数据挖掘 = 业务知识 + 自然语言处理技术（ NLP，分析文字 ） + 计算机视觉技术（ CV，分析图片 ） + 机器学习 / [**深度学习**](http://lib.csdn.net/base/deeplearning)（ ML/DL）

**需求调研**、**数据挖掘阶段**和**模型发布**

游戏内的数据挖掘主要是建立用户流失率分析模型（找当前不合理的地方），对未来可能出现的流失做一个预警，除此之外，还此分析过程可能对于游戏相关的运营活动提供临时性的指导意见。就是分析玩家的倾向，数值的平衡，系统和内容的合理性，销售的倾向性和推断出一些我们难以了解到的原因。

**WEKA**作为一个公开的机器学习、数据挖掘工作平台，集合了大量能承担数据挖掘任务的机器学习算法，包括对数据进行预处理，分类，回归、聚类、[关联规则](https://baike.baidu.com/item/%E5%85%B3%E8%81%94%E8%A7%84%E5%88%99)以及在新的交互式界面上的可视化。如果想自己实现数据挖掘算法的话，可以参考weka的接口文档。在weka中集成自己的算法甚至借鉴它的方法自己实现可视化工具并不是件很困难的事情。

将中文内容先翻译成英文然后获取关键词，这种想法倒也是神奇。

数据挖掘的关键在于分类以及聚类，如果是网咯数据挖掘还需加上“爬虫”（对IP的范围定位，需要多台服务器使用多种技术混合定位）来获取数据。