2. 流挖掘算法：

数据流聚类：

将传统的聚类算法应用于一次通过的数据集是很困难的。基于整体数据集分割的k-means算法是为了创建在整体数据流之上的聚类。在数据流的环境下，我们更希望在用户自定义的范围内来确定聚类而不是整体的数据集。第2章中讨论了决定整体数据集的小型聚类技术。同时介绍了一些小型聚类的应用：扩展数据流的分类问题、主观的数据挖掘问题像隐私数据的挖掘或查询估计。

数据流分类：

第3章讨论了数据流分类算法。一些是仅仅是传统分类算法的一次通过的适应，另一些在考虑潜在数据流的演变上更加有效。

频繁模式挖掘：

在数据流的情况下，人们可能希望在滑动窗口以及整体数据流的频率项目集。

数据流中的变化探测：

数据流的模式可能会随时间演变，在一些情况下可能希望追踪并分析这些随时间改变的的改变的本质。 同时，数据流的演变可能会影响潜在数据挖掘算法的行为，因从结果将会随着时间的变化而失效。

多维度流的流立体分析：

许多数据流存在于多维度的空间中并且有相当低程度的抽象，很多分析师致力于相对高等级的维度结合的动态改变。为了发现高等级的动态以及演变特性，人们需要执行流数据的多等级、多维度的在线分析过程（OLAP）。对于新架构的调查是很有必要的，这可能便利了多维度数据流的在线分析进程。

第六章中，一个有效执行多维度流数据的在线局部聚集的有趣的流立体架构捕获了数据流的必要的动态以及演变特性。这同时也形成了在线流挖掘的初步架构。

数据流中的负载均衡：

由于数据流是由跟流处理应用无关的进程所聚集的，要控制进入流的速率是不可能的。系统具有快速适应多种进入流处理速率的能力是必要的。第7章讨论了一个特殊了自适应类型：当系统的需求空间不能被全部可用资源所满足时，通过降低负载（通过丢弃未处理的元组）来温和地降低性能的能力。为了专注于聚集查询，章节提出了一个算法：决定在一个查询计划的哪个点上降低负载被执行以及为了最小化给查询结果带来的的不确定性每个节点所降低的负载的数量。

数据流中的滑动窗口计算：

大部分的结构摘要讨论使用整体数据流来构建相应的结构摘要。计算的滑动窗口模型被一个假定所激励，该假定是：在数据流计算中使用最近的数据是很重要的。第8章使这种计算模型具有了一定的形式同时回答了在滑动窗口模型下为了解决特定的问题多少空间以及计算时间是被需要的问题。

数据流中的结构摘要：