

基于电子商务中的数据挖掘技术研究计算机等级考试 PDF 转换可能丢失图片或格式，建议阅读原文

https://www.100test.com/kao_ti2020/644/2021_2022__E5_9F_BA_E4_BA_8E_E7_94_B5_E5_c98_644631.htm

1 引言 国际互联网的日益普及，各种形式的信息大量地产生和收集导致了信息爆炸。现代社会的竞争趋势要求对这些信息进行实时的和深层次的分析，虽然现在有了更强大的信息存储和检索系统，但是用户在分析和使用所拥有的信息方面却变得越来越困难。如何对大量信息进行有效组织利用，使用户能够从大量繁杂的信息中找出真正有价值的信息和知识，帮助企业制定更好的营销策略。信息处理技术有了新的应用研究课题数据挖掘。主要讨论采用数据挖掘技术在新兴的电子商务领域的应用研究。

2 电子商务与数据挖掘简介 电子商务是指个人或企业通过Internet网络，采用数字化电子方式进行商务数据交换和开展商务业务活动。目前国内已有网上商情广告、电子票据交换、网上订购、网上银行、网上支付结算等多种类型的电子商务形式。电子商务正以其成本低廉、方便、快捷、安全、可靠、不受时间和空间的限制等突出优点而逐步在全球流行。服务范围可归类为：商业—商业(B2B)，商业—消费者(B2C)，商业—政府(B2G)。数据挖掘是通过挖掘数据仓库中存储的大量数据，从中发现有意义的新的关联模式和趋势的过程。从商业的角度定义，数据挖掘是一种新的商业信息处理技术，其主要特点是对商业数据库中的大量业务数据进行抽取、转换、分析和其他模型化处理，从中提取辅助商业决策的关键性数据。数据挖掘最吸引人的地方是它能建立预测模型而不是回顾型的模型。利用功能强大的数据挖掘

技术，可以使企业把数据转化为有用的信息帮助决策，从而在市场竞争中获得优势地位。数据挖掘与传统的数据分析的不同是在没有明确假设的前提下去挖掘信息、发现知识。数据挖掘所得到的信息应具有先前未知、有效和可实用3个特征。

3 数据挖掘在电子商务中的作用

数据挖掘技术之所以可以服务电子商务，是因为它能够挖掘出活动过程中的潜在信息以指导电子商务营销活动。在电子商务中其作用有4个方面：

- (1)挖掘客户活动规律，针对性的在电子商务平台下以提供“个性化”的服务。
- (2)可以在浏览电子商务网站的访问者中挖掘出潜在的客户。
- (3)优化电子商务网站巾的信息导航，方便客户浏览。
- (4)通过电子商务访问者的活动信息的挖掘，可以更加深入的了解客户需求。

4 数据挖掘过程

电子商务中的数据挖掘的过程一般由3个主要的阶段组成：数据准备、数据挖掘、结果解释和评价。

- (1)数据准备数据准备又可分为数据选取、数据预处理。数据选取的目的是确定发现任务的操作对象。即目标数据，是根据用户的需要从原始数据库中抽取的一组数据。数据预处理一般包括消除噪声、推导计算缺值数据、消除重复记录、完成数据类型转换(如把连续型数据转换为离散型数据，以便于符号归纳；或是把离散性数据转换为连续型数据，以便于神经网络计算)以及对数据降维(即从初始特征中找出真正有用的特征以减少数据挖掘要考虑的变量个数)。
- (2)数据挖掘阶段首先要确定数据挖掘的目标和挖掘的知识类型。确定挖掘任务后，根据挖掘的知识类型选择合适的挖掘算法；最后实施数据挖掘操作，运用选定的挖掘算法从数据库中抽取所需的知识。
- (3)结果的解释和评价数据挖掘阶段发现的知识，经过评估，可能存在冗余或无关的知识，

这时需要将其剔除：也有可能知识不满足用户的要求，需要重复上述挖掘过程重新进行挖掘。另外，由于数据挖掘最终要面临用户，因此，还需要对所挖掘的知识进行解释，以一种用户易于理解的方式(如可视化方式)供用户所用。可以看出，以上整个数据挖掘过程是不断地循环和反复的，因而可以对所挖掘出来的知识不断求精和深化。最终达到用户所满意的结果。

5 电子商务中数据挖掘的技术与方法

数据挖掘源于人工智能。它利用人工智能中成熟的技术和方法对经过处理的数据进行分析，其利用的技术方法越多，所得到的信息也就越精确。在电子商务中主要应用的数据挖掘技术和方法有：

- (1)聚类分析聚类分析可在电子商务过程中从Web查找信息中聚集出具有相似特性的客户。划分出客户群后，能够帮助企业开发和执行其市场策略，Web可根据客户群提供特定的服务。
- (2)关联规则关联规则挖掘是发现大量数据库中集之间的关联关系，这些关联关系可以帮助许多商务决策的制定。
- (3)分类规则分类是找出一个类别的概念描述，它代表了这类数据的整体信息，一般用规则或决策树模式表示。其目的是把新的记录分配到预先定义好的类中。
- (4)时间序列模式的发现按时间顺序查看时间事件数据库，从中找出另一个或多个相似的时序事件，通过时间序列搜索出重复发生概率较高的模式。发现序列模式便于进行电子商务组织预测客户的查找模式，从而对客户进行针对性的服务。在时序模式中，一个重要影响的方法是相似时序。
- (5)偏差分析描述和分析数据的异常或极端现象，包括不规则数据、反常实例和观测结果与期望值的偏离等。主要用于分析客户异常行为、信用欺诈甄别和数据质量控制。以及网络安全管理和故障检测等领域

。(6)预测与评价对历史数据进行综合分析和归纳，推理出数据分布的时效性和规律性，从而对未来事件发展的趋势和结果进行预测和评估。100Test 下载频道开通，各类考试题目直接下载。详细请访问 www.100test.com