1. （必填）自己提出的问题的理解（罗列全部）：
2. 提出的问题1：F序列模式挖掘出频繁序列假设为 <A,B> ,这表明A发生后隔一段时间可能发生B，但隔一段时间到底是隔多长时间呢？换句话说，如何给序列挖掘添加时限约束

讨论后的理解：序列s是序列 w=<e1e2⋯ek>的邻接子序列必须满足下面条件之一

1. s是从w删除e1或ek中一个事件后得到
2. s是从w删除至少包含两个事件的ei中删除一个事件后得到
3. s是t的邻接子序列，而t是w的邻接子序列  
   

用邻接子序列概念，可以用如下方法修改先验原理，来处理最大间隔约束。

**修订的先验原理**， 如果一个k-序列是频繁的，则它的所有邻接k-1序列也一定是频繁的。

提出的问题2：GSP算法每次计算支持度的时候都要重新遍历数据，有什么办法可以优化

讨论后的理解：SPADE的算法过程和GSP类似，只是在扫描的时候不是扫描整个数据库，而是扫描ID\_LIST.通过便利一次数据库得到的经验知识来降低多次对数据库的扫描。SPADE不仅通过减少对数据库的扫描降低I/O成本，还通过搜索ID\_list降低计算成本。

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3：prefixspan算法的核心思想是什么？

自己的理解：从长度为1的前缀开始挖掘序列模式，搜索对应的投影数据库得到长度为2的前缀。搜索长度为2的前缀对应的投影数据库，得到长度为3的前缀，一直递归到不能挖掘到更长的前缀为止。

问题4：P37页，为什么具有最小MIS的项目（设为x），不是s1中第一个项目，则s2中必然含有这个x? （例23？）

自己的理解： 如果x不在第一个项目，根据生成算法，s2中必然含有这个k-1序列。

1. （必填）读书计划

1、本周完成的内容章节：2.6-2.9

2、下周计划：5.1

四、（选做）读书摘要及理解或伪代码的具体实现（读书摘要、伪代码的具体实现代码等可以写到这个部分）

1、读书摘要及理解（选做）

2.3 Generating Rules from Sequential Patterns

This section introduces only three types, sequential rules, label sequential rules and class sequential rules, which have been used in Web usage mining and Web content mining.

## 2.3.1 Sequential Rules

A sequential rule (SR) is an implication of the form, X → Y, where Y is a sequence and X is a proper subsequence of Y. Given a minimum support and a minimum confidence, according to the downward closure property, all the rules can be generated from frequent sequences without going to the original sequence data.

## 2.3.2 Label Sequential Rules

Sequential rules may not be restrictive enough in some applications. We introduce a special kind of sequential rules called label sequential rules. A label sequential rule (LSR) is of the form, X → Y, where Y is a sequence and X is a sequence produced from Y by replacing some of its items with wildcards. A wildcard is denoted by an “\*” which matches any item. These replaced items are usually very important and are called labels. The labels are a small subset of all the items in the data.

Note that due to the use of wildcards, frequent sequences alone are not sufficient for computing rule confidences. Scanning the data is needed. Notice also that the same pattern may appear in a data sequence multiple times. Rule confidences thus can be defined in different ways according to application needs. The wildcards may also be restricted to match only certain types of items to make the label prediction meaningful and unambiguous.

## 2.3.3 Class Sequential Rules

Class sequential rules (CSR) are analogous to class association rules (CAR). Let S be a set of data sequences. Each sequence is also labeled with a class y. Let I be the set of all items in S, and Y be the set of all class labels, I ∩ Y = ∅. Thus, the input data D for mining is represented with {(s1, y1), (s2, y2), …, (sn, yn)}, where si is a sequence in S and yi ∈Y is its class. A class sequential rule (CSR) is of the form X → y, where X is a sequence, and y ∈ Y. A data instance (si, yi) is said to **cover** a CSR, X → y, if X is a subsequence of si. A data instance (si, yi) is said to **satisfy** a CSR if X is a subsequence of si and yi = y.