



泰迪智能科技

服务热线：40068-40020

顶尖数据挖掘教学实训平台 (TipDM-T6)

产 品 说 明 书

广州泰迪智能科技有限公司 版权所有

地址： 广州市经济技术开发区科学城 232 号

网址： <http://www.tipdm.com>

邮箱： services@tipdm.com

热线： 40068-40020

企业 QQ： 40068-40020

邮编： 510663

电话： （020） 82039399

目 录

1 产品简介	3
2 套件组成	3
2.1 建模软件	4
2.1.1 软件功能.....	4
2.1.2 软件界面.....	5
2.1.3 软件优势.....	7
2.2 实训教程	7
2.3 教学案例库及配套资源.....	10
2.4 教学沙盘	12
2.5 课程规划及师资培训.....	13
2.6 学生实习实训.....	14
3 产品优势	14
4 应用案例	14
4.1 主要客户	15
4.2 典型示例	15

1 产品简介

随着企业的泛互联网化和企业信息化程度的不断提高，大数据的概念被提了出来，在大数据时代，数据对企业突显出越来越重要的价值，企业对数据的价值也越来越重视。目前国内不少高校将数据挖掘引入本科教学中，在数学、电子信息、金融等专业开设了数据挖掘技术相关的课程，但目前这一课程的教学仍主要限于理论介绍。以数学类专业为例，目前高校存在的困境主要有：

- ✧ 师资队伍建设问题：缺少统计学及应用数学专业的双师型教师；
- ✧ 师范院校的统计学及应用数学专业在大数据时代该如何准确定位？
- ✧ 课程设置该怎样跟进与完善？
- ✧ 实践环节应该如何加强？
- ✧ 产学研以及院校合作怎样才能落到实处？

顶尖数据挖掘教学实训平台（TipDM-T6，简称T6）是专门针对高校数据挖掘课程中，学员实际应用能力差这一教学弱点设计开发的一系列教学模块。能够让老师在讲解数据挖掘算法理论的同时，通过实用的建模工具来模拟实际应用案例，在实验室环境下体验实际应用，由此弥补学员实践经验的不足。在教学过程中，强化应用和动手，做到理论与实践的有机结合，使学生受益，老师授课方便。

适用对象

- ✧ 有数据挖掘相关教学课程的高校、研究所和培训机构。
- ✧ 用数据挖掘进行科研的高校、研究所。
- ✧ 适用专业：统计学、数学与应用数学，以及金融、医疗、管理、电子商务等有数据挖掘教学专业。

特别说明：提供数据挖掘建模软件，同时协助指导教师构建特定专业的挖掘模型及编写相应实训指导书。

2 套件组成

顶尖数据挖掘教学实训平台（TipDM-T6，简称T6）主要包含以下内容：

核心组件：

- ✧ 数据挖掘建模平台（TipDM-TB，简称TB）：底层基于R语言、Weka和Python算法包实现的数

据挖掘建模仿真工具。

可选组件：

- ✧ 数据挖掘辅助教学实训教程：由校企联合打造的以解决企业应用为目标的案例教程，主要包括：《R语言与数据挖掘》、《Python与数据挖掘》、《MATLAB数据分析与挖掘实战》、《R语言数据分析与挖掘实战》、《Python数据分析与挖掘实战》等。
- ✧ 数据挖掘辅助教学案例库及配套资源：所有案例均来自企业在数据挖掘领域的典型应用，配套资源包括案例源数据及过程数据，R、Python、TipDM、MATLAB等环境下的程序/模型，教学PPT及微课视频等教学课件。
- ✧ 数据挖掘教学沙盘：模拟各种案例场景的演练沙盘，如菜品智能推荐沙盘、电子商务网站用户行为分析与服务推荐沙盘、窃漏电用户自动识别沙盘、智能交通仿真沙盘、管网漏损智慧水务沙盘等。
- ✧ 课程规划及师资培训：协助完成课程设置及教学规划，并为数据挖掘课程教学提供师资培训。
- ✧ 学生实习实训：提供数据挖掘实践项目，组织学生参与公司的项目实习及集中实训。

2.1 建模软件

TB建模工具基于JAVA语言开发，能从各种数据源获取数据，建立各种不同的数据挖掘模型。

2.1.1 软件功能

TipDM-TB 包含了目前主流的各类算法，方便学员寓教于学，快速理解和学习算法，并可快速构建企业级数据挖掘模型，具体功能如下：

- ✧ 平台可用于统计学分析、数据挖掘、预测分析和决策支持任务。数据分析支持工作流，包括工作流定义、流程属性配置、过程数据可视化、流程服务、进度监控；平台支持数据挖掘流程所需的主要过程，包括数据的录入、数据探索与预处理、建模应用、模型评价；支持结果多样图形化展示；
- ✧ 平台所有算法提供 API（JAR 包）和 Webservice 接口，可实现平台与第三方业务系统之间的交互，实现企业级数据挖掘应用的快速开发；
- ✧ 平台支持自定义算法，自定义算法的类型具体包括 R 算法、Weka 算法等，用户可根据自身的研究应用将自定义算法上传，并与已有算法进行对比；
- ✧ 平台支持 Excel、txt、关系型数据库，支持数据源合并；支持建模中间结果在线预览，分

析结果能导出到文件及关系型数据库；数据源管理支持字段选择、关联条件、过滤条件、字段排序、条件生成；

- ◇ 支持相关性分析、单因素方差分析、标识重复记录、主成分分析、因子分析等基础统计分析；支持中值插值法、众值插值法、回归插值法、拉格朗日插值法、牛顿插值法、方差齐性检验、独立性样本 T 检验、二项分布检验、正态性检验、卡方检验、相关性检验、精确性检验、Granger 因果检验等挖掘建模算法；
- ◇ 支持特征提取、标准化处理、正交化处理、特征选择、异常值处理、离散处理、空值处理、数学类函数、属性选择等数据预处理算法；
- ◇ 支持线性回归、逐步回归、逻辑回归、Cart 回归树、保序回归、C4.5 回归树、RPart 等高级统计分析方法；
- ◇ 支持朴素贝叶斯网络、贝叶斯信念网络、决策表、ID3 决策树、CART 决策树、C4.5 决策树、BP 神经网络、RBF 神经网络、AdaBoostM1 算法、KStar 算法、SVM 支持向量机、K-最近邻分类等分类与预测算法；
- ◇ 支持 k-Means 算法、快速 k-Means 算法、EM 最大期望算法、DBScan 密度算法、KMedoids 聚类、系统聚类等聚类算法；
- ◇ 支持 Apriori、HotSpot、时序模式关联等关联规则分析；
- ◇ 支持 GM 灰色理论、ARIMA 模型、指数平滑等时序模式。

说明：版本不断更新，更多算法详见：www.tipdm.cn

2.1.2 软件界面



TipDM-TB | 数据挖掘建模平台

1.模型管理 2.数据管理 3.数据探索 4.预处理 5.挖掘建模 6.模型评估

系统菜单

- 应用模型
- 预处理
- 基础统计
- 高级统计
- 分类与预测
 - 朴素贝叶斯网络
 - 贝叶斯信念网络
 - 决策表
 - ID3决策树
 - CART决策树
 - C4.5决策树
 - BP神经网络
 - RBF神经网络
 - AdaBoostM1算法
 - KStar算法
 - K-最近邻分类
 - SVM支持向量机
- 时序模式
- 聚类分析
- 关联分析
- 自定义算法
- 分布式算法
- 系统管理

模型管理 结构化数据管理 BP神经网络

BP神经网络分类(R语言)

训练数据

index	time	id	electricity_Grow	line_Loss	alarm	steal_Electricity
1	2014-09-06 00h	9.900667154E9	4.0	1.0	1.0	1.0
2	2014-09-20 00h	9.900639431E9	4.0	0.0	4.0	1.0
3	2014-09-17 00h	9.900585516E9	2.0	1.0	1.0	1.0
4	2014-09-14 00h	9.900531154E9	9.0	0.0	0.0	0.0
5	2014-09-13 00h	9.900461501E9	2.0	0.0	0.0	0.0
6	2014-09-22 00h	9.900412593E9	5.0	0.0	2.0	1.0
7	2014-09-20 00h	9.90036618E9	3.0	1.0	3.0	1.0
8	2014-09-19 00h	9.90032296E9	3.0	0.0	0.0	0.0
9	2014-09-09 00h	9.900254673E9	4.0	1.0	0.0	0.0
10	2014-09-18 00h	9.900196505E9	10.0	1.0	2.0	1.0
11	2014-09-16 00h	9.900145248E9	10.0	1.0	3.0	1.0
12	2014-09-06 00h	9.900137535E9	2.0	0.0	3.0	0.0
13	2014-09-09 00h	9.110103867E9	3.0	0.0	0.0	0.0
14	2014-09-21 00h	8.91010184E9	9.0	0.0	3.0	1.0
15	2014-09-11 00h	8.910101209E9	0.0	0.0	2.0	0.0

模型类型: BP神经网络分类(R语言)
模型输入:
electricity_Grow
line_Loss
alarm
steal_Electricity
样本: 开始行号 1 结束行号 50
模型参数:
隐层节点数:10
衰减系数:0.0
随机种子:1234
训练的方法: bpnet
测试样本: 开始行号 1 结束行号 3

模型信息
a 3-10-1 network with 51 weights
options were - entropy fitting
b->h1 i1->h1 i2->h1 i3->h1
-26.21 -36.12 147.52 156.01
b->h2 i1->h2 i2->h2 i3->h2
127.46 -4.67 -33.49 -30.22
b->h3 i1->h3 i2->h3 i3->h3
-36.55 37.05 6.62 29.42
b->h4 i1->h4 i2->h4 i3->h4
0.26 4.45 -12.86 36.30
b->h5 i1->h5 i2->h5 i3->h5
80.88 -28.71 -34.53 -6.05
b->h6 i1->h6 i2->h6 i3->h6
16.30 -19.85 -7.17 0.89
b->h7 i1->h7 i2->h7 i3->h7
-0.85 56.94 23.62 -102.08
b->h8 i1->h8 i2->h8 i3->h8
117.19 -26.65 -79.45 -9.50

每页 20 上一页 第1/3页 共50条 当前显示: 1-20条 下一页 1 页 Excel

导入数据 参数设置 模型训练 模型验证 模型预测 应用模型

当前位置 BP神经网络 结果导出 算法说明 操作流程

TipDM-TB | 数据挖掘建模平台

1.模型管理 2.数据管理 3.数据探索 4.预处理 5.挖掘建模 6.模型评估

系统菜单

- 应用模型
- 预处理
- 基础统计
- 高级统计
- 分类与预测
 - 朴素贝叶斯网络
 - 贝叶斯信念网络
 - 决策表
 - ID3决策树
 - CART决策树
 - C4.5决策树
 - BP神经网络
 - RBF神经网络
 - AdaBoostM1算法
 - KStar算法
 - K-最近邻分类
 - SVM支持向量机
- 时序模式
- 聚类分析
- 关联分析
- 自定义算法
- 分布式算法
- 系统管理

模型管理 结构化数据管理 BP神经网络

BP神经网络结构图

训练数据

electricity_Grow	line_Loss	alarm	steal_Electricity
4.0	1.0	1.0	1.0
4.0	0.0	4.0	1.0
2.0	1.0	1.0	1.0
9.0	0.0	0.0	0.0
2.0	0.0	0.0	0.0
5.0	0.0	2.0	1.0
3.0	1.0	3.0	1.0
3.0	0.0	0.0	0.0
4.0	1.0	0.0	0.0
10.0	1.0	2.0	1.0
10.0	1.0	3.0	1.0
2.0	0.0	3.0	0.0
3.0	0.0	0.0	0.0
9.0	0.0	3.0	1.0
0.0	0.0	2.0	0.0

模型类型: BP神经网络分类(R语言)
模型输入:
electricity_Grow
line_Loss
alarm
steal_Electricity
样本: 开始行号 1 结束行号 50
模型参数:
隐层节点数:10
衰减系数:0.0
随机种子:1234
训练的方法: bpnet
测试样本: 开始行号 1 结束行号 3

模型信息
a 3-10-1 network with 51 weights
options were - entropy fitting
b->h1 i1->h1 i2->h1 i3->h1
-26.21 -36.12 147.52 156.01
b->h2 i1->h2 i2->h2 i3->h2
127.46 -4.67 -33.49 -30.22
b->h3 i1->h3 i2->h3 i3->h3
-36.55 37.05 6.62 29.42
b->h4 i1->h4 i2->h4 i3->h4
0.26 4.45 -12.86 36.30
b->h5 i1->h5 i2->h5 i3->h5
80.88 -28.71 -34.53 -6.05
b->h6 i1->h6 i2->h6 i3->h6
16.30 -19.85 -7.17 0.89
b->h7 i1->h7 i2->h7 i3->h7
-0.85 56.94 23.62 -102.08
b->h8 i1->h8 i2->h8 i3->h8
117.19 -26.65 -79.45 -9.50

每页 20 上一页 第1/3页 共50条 当前显示: 1-20条 下一页 1 页 Excel

导入数据 参数设置 模型训练 模型验证 模型预测 应用模型

当前位置 BP神经网络 结果导出 算法说明 操作流程

scheme : 方案管理接口	展开/收起	查看列表	展开所有	Raw
workflow : workflowWebService接口	展开/收起	查看列表	展开所有	Raw
cluster : 聚类分析接口(weka)	展开/收起	查看列表	展开所有	Raw
cluster : 聚类分析接口(R语言)	展开/收起	查看列表	展开所有	Raw
classify : 分类建模, 验证, 预测接口(R语言)	展开/收起	查看列表	展开所有	Raw
preprocess : 数据预处理(weka语言)	展开/收起	查看列表	展开所有	Raw
classify : 分类建模, 验证, 预测接口(weka语言)	展开/收起	查看列表	展开所有	Raw
associations : 关联分析	展开/收起	查看列表	展开所有	Raw
mahout : Mahout算法分析接口	展开/收起	查看列表	展开所有	Raw
basestatistics : 基础统计	展开/收起	查看列表	展开所有	Raw
spark : Spark算法分析接口	展开/收起	查看列表	展开所有	Raw
timeserie : 时序预测	展开/收起	查看列表	展开所有	Raw
datavalid : 数据验证	展开/收起	查看列表	展开所有	Raw

2.1.3 软件优势

- ✧ **简单易用的可视化操作**: 与一般建模工具相比, 用户不需要编程, 只需要进行点选、拖、拉、拽的方式进行操作;
- ✧ **支持自定义算法**: 用户可自行编写算法并将自定义算法上传, 并与已有算法进行对比, 进行算法优化研究;
- ✧ **提供大量算法并提供 API (JAR 包)**: 软件涵盖大量数据挖掘算法, 用户不仅可以运用不同的算法进行数据挖掘分析, 也可根据具体的业务情况, 与第三方业务系统之间进行交互。

2.2 实训教程

结合课程教学需要, 数据挖掘教学实训平台(TipDM-T6)提供如下实训教程:



- ✧ 《R语言与数据挖掘》、《Python与数据挖掘》从初学者的角度出发，内容由浅入深，循序渐进。全书从安装到基础函数的使用，对各个操作步骤详细叙述，凡涉及的常用参数均加以说明，每个操作函数均有实际的示例，极大程度地降低了初学者使用函数的难度。书中通过理论说明+实践操作的方式，极大提高了书本的实用性和可读性，介绍了分类与预测、聚类分析、关联规则、智能推荐和时间序列等分析算法，帮助读者快速掌握应用R/Python语言进行分析挖掘建模的方法。
- ✧ 《MATLAB数据分析与挖掘实战》、《R语言数据分析与挖掘实战》、《Python数据分析与挖掘实战》以大家熟知的数据挖掘建模工具（MATLAB、R、Python）语言来展开，以解决某个应用的挖掘目标为前提，先介绍案例背景提出挖掘目标，再阐述分析方法与过程，最后完

成模型构建，在介绍建模过程中穿插操作训练，把相关的知识点嵌入相应的操作过程中，使读者轻松理解并掌握相关的理论和知识点。

图书内容示例：

文档结构图

- 实战篇
 - 第6章 电力窃漏用户自动识别
 - 6.1 背景与挖掘目标
 - 6.2 分析方法与过程
 - 6.2.1 数据抽取
 - 6.2.2 数据探索分析
 - 1. 分布分析
 - 2. 周期性分析
 - 6.2.3 数据预处理
 - 1. 数据清洗
 - 2. 缺失值处理
 - 3. 数据变换
 - 6.2.4 构建专家样本
 - 6.2.5 构建模型
 - 1. 构建窃漏电用户识别模型
 - 2. 模型评价
 - 3. 进行窃漏电诊断
 - 6.3 上机实验
 - 1. 实验目的
 - 2. 实验内容
 - 3. 实验方法与步骤
 - 4. 思考与实验总结
 - 6.4 拓展思考
 - 6.5 小结

1. 构建窃漏电用户识别模型

在专家样本准备完成后，需要划分测试样本和训练样本，随机选取 20%作为测试样本，剩下的作为训练样本。窃漏电用户识别可通过构建分类预测模型来实现，比较常用的分类预测模型有 LM 神经网络和 CART 决策树，各个模型都有各自的优点，故采用这两种方法构建窃漏电用户识别，并从中选择最优的分类模型。构建 LM 神经网络和 CART 决策树模型时输入项包括电量趋势下降指标、线损类指标和告警类指标，输出项为窃漏电标识。

(1) 数据划分

对专家样本随机选取 20%作为测试样本，剩下的 80%作为训练样本。其代码如代码清单 6-2 所示。

代码清单 6-2 原始数据分为训练数据测试数据

```
%% 把数据分为两部分：训练数据、测试数据
clear;
% 参数初始化
datafile = './data/model.xls'; % 数据文件
trainfile = './tmp/train_model.xls'; % 训练数据文件
testfile = './tmp/test_model.xls'; % 测试数据文件
proportion = 0.8; % 设置训练数据比例

%% 数据分割
```

文档结构图

- 实战篇
 - 第6章 电力窃漏用户自动识别
 - 6.1 背景与挖掘目标
 - 6.2 分析方法与过程
 - 6.2.1 数据抽取
 - 6.2.2 数据探索分析
 - 1. 分布分析
 - 2. 周期性分析
 - 6.2.3 数据预处理
 - 1. 数据清洗
 - 2. 缺失值处理
 - 3. 数据变换
 - 6.2.4 构建专家样本
 - 6.2.5 构建模型
 - 1. 构建窃漏电用户识别模型
 - 2. 模型评价
 - 3. 进行窃漏电诊断
 - 6.3 上机实验
 - 1. 实验目的
 - 2. 实验内容
 - 3. 实验方法与步骤
 - 4. 思考与实验总结
 - 6.4 拓展思考
 - 6.5 小结

```
%% 保存数据
save(treefile,'tree'); % 保存决策树模型

output = vec2ind(output);
output = output';
xlswrite(trainoutputfile,[txt,'模型输出',num2cell([data,output-1])]);
disp('CART 决策树模型构建完成！');
```

*代码详见：示例程序/code/construct_dt_model.m

Confusion Matrix

	1	2	
1	199 85.8%	8 3.4%	96.1% 3.9%
2	3 1.3%	22 9.5%	88.0% 12.0%
	98.5% 1.5%	73.3% 26.7%	95.3% 4.7%
	1	2	

Output Class

Target Class

图 6-9 利用训练样本构建 CART 决策树的混淆矩阵

2.3 教学案例库及配套资源

顶尖数据挖掘教学实训平台之案例库（TipDM-C10）是泰迪科技在数据挖掘领域探索10余年和高校资深讲师联合经验总结之作，内容涵盖电智能电网、移动通信、医疗健康、网络舆情、电子商务、生产制造、金融保险、交通运输、信息安全等诸多行业，特别适合有数据挖掘相关课程教学的高校、研究所和培训机构，也可作为个人学习数据挖掘技术的最佳素材。

数据挖掘行业案例主要包括：

行业类别	案例名称	行业类别	案例名称
智能电网	电力窃漏电用户自动识别	生产制造	家用电器用户行为分析与事件识别
	气象与输电线路缺陷关联分析		基于数据挖掘的生产质量控制
	电能计量异常智能诊断		客户空调使用习惯数据挖掘

移动通信	基于基站定位数据的商圈分析	金融保险	信用卡风险管理与控制
	电信客户流失预测		贷款产品申请审批结果预测
	电信用户交往圈构建和用户识别		资金流入流出预测

医疗健康	中医证型的关联规则挖掘	交通运输	航空公司客户价值分析
	人口与医疗需求预测		城市公交站点设置的优化分析
	脑中风发病因素分析		铁路旅客流量预测

网络舆情	面向网络舆情的关联度分析	信息安全	违规数据库连接自动识别
	用户浏览行为分析及个性化新闻推荐		基于数据挖掘的网站入侵信息监测
	贵阳网络新闻舆情分析		基于数据挖掘的个人信息防护

电子商务	电子商务智能推荐服务	其它行业	应用系统负载分析及磁盘容量预测
	电商评论数据情感分析		基于水色图像的水质评价
	酒店客户数量预测		石油勘探中的油气层自动评价

说明：案例库不断更新中……。

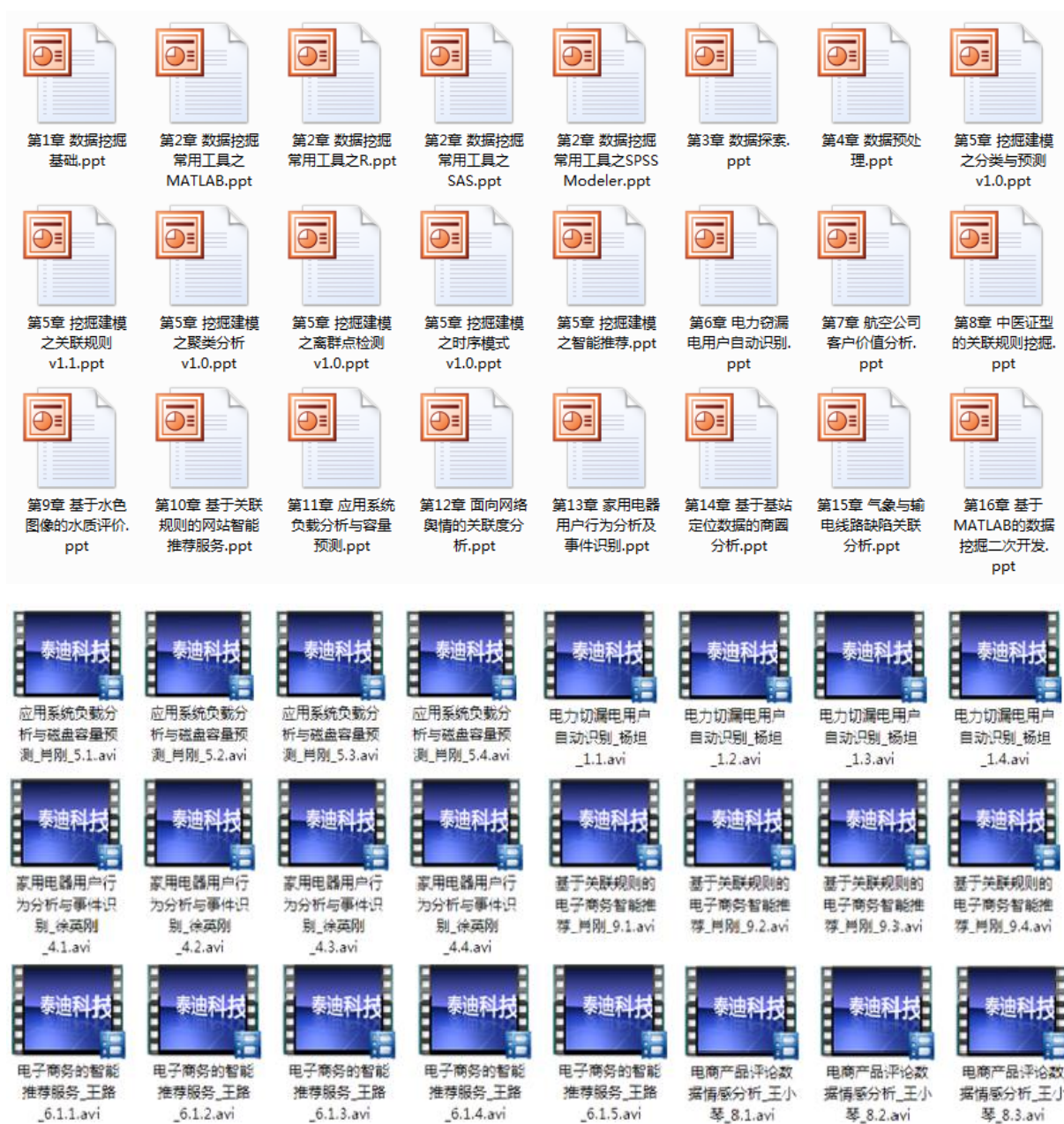
案例配套资源包括：源数据及过程数据，MATLAB、R、Python、TipDM等环境下的程序/模型。

04-Matlab数据分析与挖掘实战	名称	修改日期	类型	大小
01 - 图书内容	bin2ind.m	2015/3/6 18:29	M 文件	1 KB
02 - 数据及程序	cal_successrate.m	2015/3/6 18:29	M 文件	1 KB
chapter6	decision_tree.m	2015/3/6 18:29	M 文件	2 KB
上机实验	decision_tree_test.m	2015/3/6 18:29	M 文件	1 KB
code	ind2bin.m	2015/3/6 18:29	M 文件	1 KB
data	lm_net.m	2015/3/6 18:29	M 文件	2 KB
tmp	plotroc_confusion.m	2015/3/6 18:29	M 文件	1 KB
上机实验拓展	ployinterp_column.m	2015/3/6 18:29	M 文件	2 KB
示例程序	polyinterp.m	2015/3/6 18:29	M 文件	1 KB
code	preprocess_interp.m	2015/3/6 18:29	M 文件	1 KB
data	saveconfusion_roc.m	2015/3/6 18:29	M 文件	1 KB
tmp	split2train_test.m	2015/3/6 18:29	M 文件	1 KB
拓展思考	test_decisontree_lm.m	2015/3/6 18:29	M 文件	1 KB
03 - PPT	test_net.m	2015/3/6 18:29	M 文件	1 KB
04 - 视频资料	test_preprocess_interp.m	2015/3/6 18:29	M 文件	1 KB
	train_net.m	2015/3/6 18:29	M 文件	1 KB

chapter12	名称	修改日期	类型	大小
上机实验	Clean.R	2015/11/2 16:19	R 文件	1 KB
code	DataChange.R	2015/11/2 16:19	R 文件	2 KB
data	DataSplit.R	2015/11/2 16:19	R 文件	1 KB
tmp	RConnectDatabase.R	2015/11/2 15:51	R 文件	2 KB
示例程序	RecommenderSystem.R	2015/11/2 15:23	R 文件	3 KB
chapter13				

上机实验	名称	修改日期	类型	大小
上机实验拓展	missing_data.xls	2015/3/6 18:29	Microsoft Office...	25 KB
示例程序	model.xls	2015/3/6 18:29	Microsoft Office...	40 KB
code	test_model.xls	2015/3/6 18:29	Microsoft Office...	26 KB
data	train_model.xls	2015/3/6 18:29	Microsoft Office...	36 KB
tmp				
拓展思考				

针对教师授课需要，T6同时提供配套的全过程教学用PPT及微课视频。



2.4 教学沙盘

数据挖掘教学沙盘能够让老师与学生在实验室模拟案例现场真实数据采集过程，在讲解数据挖掘算法理论的同时，通过实用的建模工具来模拟实际应用案例，在实验室环境下体验真实应用，由此弥补学员实践经验的不足。

“沙盘模拟”实验课程改变传统的教学模式，将理论与实践融为一体的设计思路，使学生在参与、体验中完成从知识到技能的转化，这种体验式教学方式完全不同于传统理论教学及案例教学，是教学方式的一大创新。沙盘模拟过程中，学生经历了一次从理论到实践再到理论的上升过程，实现从感性到理性的飞跃。在全面深刻理解分析设计、数据取样、数据清洗、模型构建、推断思考等

数据挖掘建模问题。下面以智能交通仿真沙盘为例展开说明：

智能交通仿真沙盘系统由硬件层和软件层组成。硬件层，即公交系统运行实物模型，通过运行平台，地磁感应器，RFID Reader 刷卡机，公交卡，可充电公交车模型实现公交车在运行过程中乘客刷卡上车的仿真模拟。软件层，即信息采集，将模型所得到的信息通过数据线连接方式在数据接收终端（计算机）上得到展现。通过硬软件结合方式，形成一个可持续观测的公交模拟系统。



采集的交通信息经过 ETL 数据迁移，传输到数据中心后，即可通过数据挖掘建模，实现城市公交服务水平动态评价、公交线路及站点优化等应用。

2.5 课程规划及师资培训

为适应大数据时代及数据挖掘实践教学需求，泰迪科技可参与到高校教师队伍建设，协助完成课程设置及教学规划，并为数据挖掘课程教学提供一对一的师资培训。培训课程主要包含以下模块：

模块	课程主题	主要内容
模块一	✧ 数据挖掘基本理论 ✧ 数据挖掘工具 ✧ 数据挖掘建模过程	1、 分类与预测、聚类分析、关联规则、时序模式、智能推荐 2、 TipDM、R、Python 3、 数据探索、数据预处理、挖掘建模、模型评价
模块二	✧ 全真案例详解一	电力窃漏电用户自动识别

(示例)	✧ 全真案例详解二	航空公司客户价值分析
	✧ 全真案例详解三	中医证型关联规则挖掘
	✧ 全真案例详解四	基于水色图像的水质评价
	✧ 全真案例详解五	电子商务网站用户行为分析与服务推荐
	✧ 全真案例详解六	应用系统负载分析与磁盘容量预测
	✧ 全真案例详解七	面向网络舆情的关联度分析
	✧ 全真案例详解八	家用电器用户行为分析及事件识别
	✧ 全真案例详解九	基于基站定位数据的商圈分析
	✧ 全真案例详解十	气象与输电线路缺陷关联分析
模块三	✧ 数据挖掘二次开发	基于 R、Python 的数据挖掘二次开发

2.6 学生实习实训

结合课程教学需要，可组织学生到泰迪公司进行集中实训或实习，实习项目为从在建的企业项目中剥离出部分数据挖掘应用子课题，让学生真切参与到企业数据挖掘项目建设中，实习实训包括校内和校外（在泰迪公司现场）两种模式。

3 产品优势

T6 与其它同类产品相比，具有如下明显优势：

- ✧ 从教学实践出发，结合大量数据挖掘工程案例及教学经验；
- ✧ 以真实案例为主线，深入浅出介绍数据建模过程中的相关任务；
- ✧ 通过教学沙盘使学生轻松获取真实的生产环境，使用目前流行的 MATLAB、R、Python 等建模工具对样本数据进行处理并进行挖掘建模；
- ✧ 提供真实的原始建模数据及数据探索、数据预处理、模型构建及评价等不同阶段的程序代码；
- ✧ 提供数十个完整企业数据挖掘案例及配套源数据、PPT 课件、视频课件等资源，方便教师授课需要。

4 应用案例

4.1 主要客户

- ◇ 广东工业大学
- ◇ 珠江水产研究所
- ◇ 华南师范大学
- ◇ 广东中医药大学
- ◇ 武汉理工大学
- ◇ 韩山师范学院
- ◇ 广西财经学院
- ◇ 广东技术师范学院
- ◇ 深圳职业技术学院
- ◇

4.2 典型示例

2014 年，为提高特色教学优势，培养应用型社会人才，广州某大学应用数学学院与泰迪科技合作建设数据挖掘建模实验室，培养大数据挖掘人才。

1. 解决方案

总共 60 个实验机位，要求每个学生至少拥有 1 个独立主机的实验环境，每个机位采用一台带显示器和键鼠的瘦终端系统，设计使用 1 台安装有 Linux 操作系统的服务器，用以支撑多达 60 个客户机的实验环境，以完成数据采集、数据清洗、数据整合、数据挖掘的学习实验。

2. 实验室配置

序号	配置	明细	参数	数量
1	硬件	服务器	IBM System x3650 M4 CPU: Intel E5-2600v2 2.6GHz 2颗 内存: 64G 硬盘: 500G	1台
2		客户机	CPU: Intel酷睿i5 内存容量: 4GB DDR3 1600MHz 硬盘: 500G	60台
3		投影仪	2500流明以上 分辨率: 1280*800	1台
4	软件	建模工具	详见软件功能	1套
5	配套资源及服务	实训教程	《R语言与数据挖掘》、《R语言数据分析与挖掘实战》	各10本
6		教学案例库及配套资源	智能电网、移动通信、医疗健康、网络舆情、电子商务、生产制造、金融保险、交通运输、信息安全等行业10个案例及配套源数据、PPT课件、视频课件等资源	1套

7		教学沙盘	餐饮菜品智能推荐沙盘、防窃漏电用户自动识别沙盘、智能交通仿真沙盘、管网漏损智慧水务沙盘、	4套
8		师资培训	针对建模基本理论、R语言建模工具及企业案例培训	10人. 天
9		学生实习实训	学生校内实习实训	10天现场

3. 资源分配

实验室由泰迪科技与学校共同建设，在实验室设施上主要依托高校，在教学人员方面由泰迪科技派出数据挖掘专家，合作教学。课程设置由泰迪科技专家组与学校相关教师共同讨论完成。