

决策支持系统中客户价值预测模型的设计与分析

顾美芳, 陈建明

(苏州大学 智能信息处理研究所, 江苏 苏州 215006)

摘要:在 Internet 时代, 企业可以从外界获取大量的客户数据资料, 如何在这样丰富的客户数据资源中寻找企业的高价值客户? 基于这样的需求, 文中讨论了使用微软的 VS. Net 平台和 SQL Server Analysis Services 中的 DSO 组件来建立模型的方法, 通过对原有客户资料的分析建立客户价值预测模型, 并由此模型进行新客户和潜在客户的价值预测。此模型在客户价值预测方面有一定的参考价值。

关键词:数据挖掘; 客户预测模型; DSO 组件

中图分类号: TP311.1

文献标识码: A

文章编号: 1673-629X(2006)07-0052-03

Design and Analysis of Customer Predictability Model in DSS

GU Mei-fang, CHEN Jian-ming

(Institute of Computer Science and Technology, Soochow University, Suzhou 215006, China)

Abstract: In Internet times, companies can obtain much information about customers and want to seek higher valuable customers from it. To meet this need, here discuss a method to build a predict model based on VS. Net platform and DSO component of SQL server analysis services. First establish the customer value prediction model by original information. Then extract the value customer between new and potential customers. Thus this model is helpful to customer value prediction.

Key words: data mining; customer predictability model; DSO component

0 引言

在 Internet 时代, 企业可以从外界获取大量的客户数据资料, 如何利用这样丰富的客户数据资源为企业创造更多的利润, 这是所有的企业高层都希望的结果。通过对已有客户价值的挖掘, 建立客户价值预测模型, 将此模型运用于外界大量的客户资料, 预测潜在价值客户, 使其成为企业的新客户, 为企业创造更多的利润^[1]。

数据挖掘就是从大量的数据中自动获取有用信息和知识的过程, 针对零售业中已经存在的客户数据资料进行分析, 利用数据挖掘算法, 建立客户价值预测模型, 发掘不同特征的客户的不同价值, 针对新的客户数据资料进行预测, 发掘潜在盈利客户, 使其成为可以为企业创造利润的价值客户^[2]。

1 数据挖掘的工具及体系结构

随着数据挖掘应用的推进, 目前市场上已出现一些比较成熟的数据挖掘工具, 例如 SAS, SPSS, IBM Miner, 微软的 MS SQL Server Analysis Services 等。SAS, SPSS,

IBM Miner 等数据挖掘工具比较适应于企业型, 需要高处理能力、高网络容量, 这类数据挖掘工具的特点是它通常提供了多种数据挖掘算法, 并有能力解决多种应用问题。而微软的 MS SQL Server Analysis Services 作为直接应用在商业的数据挖掘工具来帮助项目实施, 既节省了大量的开发费用, 缩短开发周期, 又可以节约维护和升级的开销, 所以笔者选择 MS SQL Server Analysis Services 作为开发工具。

MS SQL Server Analysis Services 是集成在 SQL Server 2000 中的一套 OLAP(在线分析技术)系统, 它提供的 DSO(决策支持对象)是一种组件对象模型(COM)的类库及其接口, 这些类库及其接口能够访问 Analysis Services 引擎的核心部分。客户端可用编程的方式通过 DSO 来创建立方体、数据库和数据挖掘模型, 并管理服务器。而 DSO 是以映射 Analysis Services 中功能和组件的对象以及集合的某个层次等级形式来实现的。这就允许应用程序与底层引擎简单集成, 无需理解其内部的复杂性^[3]。

图 1 给出了通过 Analysis Services 的 DSO(决策支持对象)组件来创建 OLAP 多维数据集和客户模型的系统体系结构^[4]。

2 客户价值预测模型的设计与分析

2.1 模型的设计

数据挖掘的一般步骤包括开发准备、数据处理、建立

收稿日期: 2005-11-02

基金项目: 苏州市科技攻关基金资助项目(SG0406)

作者简介: 顾美芳(1980-), 女, 江苏苏州人, 硕士研究生, 研究方向为智能化信息处理; 陈建明, 副教授, 研究方向为智能化信息处理、软件工程。

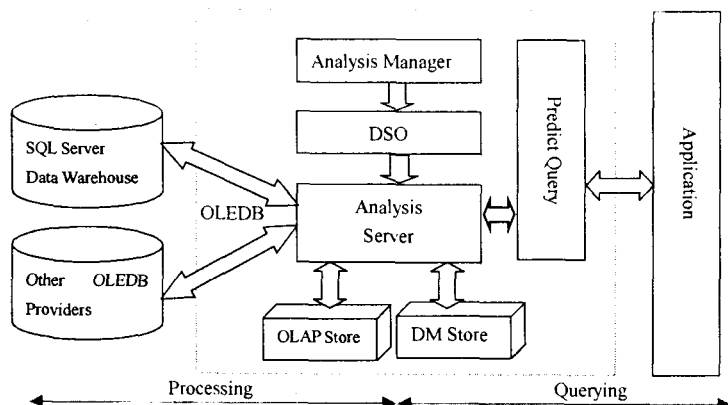


图1 基于 Analysis Services 决策分析系统的体系结构
数据仓库、模型建立、模型的评估和实施等过程。通过使用微软的工具 MS SQL Server Analysis Services 和 C#.NET 作为平台,建立客户价值预测模型的整个过程框架,如图2所示。

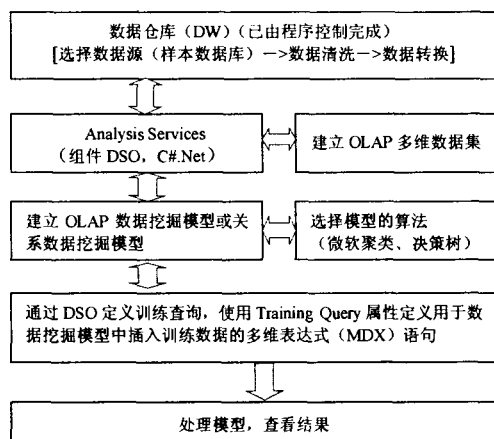


图2 建立挖掘模型的流程

根据图2的建模框架,建立客户预测模型的整个过程由客户条件及预测对象的确定和客户预测模型的建立两个步骤进行。

(1) 客户条件及预测对象的确定。

①分析客户的基本资料,将客户的年龄、职业、教育程度、薪水、婚姻状态等属性作为客户价值预测的前提输入条件。

②将客户拥有的会员卡等级作为预测对象,通过决策树算法来预测不同特征的客户的价值。通常认为拥有金卡的客户是消费金额最高的客户群体,即贵宾客户,其次是银卡客户。

(2) 客户预测模型的建立。

用 DSO 和 C#.NET 建立客户预测模型需要如下 6 个主要步骤。

①初始化 DSO 中将要使用到的对象:

```
DSO.ServerClass dsoServer = new DSO.ServerClass();
DSO.DatabaseClass dsoDb = new DSO.DatabaseClass();
DSO.CubeClass dsoCube = new DSO.CubeClass();
DSO.DataSourceClass dsoDs = new DSO.DataSourceClass();
DSO.MDStoreClass dsoMd = new DSO.MDStoreClass();
```

②连接一个 Analysis Services 服务器,并得到一个指向希望创建新的挖掘模型的数据库的对象指针:

```
dsoServer.Connect("webserver1");
dsoServer.Update();
dsoDb = (DSO.DatabaseClass)dsoServer.MDStores.Item("customer");
dsoDs = (DSO.DataSourceClass)dsoDb.DataSources.Item("cs");
```

其中 Analysis Services 服务器名为“webserver1”,新建的挖掘模型的数据库名为“customer”,数据源为“cs”。

③创建 Mining Model 对象:

```
DSO.MiningModel mineModel = new DSO.MiningModelClass();
MineModel = (DSO.MiningModelClass)dsoServer.CreateObject(
DSO.ClassTypes.clsMiningModel, DSO.SubClassTypes.sbclsOlap);
```

④建立模型 model:

```
If (dsoDb.MiningModels.Find("sortmodel"))
{dsoDb.MiningModels.Remove("sortmodel");}
dsoDb.MiningModels.AddNew("sortmodel", DSO.SubClassTypes.sbclsOlap);
MineModel = (DSO.MiningModel)dsoDb.MiningModels.Item("sortmodel");
mineModel.DataSources.AddNew("cs", DSO.SubClassTypes.sbclsRegular);
```

⑤选择微软决策树算法 Microsoft Decision Trees 和分析所需的输入列和预测列,选择算法:

```
String algorithm = "Microsoft Decision Trees";
mineModel.set_MiningAlgorithm(ref algorithm);
```

添加预测所需的输入列:

```
myModel.Columns.AddNew("psalary", DSO.SubClassTypes.sbclsRegular);
DSO.Column dsoCol2 = (DSO.Column)myModel.Columns.Item("psalary");
string salary = "p.member.pmb.salary";
dsoCol2.set_Content(ref dsoCol2);dsoCol2.set_SourceColumn(ref salary);dsoCol2.set_DataType(ref tc);dsoCol2.set_IsInput(ref t);dsoCol2.set_IsDisabled(ref f);dsoCol2.set_IsPredictable(ref f);
```

客户的会员卡等级作为预测列时,只需将预测列的参数设为 true 即可。

⑥使用 Mining Model.Update 保存模型,并用 Mining Model.Process 方法培训模型。使用 Mining Model.Lock-Object 方法锁定挖掘模型,防止在培训时被更新。

2.2 客户预测模型的结果及预测分析

(1)首先将客户的薪水等级分为 A,B,C,D,E,F 六类。

A 表示年薪在 3w~5w 之间的客户;B 表示年薪在 5w~7w 之间的客户;C 表示年薪 7w~10w 之间的客户;D 表示年薪在 10w 以上的客户。

将客户的婚姻状况表示为 M, S 两种。M 表示已婚, S 表示未婚。

会员卡等级分为金卡、银卡、铜卡 3 种等级。

(2) 模型结果分析如表 1 所示(下面只列举了部分数据)。

表 1 模型结果分析

条件 预测	条件 1	预测可能性 (所占比率%)	条件 2	预测可能性 (所占比率%)
金卡	A&M	22.32	A&S	3.88
	D&M	83.33	D&S	3.33
银卡	A&M	10.13	A&S	9.39
	D&M	9.72	D&S	86.67
铜卡	A&M	62.26	A&S	81.45
	D&M	4.17	D&S	6.67

客户分类预测模型的结果分析如表 1 所示:按照客户的薪水和婚姻状态作为预测的输入条件,将客户的会员卡等级作为预测的对象。例如:条件 D&M 表示年薪在 10w 以上的已婚客户,可能成为金卡客户的概率为 83.33%;可能成为银卡客户的概率为 9.72%。

(3) 模型的预测分析。

对新客户资料的数据进行分析,得出的预测分析结果如表 2 所示。

表 2 预测分析结果

年薪	婚姻状况	预测等级	概率	最大支持数
7~9 万元	已婚	铜卡	0.708745874	858
10 万以上	已婚	银卡	0.473684210	107

这个模型的预测结果为潜在价值客户的挖掘和新客户的获取方面提供了参考。例如:假设商家有很多新客户资料,但如何使商家有目的地去拓展新的客户群体?如何发现价值客户?那么此预测模型就是一个很好的参考。通过此模型,可以对新客户资料的自然属性进行分析,预测客户的价值,发掘那些最可能成为商家盈利的价

(上接第 51 页)

4 结束语

实践证明,通过构造算法演示引擎 AAE,可以实现交互式的动态演示。该算法演示引擎已在本工作室开发的算法演示软件中应用,取得良好的效果。但是,在现阶段,AAE 的智能性还有待提高。针对 AAE 的不足之处,未来的工作就是实现一个通用的算法演示引擎。通用的算法演示引擎可以自动地生成解释方法,这些解释方法可以对具有一定规范的伪码语言进行解释。这样,只需要实现算法演示引擎提供的接口,就可以实现动态演示。构造算法演示引擎的思想不仅为实现算法的动态演示提供了思路,

值客户。商家针对这些客户,可以采取一定的措施(如短信、赠品、积分等方式)来吸引客户,使他成为商家真正的价值客户^[5]。

这种建模的方法同样也适用于信贷风险投资中。假设客户的属性特征为输入列,以客户的信用风险度为预测对象建立客户风险预测模型,当有新的客户要申请信贷的时候,就可以使用预测模型,对客户进行信用风险度的预测,从而降低信贷风险。

3 结束语

文中主要讨论了利用微软的 MS Analysis Server 分析工具、.NET 开发平台和微软的决策树算法,通过对客户自然属性的分析,以会员卡等级作为预测对象来建立客户分类预测模型过程和方法,特别是通过对模型训练所得出的结果在客户的潜在价值预测和新客户的发掘方面具有一定的实用价值。

微软的工具在商业项目的开发方面与其他工具相比具有操作方便、开发周期短等优点。不足的是微软的分析工具仅提供了决策树算法和聚类二种数据分析算法,使其应用面受到一定的限制。但对于设计一般要求的分析系统而言,它不失为一种高效、快捷的好方法。

参考文献:

- [1] 张学军,吴 潇.CRM 实施宝典[M].北京:国防工业出版社,2005.
- [2] 贾 琳,李 明.基于数据挖掘的电信客户流失模型的建立与实现[J].计算机工程与应用,2004(4):185-187.
- [3] Gunderloy M. SQL Server 开发指南(联机分析处理)[M].张 伟,宋 霞译.北京:电子工业出版社,2001.
- [4] Seidman C. SQL Server 2000 数据挖掘技术指南[M].刘 艺,王鲁军等译.北京:机械工业出版社,2002.
- [5] 史威福特 R S. 客户关系管理[M].杨东龙等译.北京:中国经济出版社,2001.

而且也对实现其他动态演示系统起一定的启发作用。

参考文献:

- [1] 吴伟民.数据结构和算法的可视化教学研究与实践[J].现代计算机,1999(3):35-37.
- [2] 林桂伍,黄 峰,陈峥杉,等.可视化环境下 CAI 课件系统设计与实现[J].计算机应用研究,1999(4):106-107.
- [3] 高 翔.用 Java 多线程实现数据结构算法动态演示[J].北京联合大学学报,2002,16(2):63-66.
- [4] 吴志刚,韩瑶香,苏安婕.《数据结构》多媒体 CAI 系统的研制[J].郑州纺织工学院学报,2001,12(4):46-49.
- [5] 彭玉青,肖国玺,杨 昕.数据结构算法动态演示 CAI 软件的实现[J].河北工业大学学报,2000,15(1):1-4.