

数据挖掘及其应用心得

16337341 朱志儒

数据的爆炸性增长：从 TB 到 PB；数据收集和数据可用性：自动的数据收集工具，数据库系统，Web，计算化的社会；丰富数据的主要来源：商业：Web，电子商务，事务，股票，…；科学：遥感，生物信息学，科学仿真，…；社会：新闻，数码相机，…。我们正淹没在数据中，但却渴望知识！“需要是发明之母”——数据挖掘——自动化的大规模数据集分析。

数据库技术的演化：

1960s: 数据收集，数据库创建，信息管理系统 IMS 和数据库管理 DBMS；

1970s: 关系数据模型，关系 DBMS 实现；

1980s: RDBMS，高级数据模型（扩展关系的，OO，演绎等。）面向应用的 DBMS（空间，科学，工程等。）；

1990s: 数据挖掘，数据仓库，多媒体数据库，和 Web 数据库

2000s: 流数据管理和挖掘；数据挖掘和它的应用 Web 技术 (XML, 数据集成) 和全球信息系统。

数据分析和决策支持：市场分析和决策支持：目标市场定位，客户关系管理 (CRM)，购物篮分析，交叉销售；风险分析和管理：预测，客户保持，质量控制，竞争分析；欺诈检测和不寻常模式的检测（离群点）。

其他的应用：文本挖掘（新闻组，email，文档）和 Web 挖掘流；数据挖掘；生物信息学和生物数据分析。

数据来源：信用卡交易事务，会员卡，优惠券，客户投诉电话，公众生活方式研究。

目标市场：寻找“榜样”客户的聚类，他们共享相同的特征：兴趣，收入水平，消费习惯等；确定客户在一段时间的购买模式。

交叉市场分析——寻找产品销售之间的关联/相关性，以及基于这些关联进行预测。客户轮廓——什么类型的客户买什么产品（聚类或分类）。

客户需求分析：为不同的用户识别最好的产品；预测什么因素将吸引新的客户。

摘要信息提供：多维摘要信息报告；统计学总结信息（数据中心的趋势和变化）。

财经计划和资产评估：现金流分析和预测；资源计划：总结比较资源和开销；竞争：监控竞争对手和市场方向；细分客户类别，制定基于类别的定价过程；在激烈竞争市场中建立价格策略。