这是论文《Mining Significant Co-location Patterns From Spatial Regional Objects》的报告

1. 研究动机：

这篇论文偏一点模糊挖掘的方向，以往的传统模式挖掘中，都需要给出参与度阈值才能挖掘，这篇论文提供了一种新的挖掘方法，不需要参与度阈值，此次，以往的挖掘的实例都是一个个的点，但是实际上，一个饭店，一个学校等等都是有面积的，他们用一个有面积的区域来表示比用点来表示更准确些。基于以上的两点，提出了新的挖掘方法。

1. 主要解决什么问题：

重新定义了衡量空间频繁模式的方法，不需要参与度阈值，同时提出了画网格的方法来加快速度。

1. 提出的主要方法是什么：

首先，在这篇文章中，所有的实例都将不再是点，而是一片区域,就像下图中的五边形o一样，他取代了以往的实例概念。而这个o外面的矩形称作缓冲区。

缓冲区获得方法：先将边长计算出来，然后除以边数计算出平均长度，之后以平均长度为边长，在o上画等边三角形，就像图中一样。如果该边剩余的长度小于了之前计算的平均长度，那么就画成等腰三角形（底是剩余长度，两腰是平均长度）。就这样o就会被三角形所包围。然后在外面画矩形，矩形是正好能够包围住所有的三角形的最小矩形。

缓冲区就这样完成了。

形状

描述已自动生成

接下来，为了判断两个实例之间是否临近，计算两个实例之间的邻近度。先计算出两个实例之间的NR，公式如下：

文本

描述已自动生成

其中相交的面积可以通过以下公式计算：

文本

低可信度描述已自动生成

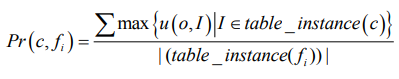


其中a.p1.x表示的是实例a在p1这个角的x轴坐标。

通过实例之间的NR，就可以计算出行实例NR(I)了，

比如说要计算{a1,b1,c1}的行实例，就可以计算出两两之间的NR，取最小的NR获得。

获得了NR（I）后，根据以下公式计算出参与率：



其中的u就代表了每个实例的NR（I），之后选取最小的参与率作为参与度PI，这里和以往的概念一样。之后我们通过数学上的置信系数来代替参与度阈值的判断,先根据这个公式计算出置信系数：

文本

描述已自动生成

之后有了置信系数，将PI除以置信系数，会得到一个值，将这个值与3比较，如果大于，那么就认定为是频繁模式。

1. 加快速度的方法：

这篇文章后面也讲了下网格法，概念差不多，就是将区域的中心点当作了之前的实例来计算。

1. 关键结果及结论是什么：

定义了一个新的概念，就是不再以点来当作实例，而是用更加实际的区域来计算，同时引入置信系数的方法，用户就不用再定义参与度阈值，为了匹配这个算法，也重新设计了网格法。