窦一蒲-2020年2月10日-读书报告

读书报告内容：

1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1：当人们想要知道两种物品的关系时，需要计算x->y和y->x的一个support和两个不同的confidence吗？

讨论后的理解：根据实际的需求来计算，但x->y与y->x不同。

1. 提出的问题2：对于table data set的join step如何在原先生成candidate算法的基础上进行调整，使得满足一个不会产生一个candidate itemset containing two items from the same attribute,这里如何调整？

讨论后的理解：如果直接使用的话，它会把attribute-value当一个独立的item来用。而一个属性不能有两个值，因为这不符合需求，所以得用attribute-value来操作。

1. 提出的问题3：反证法证明2.4.3的lemma的时候，为什么a属于x就一定是一个frequent itemset？

讨论后的理解：a属于x说明最小mis不变，而子集一般比全集的count更大，所以一定满足frequent的条件。

1. （必填）别人提出的问题的理解（选择几个问题罗列，并给出理解）：
2. 问题4： 为什么在MS candidate函数中每次循环中使|sup(ik-1) − sup(i’k−1)| ≤ ϕ 就能令一个itemset中maxsup与minsup的差值在ϕ的范畴内？

自己的理解：如果Fk-1是frequent itemset的话 那么对于Fk-1中的任何一对item都满足|sup(x)-sup(y)|<=φ 那么唯一可能不满足的情况就是|sup(ik-1) − sup(i’k−1)| ≤ ϕ ，因此只需要考虑这两个item就行。

1. 问题5：为什么对于使用MIS的item来说，向下包闭性质不再适用？

自己的理解：当MIS相差很大时，会出现类似例9的问题。

1. 问题6：为什么在Level2-candidate-gen函数中，第五行sup(l)不会比sup(h)小？

自己的理解：因为sup是实际数据情况，MIS是设置的，两者没有直接关系。

1. （必填）读书计划

1、本周完成的内容章节：如2.1-2.6

2、下周计划：第三章3.1-3.3