作业4

朱志儒 SA20225085

6.2

（a）仅使用G和他们的支持度计数不能确定项集A是否频繁，需要闭频繁项集的信息。

算法：

输入：数据集D上所有闭频繁项集的集合G，每个闭频繁项集G[i]的支持度计数C[i]，支持度阈值e，给定项集X；

输出：项集X是否为频繁的，如果是，给出X的支持度；

过程：对G中的闭频繁项集按支持度递增的次序排序，得到一个有序的闭频繁项集的集合，用分别表示第i个闭频繁项集和他的支持度计数

:

统计中的支持度，中所有集合的支持度计为

项集X不是频繁的

（b）项集X是数据集D上的闭项集，如果不存在真超项集Y使得Y与X在D中具有相同的支持度计数。

项集X是数据集D上的生成元，如果不存在真子集使得。

显然，闭项集考虑的是真超项集，而生成元考虑的是真子集。闭频繁项集包含对应频繁项集的完整支持度信息，而频繁生成元不包含对应频繁项集的完整支持度信息。

6.6

：项集 支持度计数 ：项集 支持度计数

{M} 3 {M} 3

{O} 4 {O} 4

{N} 2 {K} 5

{K} 5 {E} 4

{E} 4 {Y} 3

{Y} 3

{D} 1

{A} 1

{U} 1

{C} 2

{I} 1

：项集 支持度计数 ：项集 支持度计数

{M,O} 1 {M,K} 3

{M,K} 3 {O,K} 3

{M,E} 2 {O,E} 3

{M,Y} 2 {K,E} 4

{O,K} 3 {K,Y} 3

{O,E} 3

{O,Y} 2

{K,E} 4

{K,Y} 3

{E,Y} 2

：项集 支持度计数 ：项集 支持度计数

{O,K,E} 3 {O,K,E} 3

{K,E,Y} 2

频繁项集：{M}，{O}，{K}，{E}，{Y}，{M,K}，{O,K}，{O,E}，{K,E}，{K,Y}，{O,K,E}