1. 资源下载

1.1 数据集

<http://pan.baidu.com/s/1i3Fm477>

2. 项目配置

新建控制台工程然后添加文件

项目结构:

main.cpp

inc/

main.h

global.h

…其他所有.h/.cpp文件

3. 算法总体过程

3.1 概述

1选择要处理的用户数量

2对每个用户

2.1读取数据集点

2.2处理转换成停留点

3对所有停留点进行聚类

4聚类结果优化

5对每个用户生成相应的簇序列

3.2 备注

步骤1,2,4,5由项目完成,步骤3使用数据挖掘平台(ELKI)完成.

4. 代码解析

具体步骤代码部分都作了注释.如果不了解算法参数请查看相关论文.

main()里

执行test\_paper()即为执行步骤1,2

执行test\_paper2()即为执行步骤4

执行test\_paper3()即为执行步骤5

5. 具体步骤

5.1 生成停留点

main()里使用test\_paper()编译生成exe,放到数据集里的Data文件夹下执行,用户数为182.

**输出(文件):**

allspex\_latlng.txt 所有停留点的经纬度信息

allspex 所有停留点的所有信息

5.2 聚类

双击elki.jar打开数据挖掘平台会有一个minigui的窗口,load设置stitp opticsxi,在上面的dbc.in处写上输入数据的地址,也就是allspex\_latlng.txt的地址,resulthandler的out处写输出的文件夹位置(假设输出的文件夹名为AAA),然后Run Task执行.因为聚类所有的停留点所以过程漫长,大概要30min-1h.

聚类结束后查看输出文件夹,因为minigui在算法结束后就会有提示,但是输出到文件还需一些时间,等所有文件都输出完了关闭minigui.

再次打开minigui,load设置stitp opticsof,输入依然是allspex\_latlng.txt,输出位置选择和刚才不一样的文件夹(假设输出的文件夹名为BBB),Run Task执行.这次过程会很快结束,同样确保文件输出完成后才可以关闭minigui,运行完控制台会有警告,无视.

关于这个平台更多的使用见官网<http://elki.dbs.ifi.lmu.de/> 这个平台有很多数据挖掘的算法.

**输出(文件夹+文件):**

AAA/ opticsxi算法的簇信息输出

BBB/ opticsof算法的各停留点of值信息

5.3聚类结果优化

将BBB文件夹里的optics-outlier\_order.txt复制到AAA文件夹,将5.1生成的allspex 改名为allspex2后放到AAA文件夹,main()里使用test\_paper2()编译生成exe,放到AAA文件夹下执行,ratio为2.

**输出(文件):**

allspex 优化后的所有停留点的所有信息

5.4对每个用户生成簇序列

新建目录(如CCC)用来保存每个用户的簇序列,将5.3生成的allspex改名为allspex2后放到CCC里,main()里使用test\_paper3()编译生成exe,放到CCC下执行.

**输出(文件夹+文件):**

CCC/ 所有用户的簇序列文件

5.5对某一个用户的序列模式挖掘和准确度分析。

步骤5.4种会生成多个停留点文件，选出任何一个用户的文件放在E:\\（可以修改），改名为xx.txt，然后执行工程PreFixSpan\_and\_Predict 生成的的exe即可。PreFixSpan算法百度即可，算法无修改。Predict算法参见附件中的论文。

6. 一些注意点

minigui使用一次之后必须重开再使用,否则点的ID会有错(因为他内部默认递增,而我们2次算法运行需要同样的ID序列).

3次生成的exe可以重复用.