

Oracle9i 数据挖掘

数据表

2001年6月

Oracle9i Data Mining™ 使公司能够建立集成的商务智能应用。利用内嵌于 Oracle9i 数据库的数据挖掘功能，应用程序开发人员可以使商务智能的提取和分布自动完成，并把结果集成到其它商务应用程序中。



Oracle9i 数据挖掘

Oracle9i Data Mining 可使公司建立高级的商务智能应用，这些程序能够挖掘公司的数据库、洞察新的信息并把这些信息整合到商务应用中。Oracle9i Data Mining 是 Oracle9i 数据库企业版（EE）的附有定价的一个选项，该版本为数据库分类、预测及关联内嵌了数据挖掘功能。通过基于 Java 的 API 可以访问所有的模型建立和得分功能。

Oracle9i Data Mining 帮助公司建立商务智能应用，这些程序可以查找公司数据中有意义的模式和关联，这些模式可以帮助您更好地理解 and 预测客户行为。利用数据挖掘 Data Mining，您可以制订策略来：

- 防止客户消耗
- 向现有客户交叉销售
- 获取新客户
- 侦测欺骗行为
- 识别最可赢利的客户
- 更精确地描述客户

利用 Oracle9i 数据挖掘技术，公司可以提取他们的数据库中的隐含信息，从而对公司的客户及公司的商务活动达到更深的了解。通过其它查询、分析、制表工具和应

解。在与客户合作期间的每一个阶段，Oracle9i Data Mining 提供的价值可以直接达您的底线。

Oracle9i Data Mining 在科学、政府、制造、医疗以及其它应用领域里也可检测到隐含的数据模式，如：

- 预测生产部件的质量
- 查找病人、药物及医疗效果之间的关联
- 识别可能的网络入侵

Oracle9i Data Mining 技术发现的内在信息具有启迪作用，是非常重要的，其价值无可估量。

内嵌于 Oracle9i 数据库的数据挖掘

Oracle9i Data Mining 简化了从大量数据中提取商务智能信息的过程。它避免了把海量数据卸载到外部专用分析服务器来做数据挖掘和得分（data mining & scoring）。通过 Oracle9i Data Mining，所有的数据挖掘功能都内嵌到了 Oracle9i 数据库；这样，数据、数据准备、模型建立以及模型得分活动都保留在数据库内进行。

Oracle9i 的可伸缩性可使 Oracle9i Data Mining 分析大量数据以侦测其中的微妙模式和关系，并提取更多的商务智能信息。用，Oracle9i Data Mining 的新的洞察力和

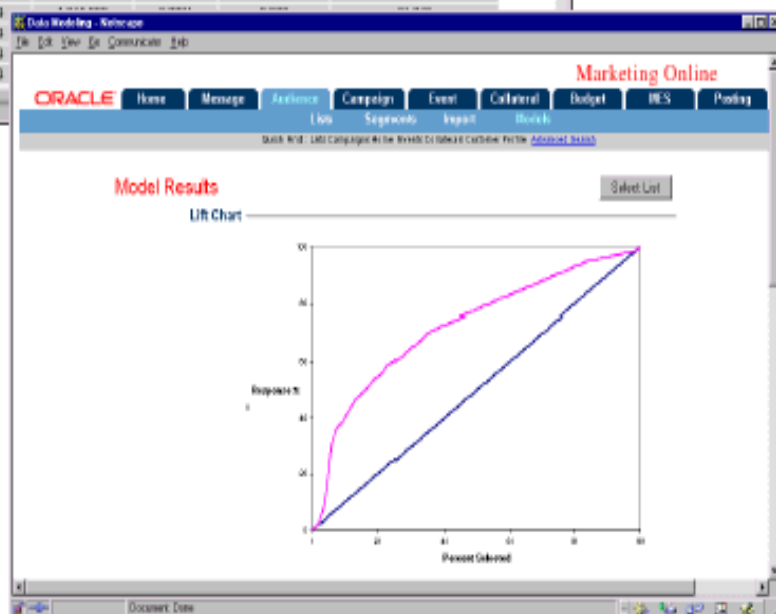
预测功能可供访问。这可使企业建立起由数据挖掘结果驱动的应用。由于 Oracle9i 数据库具有无可匹敌的性能和伸缩性，因而 Oracle9i Data Mining 为建立高级商务智能应用提供了非常理想的基础构架。利用 Oracle9i Data Mining，数据从来不必脱离数据库，所以大批量得分运算也是切实可

行的。公司可以对大型数据表进行得分，而不必把数据提取到外部指定的数据挖掘服务器来进行挖掘和得分。

通过新的商务智能信息的自动提取和分布，Oracle9i Data Mining 提供了可以直接转化为高效益低成本的结果。



数据挖掘的洞察力可以整合到其它应用中，如 Oracle CRM /Oracle Marketing Online 活动管理应用软件。



通过预测与洞察力增强应用

Oracle9i Data Mining 使公司能够在其经营范围之内，系统化地提取和集成新的商务智能信息。应用开发人员可以使用 Oracle9i Data Mining 的基于 Java 的 API(应用编程界面) 来增加数据挖掘的洞察和预测功能，从而增强商务应用功能，如客户关系管理 (CRM)、企业资源计划 (ERP)、Web 入口以及无线应用等。例如，电信公司可以使用 Oracle9i Data Mining 来建立转网应用(churn applications)，以此，在客户转而投向公司的竞争者之前，来识别出这些可能转网的客户。Oracle9i Data Mining 的预测功能在互惠互利的一对一关系中，可用来预测客户行为，并前瞻地管理客户。

零售商和数据库营销商可以使用 Oracle9i Data Mining 来建立营销活动应用，其目标是那些对报价最可能作出响应的潜在客户。Oracle9i Data Mining 能够把数据挖掘结果整合到应用中。这样的例子包括预测客户转网 (churn) 的行为、对特定报价作出响应、成为可赢利的客户、提出一项索赔、或花费大量金钱等可能性。利用 Oracle9i Data Mining，电子商务和 Web 网点可以加强 Web 搜索，提供有关的或“关联的”在用的或内容方面的其它文档或项目。

一旦对数据进行了挖掘和建立了预测模型，Oracle9i Data Mining 就可以利用该模型给其它数据“得分”以便作出预测。给数据得分是在数据库中发生的，分数随后可供其它应用使用。存储于数据库中的数据挖掘模型可对交互应用 (如呼叫中心)

的需求提供洞察和预测，从而提出“推荐做法”。例如，呼叫中心应用可以使用某个客户的历史数据以及来自一次呼叫进程中的响应，对该客户的喜好作出评估，并制作出个性化的交叉销售推荐做法。

预测与分类

Oracle9i Data Mining 提供了 Naive Bayes 数据挖掘算法来进行预测和分类。该算法适用于各种数据挖掘问题，同时也提供了高度准确性。通过查找数据中存在的模式，公司可以用相似的特征——即用过过去作为未来的预测算子——对客户未来的行为作出预测。典型的预测应用可以估计某个结果的可能性，如“0, 1”或“yes, no”或“A, B, C 或 D”。请考虑下面的例子：

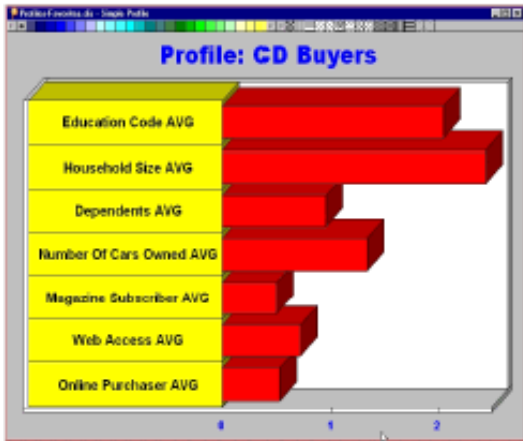
问：这个客户对我的特别报价会作出响应吗？

答：是的，客户响应的可能性是 92%。

客户 ID	响 应	概 率
4567332	YES	0.92
5463433	YES	0.92
1123444	YES	0.85
9565333	YES	0.85
2346557	YES	0.72
2257673	YES	0.55

Oracle9i Data Mining 的预测模型返回了预测的结果及其概率，这样，公司就能够前瞻地管理他们的商务。

预测模型的结果可以组合，从而提供有价值的商务智能。例如，Oracle9i Data Mining 能够建立一个客户的合作期间值 (LTV) 的模型，同时还可建立一个预测客户转网 (Churn) 的可能性的模型。把这两个预测结果的概率值相乘 (P (LTV) × P (Churn))，即可提供关于如何花费营销预算的有价值的洞察。



可以利用其它软件和应用（如此处所示的 Oracle Discoverer）检测 Oracle9i Data Mining 的预测和分类。

发现关联

Oracle9i Data Mining 提供了关联规则（Association Rules）数据挖掘算法来检测隐含在数据库里的“关联的”或同时发生的事件。对于与客户相关的产品，关联分析经常用来查找其受欢迎的相关产品，例如与“香蕉”相关的“牛奶”和“麦片”。关联还可以用来识别同时发生的项目和事件，如：

- 什么生产部件和设备装置与故障事件存在关联？
- 什么样的病人和药物属性与哪种治疗效果相关联？
- 这位客户最可能购买或最喜爱哪些物品或产品？

关联可用来预测放入购物筐的下一个物品，购物筐有助于满足客户需求和增加平均订购值。

基于 Java 的 API

应用开发人员通过基于 Java 的 API 来访问 Oracle9i Data Mining 的功能。所有数据挖掘功能的程序控制使数据准备、模型建立以及模型得分操作都实现了自动化处理。

在 Sun 公司的 Java Community Process 作为 Java 规范要求（JSR）出现之后，Java 数据挖掘（JDM）也随之成为数据挖掘的新标准。Oracle、Sun、IBM 以及其它许多公司都意识到需要一个基于 Java 的标准来规范和使用数据挖掘技术，这些公司都参与了 JDM。JDM 利用了几个已逐渐形成的标准，其中包括对象管理组（Object Management Group）的公用仓库元数据（CWM — Common Warehouse Metadata）、数据挖掘组（Data Mining Group）的预测性挖掘标记语言（PMML — Predictive Mining Markup Language）以及国际标准化组织（International Standards Organization）的用于数据挖掘的 SQL/MM。

对于建议使用的 JDM 概念和方法，Oracle9i Data Mining 的 API 提供了一个先期预览。Oracle9i Data Mining 在 JDM 发布之后，最终将遵守该标准。

Rule Number	Antecedent	Support	Probability
1	6 Ham	2,791	62%
2	6 Pie Crust	2,791	62%
3	10 Carmel Cubes	429	92%
4	10 Pop Sticks	429	92%
5	8 Oranges	1,592	95%
6	8 Grapes	1,592	95%
7	8 Bananas	1,592	95%

Oracle Discoverer 显示了 Oracle9i Data Mining 的关联分析结果

技术规范

数据准备

- 数据采集（离散化）工具，采集界限由用户自定义。
根据分位点和头 N 点采集数据的自动采集方法。
- 聚合单个记录格式或事务格式的数据。

模型评估

- 含混矩阵
- 二进制分类模型的提升表(lift tables)

数据挖掘算法

- 为所有 Oracle9i Data Mining 操作提供的缺省设置。
覆盖和规定设置的能力。

分类和预测的 Naive Bayes 算法

- 二进制结果的预测。（例如，yes 或 no）
- 多个可能结果中的最可能结果的预测（分类）。
（例如：假设 A、B 或 C 是可能的结果，而 B 发生的置信度为 60%）

识别同时发生的项目或事件的关联规则

- 找出事件和同时发生事件的可能性。
例如，Q、R 和 S 与 Z 有 n 次关联，其置信度为 m%）

平台需求

Oracle9i Data Mining 运行于所有支持平台上的 Oracle9i 数据库。
对大型数据挖掘问题推荐使用 Oracle9i Partitioning。

ORACLE®

Oracle 公司
世界总部
500 Oracle Parkway
Redwood Shores, CA 94065
U.S.A.
全球咨询:
电话: +1.650.506.7000
传真: +1.650.506.7200
<http://www.oracle.com/>
版权所有 © 2001 Oracle 公司
保留所有权利。

本档提供的內容仅用作参考信息。此处内容如有更改，恕不另行通知。若有问题，请联系 Oracle 公司。
Oracle 公司对此并没有提供任何保证，同时也没有与该文档相关联的义务。

Oracle 是已注册商标，Oracle9i 和 Oracle9i Data Mining 也是 Oracle 公司的商标。

文中所提到的所有其它公司和产品名称均只是用作鉴别目的，它们均可能为相应所有者的商标。
