1. 利用R平台熟悉社交网络点赞数据分析流程。具体R安装、使用等问题可咨询助教。

users-likes.csv文件中给出了用户和点赞的关系， 这个关系可以用矩阵表示，即User-Footprint Matrix。但是这个矩阵很大。M.csv文件通过修剪掉对过于稀疏的行和列，得到了一个19742\*8523大小的矩阵，以稀疏矩阵(i,j,1)的形式存储。

可以通过R语言命令：

M <- read.csv("M.csv")

library(Matrix)

require(Matrix)

M <- sparseMatrix(I=M$I, j=M$j, x=1)

将这个矩阵在R中加载出来。

参考如下文献并在R平台实现文中的如下结果。

a).下载并读取数据集，查看数据集前6行数据和数据特征维度。

b).构建User-likes矩阵M，行为用户ID，列为喜爱内容。修剪矩阵 M 以删除太稀有的用户和喜欢内容。

c).使用SVD分解和LDA分解，降低矩阵M的维度。

d).使用相关性分析，研究 LDA 集群或SVD 维度与心理人口特征之间的关系。用热力图表示相关矩阵。

e).建立预测模型，将用户分成 10 个组，从组 1 中获取用户并将其作为测试集，其余组数据作为训练集。从训练集中提取 SVD 维度，建立关于用户openness的线性回归模型（openness作为因变量），计算回归系数，建立关于用户gender的线性回归模型（gender作为因变量，family = "binomial" ），计算回归系数。检验两个模型的准确性（比如可以计算预测值和真实值的相关系数），尝试可视化结果。计算10 个分组交叉验证的预测精度，估计LDA聚类分数。

[1] Michael Kosinski, etc. Mining Big Data to Extract Patterns and Predict Real-Life Outcomes. Psychological Methods, 2016, Vol. 21, No. 4, 493–506.

[2] Nina Zumel, John Mount, Jeremy Howard, Rachel Thomas - Practical Data Science With R-Manning Publications (2020).

[3] 参考相关文档（《An introduction to R》）或者自行网络查询。

2. 如何理解Netflix例子中的概念Genre, 结合社交网络点赞数据的分析结果给出实例。