# 同济大学软件学院 2018年春季数据分析/挖掘课程作业之2

1. [频繁集挖掘] 以reco\_data.rar的trade文件和最新的trade\_new.csv分别作为交易数据库，挖掘其中的频繁集，要求对每个VIPNO对交易数据进行分组，按照sldat字段为标准取每个用户最早60%的交易记录作为训练数据，完成频繁集挖掘。
   1. 无需考虑时序性
      1. 分别以pluno、dptno和bndno为频繁集挖掘的item\_no，通过FP-Growth算法挖掘pluno、dptno和bndno的频繁集，其中要求support为2，4，8，16，32，64等共计6个阈值
      2. 若把每个用户的交易记录合并，这样一个用户只存在一个（合并后的）交易记录，然后再重复a)中的步骤，其中要求support为2，4，6，8，10等共计6个阈值
   2. 按照sldat字段的信息，设计sequential FP，重复上述a).i和ii的步骤。
   3. 加分题：利用上述频繁集结果预测该用户接下来的购买信息，这里以每个用户余下40%的交易记录作为ground truth评价该预测的准确性。
   4. [SPMF](http://www.philippe-fournier-viger.com/spmf/index.php?link=algorithms.php)提供参考频繁集挖掘代码实现。

提交日期：2018/05/06日 23：59PM,

提交内容发送至tongjidam18@163.com，邮件题目为hw2+学号，提交内容包括：

1. 每个作业提交内容以学号+hw2.zip作为命名方法；其中包括2-3个子目录，分别对应命名方式分别为ai、aii、bi、bii、ci、cii，每个子目录包括pluno、dptno和bndno三个items对应目的代码、运行时间和word报告。其中报告包括1）代码运行结果屏幕拷贝；2）讨论分析部分；3）性能比较图表