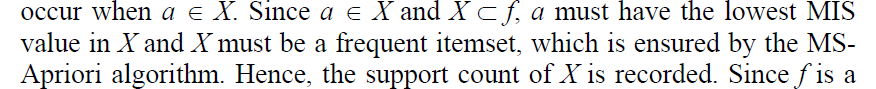
方骏-2020年2月9日-读书报告

1. 自己提出的问题的理解：
2. 提出的问题1：为什么multiple minimum supports问题不再适用于downward closure property了？能否给出一个比较完整的证明？

讨论后的理解：讨论后我认为这个问题主要出现在一个item到itemset中有两个item的问题中，因为有两个item的itemset中的一个item的support count可能会小于自己的MIN，但是它在itemset中可以不作为第一个，这样也是有可能的，但是这个item不能作为独立的itemset，所以这样downward closure property就不满足了。

1. 提出的问题2：反证法证明2.4.3的lemma的时候，为什么



a属于X那么X就一定是一个frequent itemset。

讨论后的理解：a既然是属于X的，那么就说明X中的最小的MIN就是a的MIN。那么显然所有的X中的其他item的MIN都是大于a的，一定是在之前讨论过的frequent itemset中。

1. 别人提出的问题的理解：
2. 问题3：算法中为什么记录了f – {a}就能确保所需的非frequent的condition的count被记录下来？

自己的理解：这个其实也比较容易理解，首先如果a出现在规则的左边，那么左边的itemset就一定是frequent itemset。那么至于为什么单单是c-c[1]就可以保证这是所有没算过的非frequent了呢？假设有一个c-c[1]-c[k]也是非frequent的，但其实这个在计算c-c[k]层的时候就已经计算过了，在该层对于c，只有c-c[1]没算过了，所以可以确保。

1. 问题4： 在MS-Apriori算法中，level2-candidate-gen(L, ϕ)和MScandidate-gen(Fk−1, ϕ）的内容有类似的地方，也有不同的地方，为什么两种function的内容有所不同。

自己的理解：因为这就是一个特殊的情况。对于两个item的itemset，是存在包含有其support count小于MIN的item，导致该item无法作为F1的itemset。那么如果只用MScandidate函数，这个itemset就不能作为候补C2中的一员。但是显然，这个item可以不做第一个，可以作为一对中的第二个item，这也是可以允许的。所以才会有这么一个特殊的情况存在，其他的k值就不会有此问题了。

5、 问题5：Fig.2.8的第五行中 𝑖𝑘−1<𝑖𝑘−1′ 是不是为了避免重复讨论，以及它们的大小是按什么标准比较的？（可能是字典序？）

自己的理解：其实可以说就是字典序。因为这个顺序其实在Fk已经排列好了，直接按照Fk中给出的顺序排列就可以了。

1. 读书计划

1、本周完成的内容章节：2.3-2.5

2、下周计划：第三章

四、读书摘要及理解

1、读书摘要及理解

首先是表格的问题，对于表格要满足association rule的格式，就必须要把属性和值联系在一起构成属性、值对，即<attribute, value>，这作为一个item就可以了，剩下的和之前的minging规则是一样的。

对于多support值问题，处理的方法就是为每一个item指定一个MIS值，要求每一个itemset对中，所有的item的support count都要大于itemset中的最小的MIS值。其规则的生成方法都是类似的。其中由F1生成F2有所不同，原因在前面的问题解决中有解释，因为两个item的itemset中的item的support count会小于其MIS，这样就不在F1中，但是确实在F2中，因此由L生成，不由F1生成。这是最大的区别，其余都是类似的。

在规则的完全生成中，值得注意的是对于一个itemset集c，其中最小MIS的item是c[1]，那么只需要额外统计的非frequent集只有c-c[1]。理由可以见问题3。

在有标签的规则生成模型中，标签其实就相当于是生成规则的右侧，但是这个右侧只有item，区别与正常的规则。其余都是一样的，值得注意的是，带标签的Apriori算法在计算了support count后，还要把confidence算出来才能作为真正的一个带标签的规则。