

通信工程

20120802310083

目录

1	概念	2
1.1	研究背景与意义	2
1.1.1	研究背景	2
1.1.2	研究作用与意义	2
2	目标	3
2.1	股票分层工具	3
2.2	巩固数据结构基础	3
2.3	训练相关算法	3
2.4	项目与工具	3
3	知识与应用	4
3.1	数据结构基础应用	4
3.1.1	向量	4
3.1.2	图	4
3.1.3	最小生成树	4
3.2	自动化项目管理知识应用	4
3.2.1	Make工具	4
4	方法	5
4.1	编程与写作环境	5
4.2	贪心算法求最小生成树	5
4.3	开源支持	5
4.3.1	股票数据调取项目——Tushare	5

1 概念

1.1 研究背景与意义

1.1.1 研究背景

对于风格投资的研究，国外的已有文献较多，Farrell[1]和Sharpe[1]都发现了股票收益间较高的相关性，与风格相关。一些实证研究也表明，基础价值协同性并不是收益协同性的唯一根源，Fama[1]和French的研究表明某些股票的收益协同性与基础价值协同性无关；另外，大量文献证明了初基础价值外，投资者偏好、交易地点、现金流、指数都可以导致收益协同性形成风格。Barberis等总结了已有的研究成果并提出了风格投资协同理论——风格投资产生收益协同性。对于同种风格的股票之间存在较强的相关性。

1.1.2 研究作用与意义

指数分层结构图横坐标为股票名称，纵坐标为股间距离。通过绘制指数分层结构图，可以清晰的看到股票的聚集状态，推导出风格的分布和组合样本。继而为研究风格形成因素、收益及风险水平等性质开辟了道路，为证券投资组合配置提供依据和参考。从上述可知，指数分层结构图对风格投资分析的重要性，然而在绘制指数分层结构图时用到的数据相当庞大，因此研究指数分层结构图的绘制算法具有实际的意义与价值。

2 目标

2.1 股票分层工具

制作一个简单的股票分层工具，能对沪深300中的股票进行聚类，输入股票代码能得到其同类风格的股票代码，给投资者提供分析基础。

2.2 巩固数据结构基础

数据结构是构建计算机世界的砖瓦，在本项目中，通过相关训练，巩固课堂上所学习的数据结构基础。

2.3 训练相关算法

算法的知识较为复杂，通过本文的训练，目标是能够入门写一些简单的算法程序，如最小生成树MST程序。

2.4 项目与工具

好的项目管理思想以及相关工具的使用，能够使工作事半功倍，在本项目的编码过程中，熟悉相关思想和工具使用方法。

3 知识与应用

3.1 数据结构基础应用

3.1.1 向量

本文中有沪深300股票代码数据，股票每日收盘数据，股票在某一时段内价格变化趋势序列三类主要数据，由于数据在程序的相关处理过程中并不需要动态操作，因此本文采用向量结构对数据进行存储和传递。

3.1.2 图

本文中研究的A股市场中代表性股票之间关系，任两支股票间都存在两两股间距离，故形成无向连通图，由于两两股票间的距离一定存在，故图为稠密图，本文采用邻接矩阵来表示股间距离图。

3.1.3 最小生成树

任何连通图都必存在一最小生成树，使图不成环。树结构可以转化为相应的层次结构，因此该最小生成树与股票分层结构图是一一对应的。

3.2 自动化项目管理知识应用

3.2.1 Make工具

4 方法

4.1 编程与写作环境

本文采用C/C++编译器为GCC 5.1.1，系统环境为Fedora 23，代码编辑器为VIM 7.4，项目编译工具为Make 4.0，文档编辑工具为Latex。

4.2 贪心算法求最小生成树

贪心算法一般不能解决实际问题，但是只要能使用贪心算法解决的问题，那么对于该问题而言，此算法即为最优的。

4.3 开源支持

4.3.1 股票数据调取项目——Tushare

本文中所有股票数据全部来自于Tushare财经接口工具包项目，该项目采用Python编码，从各大官方财经站点调取股票数据。

Literatur

- [1] Turing, A. M. On computable numbers, with an application to the entscheidungsproblem. *J. of Math* 58, 345-363 (1936), 5.