



**数据挖掘**

**课后作业**

|  |  |
| --- | --- |
| 专 业： | 人工智能 |
| 班 级： | 192204 |
| 学 号： | 19220432 |
| 姓 名： | 陆昊宇 |

2024年10月28日

## 课后作业

1. 考虑表1中的候选3-项集，假设hash函数为h(x)=x mod 4，叶节点最大尺寸为2，构造hash树。

表1. 候选3-项集

|  |  |  |
| --- | --- | --- |
| **候选3-项集** | | |
| {1，4，5} | {1，5，9} | {3，5，6} |
| {1，2，4} | {1，3，6} | {3，5，7} |
| {4，5，7} | {2，3，4} | {6，8，9} |
| {1，2，5} | {5，6，7} | {3，6，7} |
| {4，5，8} | {3，4，5} | {3，6，8} |

1. 事务集如表2所示，**最小支持度阈值是30%**。

表2. 事务集

|  |  |
| --- | --- |
| TID | 项集 |
| 1 | {a,b,d,e} |
| 2 | {b,c,d} |
| 3 | {a,b,d,e} |
| 4 | {a,c,d,e} |
| 5 | {b,c,d,e} |
| 6 | {b,d,e} |
| 7 | {c,d} |
| 8 | {a,b,c} |
| 9 | {a,d,e} |
| 10 | {b,d} |

根据表2的事务集，在格结构上对**每个**结点添加**所有**符合条件的字母标记：

**N**：如果该项集被Apriori算法认为不是候选项集。（一个项集不是候选项集有两种可能的原因，一个是它们没有在候选项集产生步骤产生，另一个是它虽然在候选项集产生步骤产生，但是在剪枝步骤被丢掉）

**I**：如果计算支持度计数后，该候选项集被认为是非频繁的。

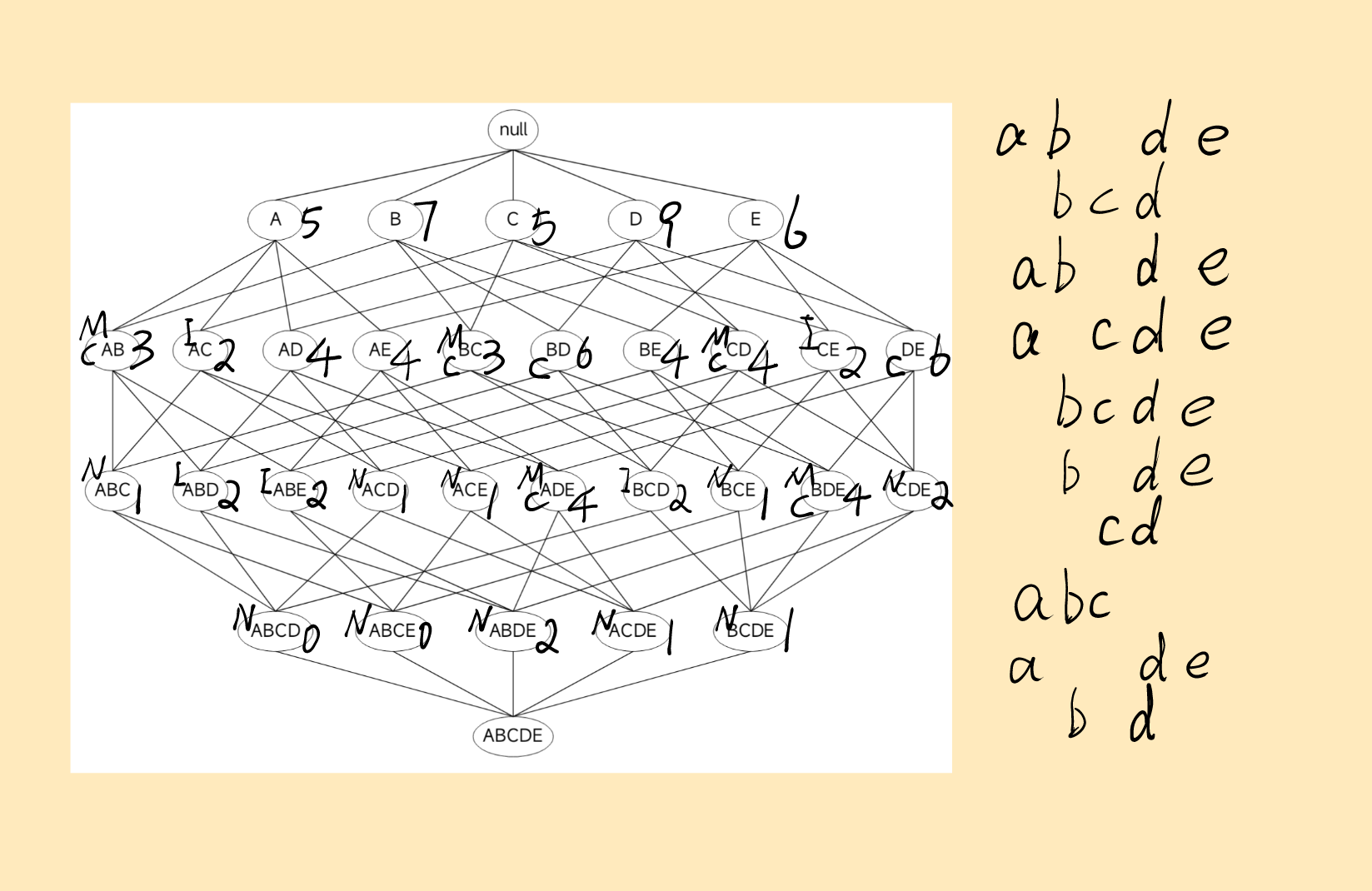
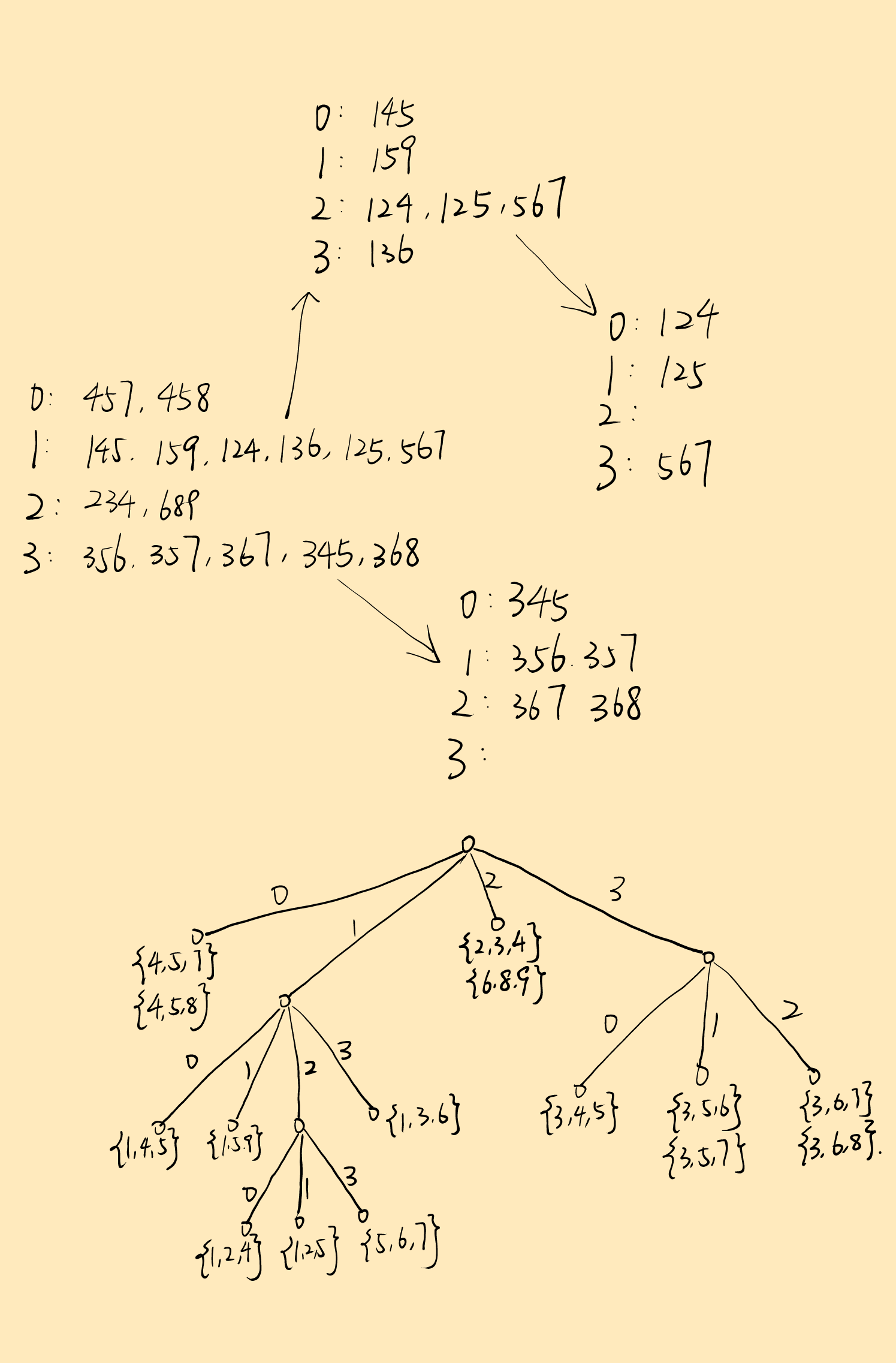
**M**：极大频繁项集。

**C**：闭频繁项集。



图1. 格结构

## 回答区域



**提交入口：**

https://send2me.cn/xvaluzS-/TcyHzhe\_Ut-tbA