# 引言

（这一节介绍思路如下：先通过介绍关联分析的起源及应用来说明关联分析很有用，再通过介绍关联分析的定量属性的分割问题及其挑战，从而引出我们方案的大致思路以及我们方案的贡献点。）

1. 介绍关联分析的起源，应用。
2. 介绍关联分析的连续值问题及其挑战。
3. 对于一个整体的区间如何划分出新的粒度。
4. 对于某个划分出来粒度进行评估。
5. 我们的解决方法及贡献点。
6. 提出了一个可视化系统迭代式对属性进行粒度的划分
7. 提出一个区间整合算法将区间划分成不同的粒度
8. 提出几个指标综合展示出粒度信息

# 背景知识与相关工作：

（这一节先介绍关联分析时的一些属于(如支持度，贡献度等)。然后再介绍当前常见的处理方法，主要分成两个大的方面：通过可视化展现规则以及对于定量属性的自动划分的算法及其优缺点）

## 可视化展现关联规则

## 自动划分算法

# 总体介绍

（这一节的介绍思路如下：首先通过我们的目的（引导用户进行迭代式的粒度划分）引出我们下一步的设计要求，并详细介绍该设计要求为什么是需要的且对我们的目标有帮助，接下来则是根据设计要求设计出了两个模块以及分析流程，并说明该系统是如何满足设计要求的。）

## 设计要求

* 1. 展现出所有频繁集的大致信息以及各频繁集中的对比信息，让用户对频繁集有一个全面的了解
  2. 将频繁集的全局展现与具体频繁集细节信息展示相结合
  3. 用户可以输入自己的知识从而给粒度划分提供参考
  4. 给出信息，让用户可以判断当前粒度是否可以被接受

## 系统总览

基于以上的设计要求，所以我们设计出一个系统来解决以上问题。该系统包含以下两个模块以及一个流程：

* + 概览视图：展现当前所有频繁集。为用户理解问题提供一个全局信息(设计要求a)以及当前选中的频繁集的上下文信息(设计要求b)
  + 频繁集视图：展现被用户选中的某个具体的频繁集，为用户的决策（如：继续迭代或者该频繁集满足条件终止迭代）提供辅助信息（设计要求c,d）。

系统流程如下所示：（画出系统流程图）

# 概览视图

（这一节的介绍思路是：首先通过介绍概览视图的设计目的并据此提出如果想达到这一目的需要解决哪些问题。再用各个小节来具体介绍为了解决问题我们的分析思路以及方案。）

概览视图的目的是为了帮助用户挑选自己感兴趣的频繁集，为了实现这一目标，我们需要解决两个问题：用户可能对哪些信息感兴趣，我们应该如何展现这一信息。（主要思想）

## 在概览图中需要展现哪些信息

（这一小节的介绍思路是：因为我们是为了找到感兴趣的频繁集，所以我们从频繁集的产生过程进行分析，从频繁集本身性质以及频繁集之间的关系入手，最终明确了需要展现的两种信息。）

我们需要频繁集本身的一些性质来刻画该频繁集。

因为有些频繁集很相似，需要通过频繁集之间的关系来提高用户的选择效率。（主要思想）

### 频繁集本身的性质

因为频繁集本身就是通过支持度与置信度来确定的，所以采用展现频繁集支持度和置信度的方式来刻画该频繁集。（主要思想）

### 频繁集之间的关系

介绍为什么要采用两个频繁集的数据集的重合度作为距离作为两个频繁集的比较。

## 如何展现出这样的信息

（介绍思路：从上小节总结出来的展现信息开始，说明我们为什么要采用这样的图表来展现上述信息，以及一些细节设计的原因）

因为我们要展现频繁集中两两之间的相似度，所以将频繁集投影到二维平面中更加适合。所以采用散点图进行展现，每个频繁集的支持度和置信度分别用散点图圆圈的大小和颜色的深浅来进行表示。

为了展现与右边频繁集视图的细节信息进行关联，将当前选中的频繁集用另一种形状表示。（设计要求b）。

# 频繁集视图

频繁集视图的作用是为了提供给用户信息，让用户可以进一步判断该频繁集是否应该被接受，以及哪个区间可以进一步迭代。

所以需要解决以下两个问题：

* 哪些信息可能有助于用户进行判断
* 这些信息应该如何展现
* 系统应该如何根据用户的输入作为指导得到进一步的粒度划分结果

## 哪些信息可能有助于用户的判断

（这一段的介绍思路是：那两个比例信息为什么会有助于用户的判断，以及在什么情况下代表该粒度可以被接受，什么情况下可以继续分割）

首先介绍那两个比例信息是什么（结合ＶＮ图来说明）

那两个比例信息对用户的判断意味着什么。

因为用户是希望通过该信息得知该粒度是否真的有意义。所以问题就转化为如何判断该粒度是否在频繁集中有特殊之处。

在一般情况下，各个粒度占该粒度总体数据的比例应该基本一致，如下图所示：(用viso画一个VN图)

但如果一个粒度占比明显高于其他的粒度，则说明该粒度算是一个异常值，那么该粒度有意义的概率就大。所以我们需要展现这一信息。

如果一个粒度包含了很多数据，但是其占总体数据的比例却很低，那么这个粒度能够继续分割的概率应该也会大一点，所以我们需要展现各个粒度所包含数据量之间的对比。

## 应该如何展现信息

（介绍思路：根据上一节所说的信息，介绍我们的方案，并说明为什么该方案可以有效的展现上一节所说的信息）

## 如何根据用户的输入进一步划分

从上一步的视图中，用户可能会对于同一项在不同的频繁集中选择多个区间，如果我们直接将这些区间作为新的划分标准，那么随着迭代的进行，这些区间会逐渐被划分成一些重合很多但又有一些细微差别的区间，这样对于我们的分析可能会造成一些不利的影响，所以我们希望提出一个区间整合算法来解决这个问题。

### 区间整合算法的分析

介绍区间整合算法在划分过程中问题（如果分的过多则可能过细从而导致支持度不够而造成丢失，如果分的过大有可能出现迭代停止的情况）

### 介绍我们所提出的区间整合算法

对算法进行详细的介绍并说明该方案有助于缓解上节中所提到的问题。

# 案例分析

## 采用人造数据

## 采用交通数据

# 讨论

# 结论