



张策

第9章 物联网智能决策

嵌入式系统研发中心
山东省嵌入式系统工程技术研发中心

本章内容

第一节 数据挖掘概述

第二节 数据挖掘的基本类型和算法

第三节 智能决策与物联网

第四节 物联网的其它智能化技术

第9章 物联网智能决策--9.1 数据挖掘概述

一、概述

✍ 智能决策是物联网“智慧”的来源。

0. 物联网中数据特点

- ✍ 智能信息处理指信息的储存、检索、智能化分析利用，比如利用人工智能、专家系统对感知的信息作出决策和处理等。
- ✍ 物联网的智能信息处理主要针对感知的数据，而物联网的数据具有独特的特点：

(1) 异构性

- ① 在物联网中，不仅不同的感知对象有不同类型的表征数据，即使是同一个感知对象也会有各种不同格式的表征数据。
- ① 比如在物联网中为了实现对一栋**写字楼的智能感知**，需要处理各种不同类型的数据，如探测器传来的各种高维观测数据，专业管理机构提供的关系数据库中的关系记录，互联网上提供的相关超文本链接标记语言(HTML)、可扩展标记语言(XML)、文本数据等。
- ① 为了实现完整准确的感知，必须综合利用不同类型的数据获得全面准确信息。

一、概述

0. 物联网中数据特点

(2) 海量性

- ① 物联网是网络和数据的海洋。
- ① 在物联网中海量对象连接在一起，每个对象每时每刻都在变化，表达其特征的数据也会不断地积累。
- ① 如何有效地改进已有的技术和方法，或者提出新的技术和方法，从而高效地**管理**和**处理**这些海量数据，将从这些原始数据中提取信息并进一步**融合**、**推理**和**决策**的关键。

(3) 不确定性

- ① 物联网中的数据具有明显的不确定性特征，主要包括：
 - ① **数据本身的不确定性**、② **语义匹配的不确定性**、③ **查询分析的不确定性**等。
- ① 为了获得客观对象的准确信息，需要去粗取精、去伪存真，以便更全面地进行表达和推理。

第9章 物联网智能决策--9.1 数据挖掘概述

一、概述

✎ 智能决策是物联网“智慧”的来源。

0. 数据融合——含义

✎ 数据融合是一种数据处理技术，指将多种数据或信息进行处理得出高效且符合用户需求的数据的过程。它是利用计算机对按时序获得的若干观测信息，在一定准则下加以自动分析、综合，以完成所需的决策和评估任务而进行的一种信息处理技术。

✎ 数据融合类似人类和其他动物对复杂问题的综合处理，比如在辨别一个事物的时候，通常会综合各种感官信息，包括视觉、触觉、嗅觉和听觉等。单独依赖一个感官获得的信息往往不足以对事物做出准确的判断，而综合各种感官数据，对事物的描述会更准确。

✎ 数据融合也是一个多级、多层面的数据处理过程，能完成对来自多个信息源的数据的自动检测、关联、估计及组合等的处理，并基于多信息源数据进行综合、分析、判断和决策。

0. 数据融合——分类

✎ 数据融合一般有数据级融合、特征级融合、决策级融合等层次的融合。

① 数据级融合：直接在采集到的原始数据上进行融合，是最低层次的融合，它直接融合现场数据，失真度小，提供的信息比较全面。

② 特征级融合：先对来自传感器的原始信息进行特征提取，然后对特征信息进行综合分析和处理，这一级的融合可实现信息压缩，有利于实时处理，属于中间层次的融合。

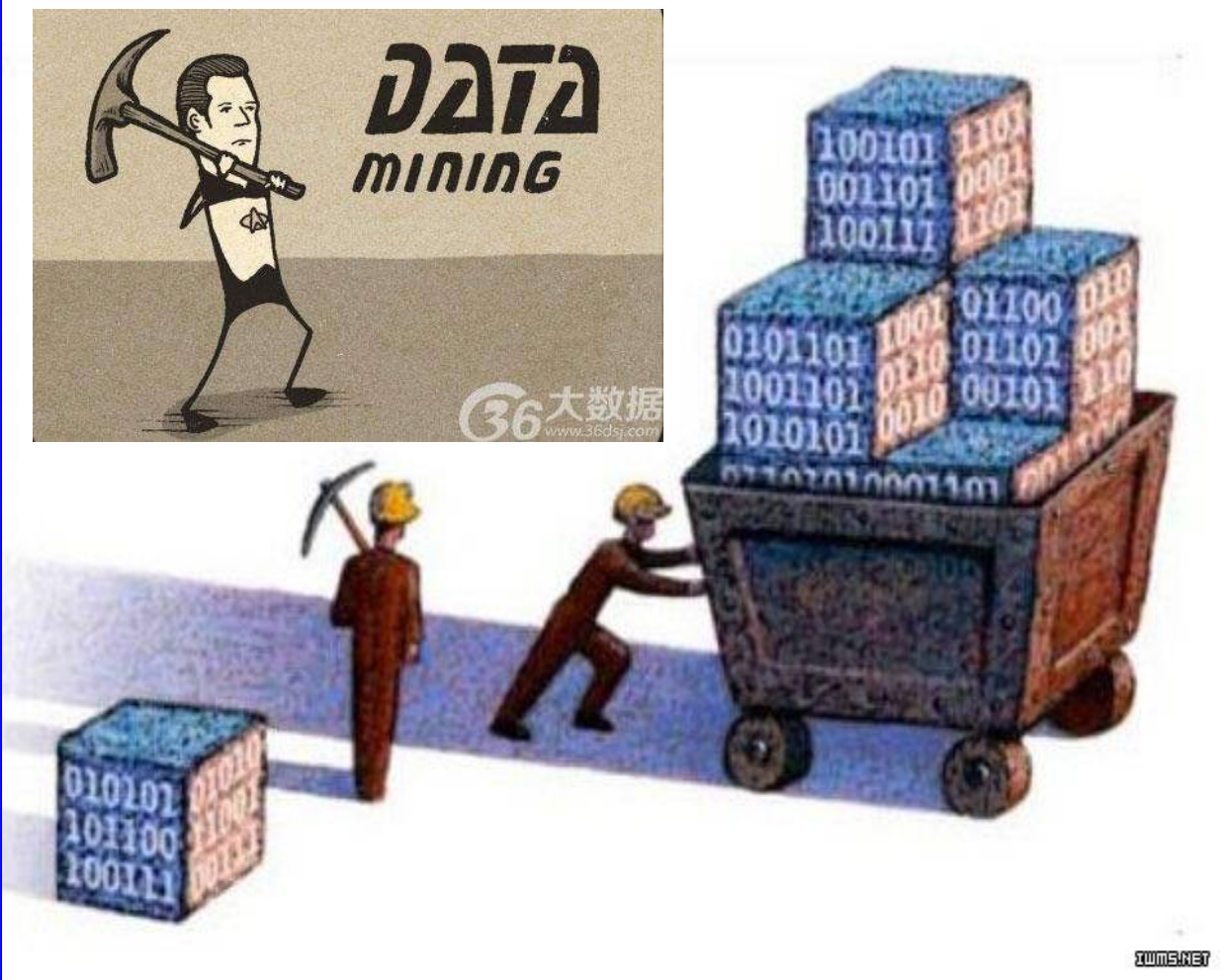
③ 决策级融合：在高层次上进行，根据一定的准则和决策的可信度做最优决策，以达到良好的实时性和容错性。

第9章 物联网智能决策--9.1 数据挖掘概述

一、概述

智能决策是物联网“智慧”的来源。

1. 数据挖掘 (Data Mining) 含义

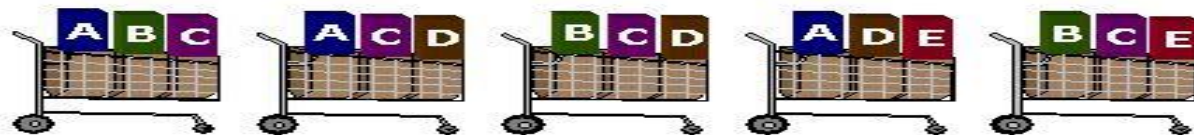


第9章 物联网智能决策--9.1数据挖掘概述

一、概述



Association Rules



Rule	Support	Confidence
$A \Rightarrow D$	2/5	2/3
$C \Rightarrow A$	2/5	2/4
$A \Rightarrow C$	2/5	2/3
$B \& C \Rightarrow D$	1/5	1/3



一、概述

智能决策是物联网“智慧”的来源。

1. 数据挖掘 (Data Mining) 含义

啤酒和尿布

- 在沃尔玛超市里，有一个有趣的现象：尿布和啤酒赫然摆在一起出售，而且尿布和啤酒的销量双双增加了。原来，美国的妇女们经常会嘱咐她们的丈夫下班以后要为孩子买尿布。而丈夫在买完尿布之后又要顺手买回自己爱喝的啤酒，因此啤酒和尿布在一起购买的机会还是很多的。

是什么让沃尔玛发现了尿布和啤酒之间的关系呢？正是商家通过对超市一年多原始交易数字进行详细的分析，才发现了这对神奇的组合。

第9章 物联网智能决策--9.1 数据挖掘概述

一、概述

✍ 智能决策是物联网“智慧”的来源。

1. 数据挖掘 (Data Mining) 含义

- ✍ 一个从大量数据中获取潜在有用的并且可以被人们理解的模式的过程；
- ✍ 数据挖掘是从大量的、不完全的、有噪声的、模糊的、随机的数据中提取潜在的、事先未知的、有用的、能被人理解的有用的信息和知识的数据处理过程；
- ✍ 是一个反复迭代的人机交互和处理的过程，历经多个步骤，并且在一些步骤中需要由用户提供决策；
- ✍ 与数据挖掘相近的同义词有数据融合、数据分析和决策支持等。

① 数据挖掘的数据源必须是真实的、大量的、含噪声的；① 发现的是用户感兴趣的知识； ② 发现的知识要可接受、可理解、可运用。

① 被挖掘的数据可以是：① 结构化的关系数据库中的数据、 ② 半结构化的文本、
③ 图形和图像数据、 ④ 或者是分布式的异构数据。

2. 数据挖掘的过程

✍ 可分为：

(1) 数据预处理、(2) 数据挖掘、(3) 对挖掘结果的评估与表示

✍ 每一个阶段的输出结果成为下一个阶段的输入

第9章 物联网智能决策--9.1 数据挖掘概述

一、概述

2. 数据挖掘的过程

(1) 数据预处理阶段

- ① **数据准备**：了解领域特点，确定用户需求（调研）
- ② **数据选取**：从原始数据库中选取相关数据或样本（采样）
- ③ **数据预处理**：检查数据的完整性及一致性，消除噪声等（去噪）
- ④ **数据变换**：通过投影或利用其他操作减少数据量（压缩）

(2) 数据挖掘阶段

- ① **确定挖掘目标**：确定要发现的知识类型（目标）
- ② **选择算法**：根据确定的目标选择合适的数据挖掘算法（算法）
- ③ **数据挖掘**：运用所选算法，提取相关知识(模式)并以一定的方式表示（挖掘）

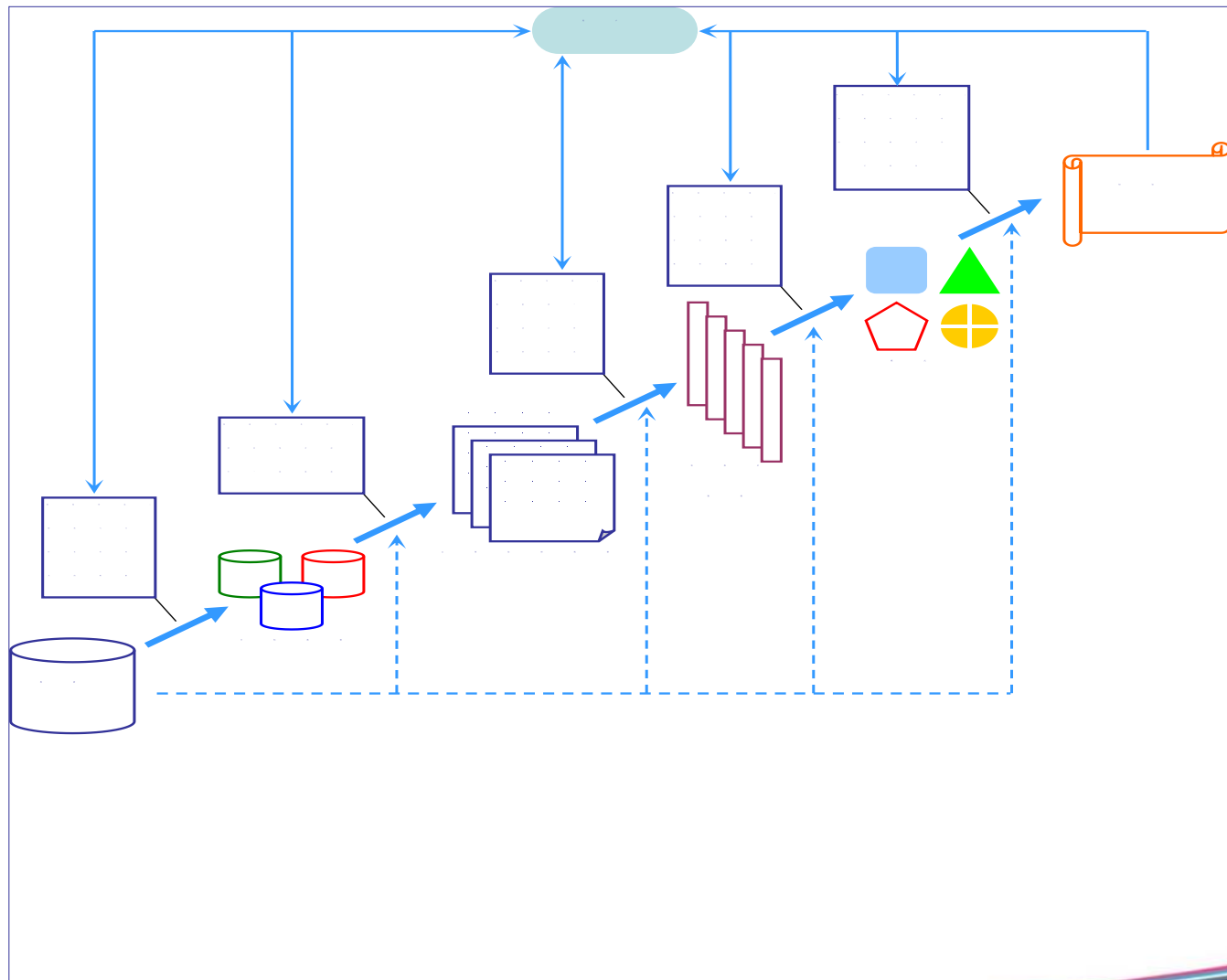
(3) 知识评估与表示阶段

- ① **模式评估**：对在数据挖掘步骤中发现的模式(知识)进行评估（评估）
- ② **知识表示**：使用可视化和知识表示相关技术，呈现所挖掘的知识（知识）

第9章 物联网智能决策--9.1数据挖掘概述

一、概述

2. 数据挖掘的过程



第9章 物联网智能决策--9.1 数据挖掘概述

一、概述

3. 数据挖掘的主要任务

数据挖掘的任务有关联分析、聚类分析、分类、预测、时序模式和偏差分析等，以下分别进行介绍。

① 关联分析(Association analysis): 两个或两个以上变量的取值之间存在某种规律性，称为关联。数据关联是数据库中存在的一类重要的、可被发现的知识。关联分为简单关联、时序关联和因果关联。关联分析的目的是找出数据库中隐藏的关联网。一般用支持度和可信度两个阈值来度量关联规则的相关性，还可引入兴趣度、相关性等参数，使得所挖掘的规则更符合需求。

② 聚类分析(Clustering): 聚类是把数据按照相似性归纳成若干类别，同一类中的数据彼此相似，不同类中的数据相异。聚类分析可以建立宏观的概念，发现数据的分布模式，以及可能的数据属性之间的相互关系。

③ 分类(Classification): 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，即该类的内涵描述，用这种描述构造模型，一般用规则或决策树模式表示。分类是利用训练数据集通过一定的算法而求得分类规则。分类可被用于规则描述和预测。

第9章 物联网智能决策--9.1 数据挖掘概述

一、概述

3. 数据挖掘的主要任务

数据挖掘的任务有关联分析、聚类分析、分类、预测、时序模式和偏差分析等，以下分别进行介绍。

④ 预测(Predication)：预测是利用历史数据找出变化规律，建立模型，并由此模型对未来数据的种类及特征进行预测。预测关心的是精度和不确定性，通常用预测方差来度量。

⑤ 时序模式(Time-series pattern)：时序模式是指通过时间序列搜索出的重复发生概率较高的模式。与回归一样，它也是用已知的数据预测未来的值，但这些数据的区别是变量所处时间的不同。

⑥ 偏差分析(Deviation)：在偏差中包括很多有用的知识，数据库中的数据存在很多异常情况，发现数据库中数据存在的异常情况是非常重要的。偏差检验的基本方法就是寻找观察结果与参照之间的差别。

二、数据挖掘的基本类型

1. 描述性挖掘任务

- (1) 关联分析(Association Analysis)
- (2) 聚类分析(Clustering Analysis)

✎ **描述性挖掘任务**：刻画数据库中数据的一般特性

2. 预测性挖掘任务

- (1) 离群点分析(Outlier Analysis)
- (2) 分类与预测(Classification and Prediction)
- (3) 演化分析(Evolution Analysis)

✎ **预测性挖掘任务**：在当前数据上进行推断和预测

三、关联分析

1. 基本含义

关联分析的目标是从给定的数据中发现频繁出现的**模式(知识)**，即**关联规则**

⇒

关联规则通常的表述形式是**X Y**，表示“数据库中满足条件X的记录(元组)可能也满足条件Y”

以计算机卖场销售记录为例：

年龄(顾客, “20~29”) ∧ 收入(顾客, “3000~5000”) ⇒ 购买(顾客, “笔记本电脑”)

[支持度=4%, 置信度=65%]

含义：**4%** (支持度)的顾客的年龄在20至29岁且月收入在3000至5000元，**且**这样的顾客中，**65%** (置信度)的人购买了笔记本电脑

三、关联分析

1. 基本含义

挖掘关联规则，需要置信度和支持度越高越好

(1) 基本概念

项集：满足若干条件的数据项的集合，如果条件数为 k ，则称 **k -项集**

① 满足年龄(顾客, "20~29")的项集是**1-项集**

① 满足年龄(顾客, "20~29")、收入(顾客, "3000~5000")的项集是
2-项集

(2) 计算步骤

① 首先找到具备足够**支持度**的项集，即**频繁项集**

① 然后由频繁项集构成**关联规则**，并计算**置信度**

三、关联分析

2. 如何寻找频繁项集(一种挖掘关联规则的频繁项集算法)

(1) Apriori算法

✎ **基本思想：**利用已求出的 k -项集来计算 $(k+1)$ -项集

① 首先计算频繁1-项集

② 然后根据两个频繁 k -项集 $\{p_1, p_2, \dots, p_k\}, \{q_1, q_2, \dots, q_k\}$ 计算频繁 $(k+1)$ -项集，其中 $p_i = q_i, 1 \leq i \leq k-1$ ，且该 $(k+1)$ -项集为 $\{p_1, p_2, \dots, p_k, q_k\}$

③ 最后判定该 $(k+1)$ -项集是否频繁即可

✎ **缺点：**可能产生大量候选项集，并需要重复地扫描数据库

(2) FP-Growth算法

✎ 利用树状结构（自创的频繁模式树（**Frequent Pattern Tree**）的数据结构）保存项集，从而减小了计算频繁项集所需的存储空间

三、关联分析

2. 如何由频繁项集构造关联规则，并计算置信度

✎ 关联规则 $A \Rightarrow B$ 的置信度

$$\text{Confidence}(A \Rightarrow B) = P(B | A) = \frac{\text{count}(A \text{ AND } B)}{\text{count}(A)}$$

其中 $\text{count}(A \text{ AND } B)$ 为满足条件 A 以及 B 的数据项数目， $\text{count}(A)$ 为满足条件 A 的数据项数目

✎ 计算步骤

- ① 对于每一个频繁项集 S ，计算 S 的所有非空子集
- ① 对于每个 S 的非空子集 F ，若 $\frac{\text{count}(S)}{\text{count}(F)}$ 大于给定置信度阈值，则得到一个关联规则 $F \Rightarrow (S - F)$

四、分类和预测

1. 基本方法

✎ **分类和预测**的目标是找出描述和区分不同数据类或概念的**模型或函数**，以便能够使用模型预测数据类或标记未知的对象

✎ 所获得的**分类模型**可以采用多种形式加以描述输出：

- ① 分类规则
- ① 判定树
- ① 数学公式
- ① 神经网络
- ① ...

✎ **分类与预测的区别**：**分类**通常指预测数据对象属于哪一类，而当被预测的值是数值数据时，通常称为**预测**

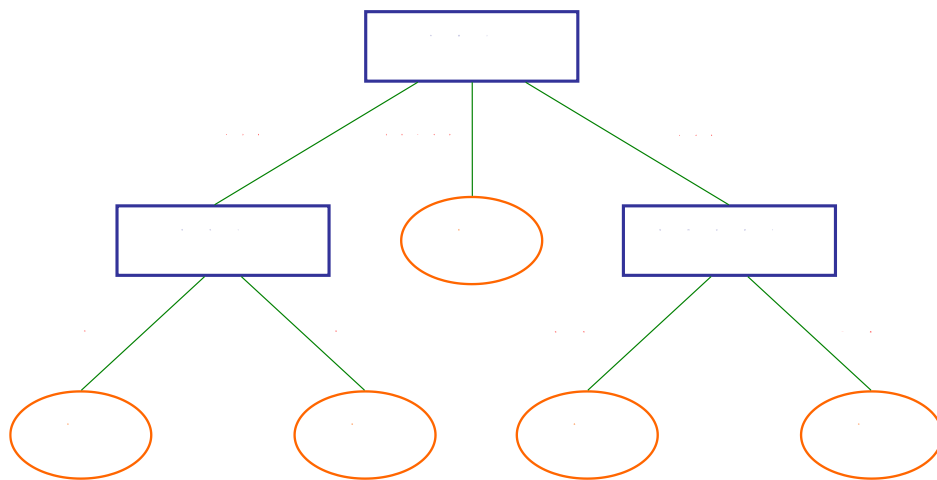
四、分类和预测

2. 以判定树方法为例，简要介绍分类的基本步骤和结果表示

✎ **问题实例：**假定商场需要向潜在的客户邮寄新产品资料和促销信息。客户数据库描述的客户属性包括姓名、年龄、收入、职业和信用记录。

① 我们可以按是否会在商场购买计算机将客户分为两类，只将促销材料邮寄给那些会购买计算机的客户，从而降低成本。

✎ 用于预测客户是否可能购买计算机的**判定树**，其中每个非树叶节点表示一个属性上的测试，每个树叶节点代表**预测结果**



四、分类和预测

2. 以判定树方法为例，简要介绍分类的基本步骤和结果表示

✎ 如何构造上述判定树？

① 基本概念： n 个客户中有 a 个购买了计算机的期望信息

$$I(a, n-a) = -\frac{a}{n} \log \frac{a}{n} - \frac{n-a}{n} \log \frac{n-a}{n}$$

① 建立树节点时，选取合适的判定属性，以最大化期望信息增益

☑ 某种属性上的信息增益大小反映了该属性区分给定数据的能力强弱

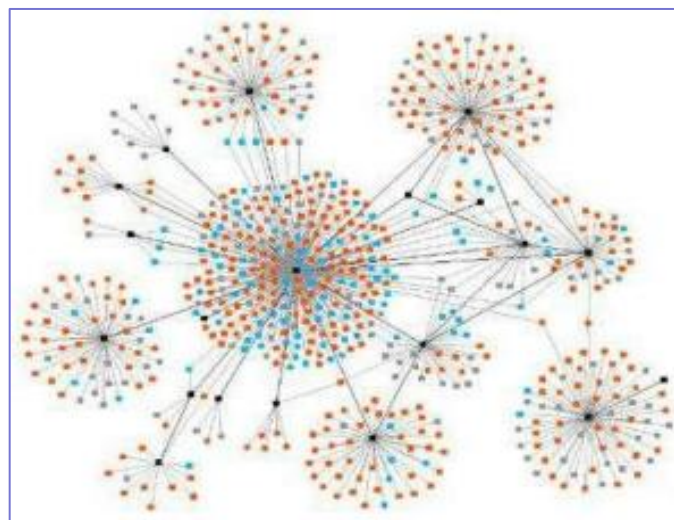
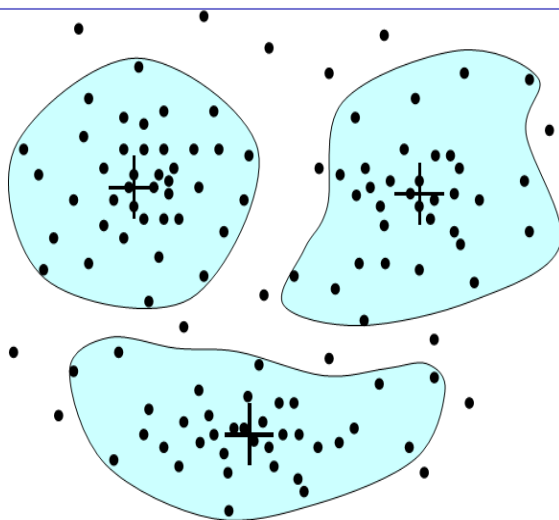
✎ 10条客户记录，其中6人购买了计算机，4人没有购买。这10位客户中有3人的职业是学生，其中有2人购买计算机，而非学生客户购买计算机的有4人。在选择区分属性以前，数据的期望信息为 $E = I(6,4) = 0.673$ 用职业区分之后的期望信息为 $E' = \frac{3}{10} I(2,1) + \frac{7}{10} I(4,3) = 0.669$ ，则选择职业作为区分属性的信息增益为 $E - E' = 0.004$

五、聚类分析

1. 聚类分析基本含义

- ✎ 聚类的目的是将数据对象划分为多个类或簇，在同一个簇中的对象之间具有较高的相似度，而不同簇中的对象差别较大
- ✎ **聚类与分类的区别：** 要划分的类是事先未知的(分类之前一无所知，聚类是有一定预知的信息)
- ✎ **聚类分析的应用**

对客户居住地进行聚类分析，可以帮助商场针对相应顾客群体采取有针对性的营销策略



五、聚类分析

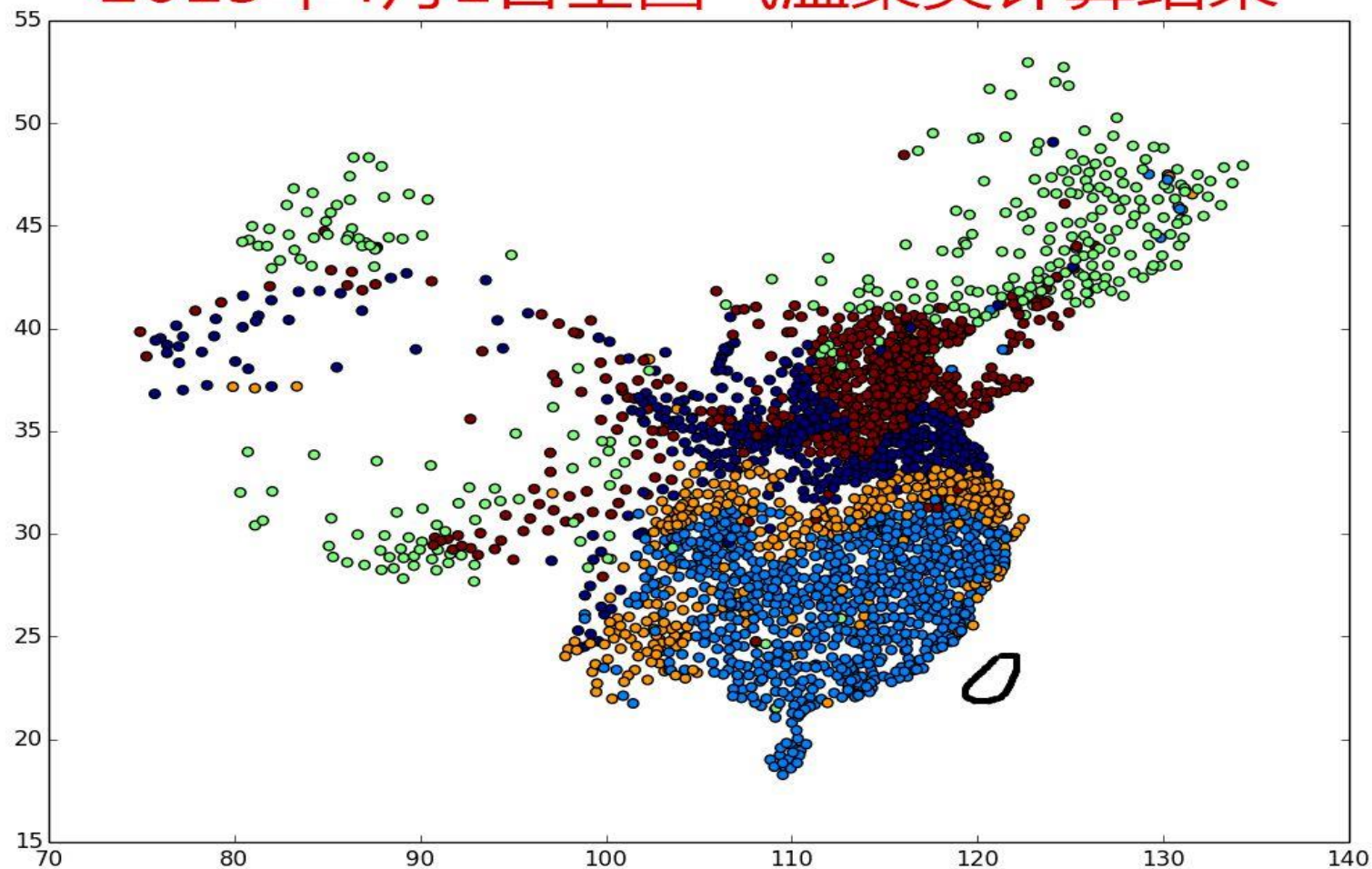
2. 聚类分析的方法

- (1) **划分方法**：要求事先给定聚类的数目 k 。首先创建一个初始划分，然后通过对划分中心点的反复迭代来改进划分。典型算法包括 k -means算法和 k -medoids算法等
- (2) **层次方法**：对给定数据集进行逐层递归的合并或者分裂，因此可以被分为合并或分裂方法。合并方法首先将每个对象都作为独立的类，然后持续合并相近的类，直到达到终止条件为止。分裂方法首先将所有数据对象置于一个类中，然后反复迭代并判定当前的类是否可以被继续分裂，直到达到终止条件为止
- (3) **基于密度的方法**：只要某区域数据密度超过阈值，就将该区域的数据进行聚类。其优势在于噪音数据下的抗干扰能力，并能够发现任意形状的聚类
- (4) **基于网格的方法**：把对象空间量化为具有规则形状的单元格，从而形成一个网格状结构。在聚类的时候，将每个单元格当作一条数据进行处理。优点是处理速度很快，因处理时间与数据对象数目无关，而只与量化空间中的单元格数目相关
- (5) **基于模型的方法**：如果事先已知数据是根据潜在的概率分布生成的，基于模型的方法便可为每个聚类构建相关的数据模型，然后寻找数据对给定模型的最佳匹配。主要分两类：统计学方法和神经网络方法

五、聚类分析

2. 聚类分析的方法

2015年4月1日全国气温聚类计算结果



六、离群点分析

1. 离群点及意义

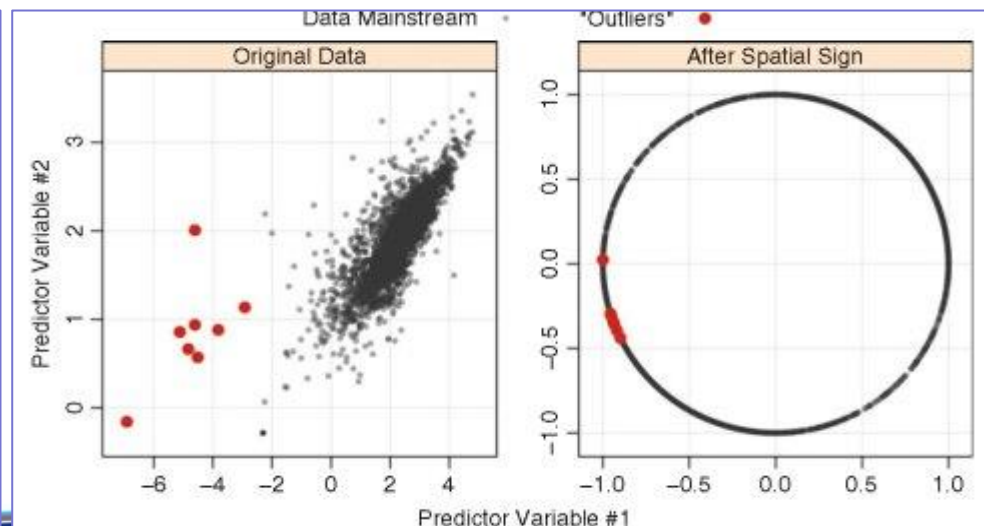
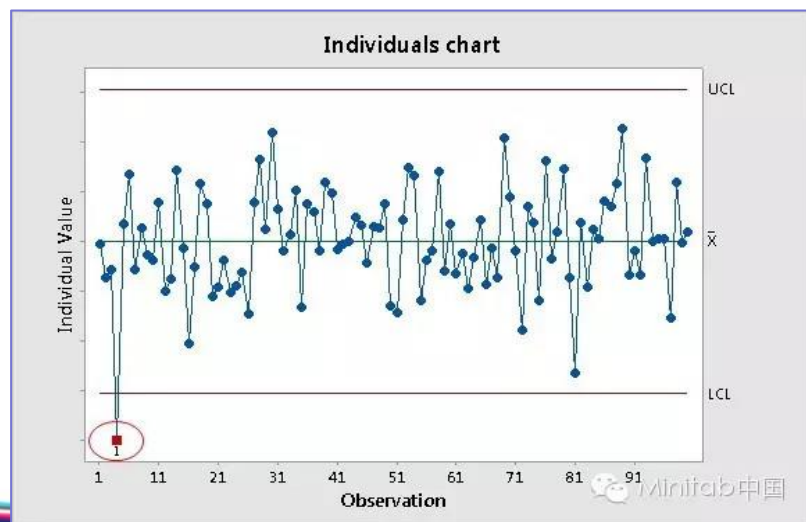
✎ **离群点(Outlier)**: 数据集合中存在的一些数据对象, 它们与其余绝大多数数据的特性或模型不一致

—— “甚至当我还是个孩子的时候, 我就知道自己与众不同”

✎ 寻找离群点的意义

① **发现信用卡诈骗**。通过检测购物地点、商品种类或者购物金额和频率, 能够发现与绝大多数正常消费不一样的记录, 这种行为就有可能属于信用卡诈骗性使用

② **预防网络诈骗**。在网络销售的时候, 诈骗者往往冒充商家, 出售报价比正常价格低出许多的商品, 这样的行为也是可以通过离群点分析被找到的



六、离群点分析

2. 寻找离群点的方法

- ✎ **基于统计的方法**：需要事先已知数据的**分布或概率模型**(例如一个正态分布)，然后根据数据点与该模型的不一致性检验来确定离群点
- ✎ **基于距离的方法**：不需要数据模型，而是将那些没有足够邻居的数据对象看作是离群点，这里的邻居是基于距给定对象的距离来定义的。现有的基于距离的离群点探测算法又分为**基于索引的算法**，**嵌套循环算法**和**基于单元的算法**，其目的都是为了减小计算和I/O开销
- ✎ **基于偏移的方法**：不采用统计检验或基于距离的度量值来确定异常对象。相反，它通过检查数据对象的一组主要特征来确定离群点。偏离事先给出的特征描述的数据对象被认为是离群点

七、演化分析

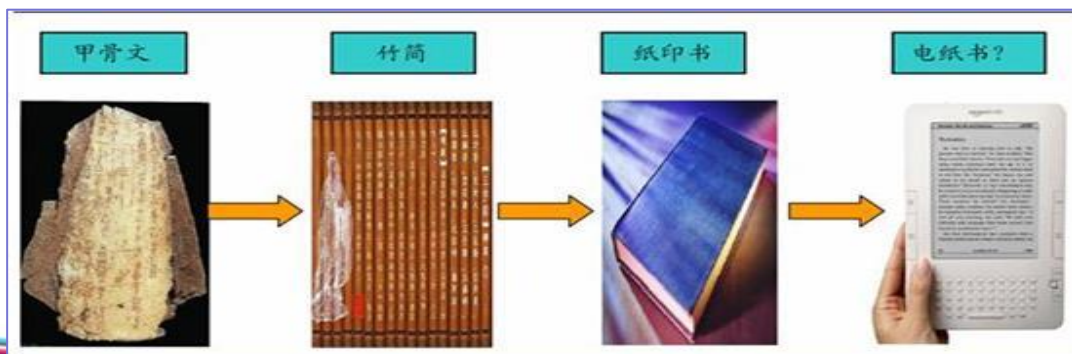
1. 演化分析

✎ **演化分析**的目的是挖掘随时间变化的数据对象的变化规律和趋势，并对其建模，进而为相关决策提供参考

✎ 演化分析的应用

- ① 对**股票**的演化分析可以得出整个股票市场和特定的公司的股票变化规律，为投资者决策提供帮助
- ① 对**生态和气候**的演化分析可以知道人类活动对自然的影响程度，为环境保护提供重要依据
- ① ...

➤ **建模方法**：除了关联分析和分类分析，还包括与时间相关的数据分析方法，主要包括趋势分析、相似搜索、序列模式挖掘和与周期分析



计算机 → 互联网 → 嵌入式系统
→ 无线传感器网络 → 物联网
→ CPS

七、演化分析

王晓光, 程齐凯. 基于NEViewer的学科主题演化可视化分析. 情报学报, 2013, 32 (9) : 900-911.



第9章 物联网智能决策--9.2 数据挖掘的基本类型和算法

七、演化分析

1. 演化分析

王晓光, 程齐凯. 基于NEViewer的学科主题演化可视化分析. 情报学报, 2013, 32 (9): 900-911.



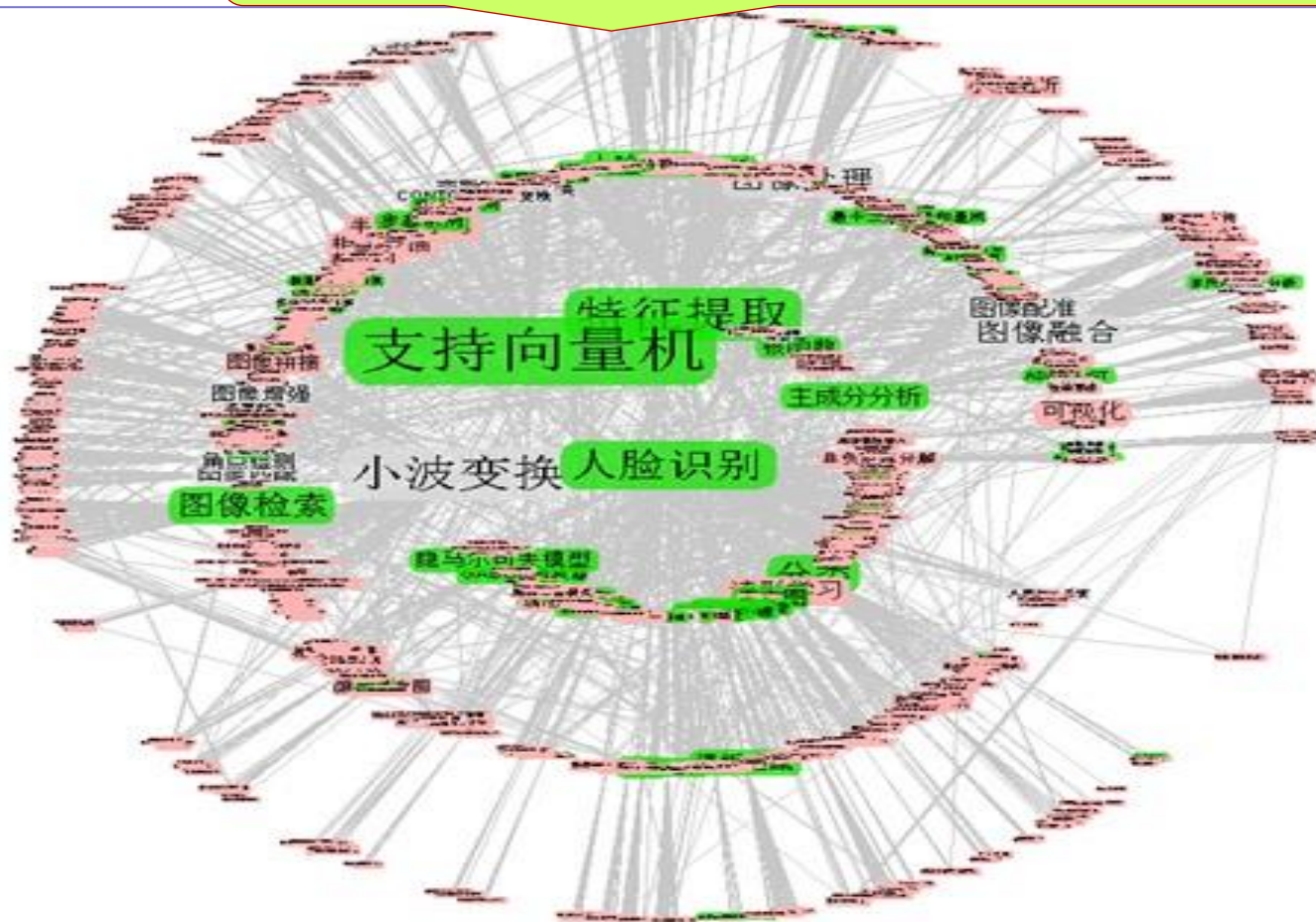
上图代表了一个特定主题的演化过程

第9章 物联网智能决策--9.2 数据挖掘的基本类型和算法

七、演化分析

1. 演化分析

王晓光, 程齐凯. 基于NEViewer的学科主题演化可视化分析. 情报学报, 2013, 32 (9): 900-911.



“人脸识别主题”的前向赋色网络

上图代表了一个特定的主题（共词网络中的社区）中不同的概念的来源，及遗传过程。

七、演化分析

2. 建模方法

✎ **建模方法：**除了关联分析和分类分析，还包括与时间相关的数据分析方法，主要包括趋势分析、相似搜索、序列模式挖掘和与周期分析

✎ 与时间相关的数据分析方法

- (1) **趋势分析：**确定趋势的常见方法是计算数据n阶的变化平均值，或者采用最小二乘法等方法平滑数据变化曲线
- (2) **相似搜索：**相似搜索用于找出与给定序列最接近的数据序列
- (3) **序列模式挖掘：**挖掘相对时间或其它维属性出现频率高的模式
- (4) **周期分析：**挖掘具有周期的模式或者关联规则，例如“若每周六公司的下班时间比平时晚半小时以上，则选择打车回家的人数大约增加20%”

第9章 物联网智能决策--9.3 智能决策与物联网

一、物联网的数据挖掘

- 数据挖掘是**决策支持**和**过程控制**的重要技术手段，是物联网中重要的一环。
- 针对**物联网具有行业应用的特征**，对各行各业的、数据格式各不相同的海量数据进行**整合、管理、存储**，并在整个物联网中提供数据挖掘服务，实现**预测、决策**，**进而反向控制这些传感网络**，达到控制物联网中客观事物运动和发展进程的目的。
- 在物联网中进行数据挖掘已经从传统意义上的数据统计分析、潜在模式发现与挖掘，转向成为物联网中不可缺少的工具和环节。

1. 物联网的计算模式

- 物联网一般有两种基本计算模式，即**物计算模式**和**云计算模式**。

① **物计算模式**基于**嵌入式系统**，**强调实时控制**，对终端设备的性能要求较高，**系统的智能主要表现在终端设备上**。物计算模式建立在对智能信息结果的利用上，对集中处理能力和系统带宽要求较低。

① **云计算**以**互联网为基础**，**目的是实现资源共享和资源整合**，其计算资源是动态、可伸缩、虚拟化的。**云计算模式通过分布式的构架采集物联网中的数据，系统的智能主要体现在数据挖掘和处理上**，需要较强的集中计算能力和高带宽，但终端设备比较简单。

一、物联网的数据挖掘

2. 物联网两种计算模式的选择

物联网数据挖掘的结果主要用于决策控制，挖掘出的模式、规则、特征指标用于预测、决策和控制。在不同的情况下，可以选用不同的计算模式。

① 物计算模式：

- ① 对于要求实时高效的数据挖掘，物联网任何一个控制端均需要对瞬息万变的环境实时分析、反应和处理，需要物计算模式和利用数据挖掘结果。

② 基于云计算模式：

- ① 以海量数据挖掘为特征的应用要处理海量的数据，需要进行数据质量控制，如多源、多模态、多媒体、多格式数据的存储与管理保证数据质量。此外，物联网计算设备和数据具有天然分布的特征。

第9章 物联网智能决策--9.3 智能决策与物联网

一、物联网的数据挖掘

3. 数据挖掘算法的选择

✎ 一般而言，数据挖掘算法可以分为分布式数据挖掘算法和并行数据挖掘算法等。

① 分布式数据挖掘算法：

① 分布式数据挖掘算法适合数据垂直划分的算法、重视数据挖掘多任务调度算法。

② 并行数据挖掘算法：

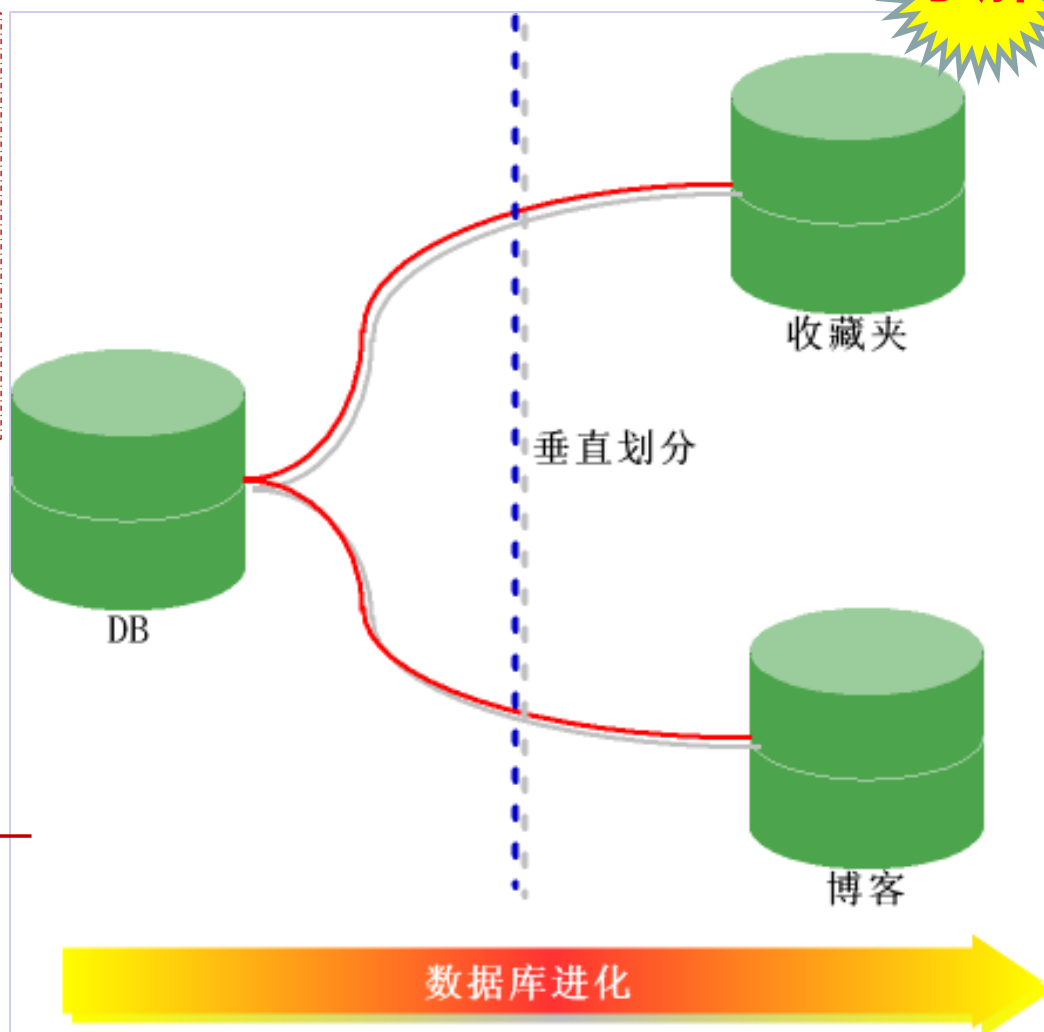
① 并行数据挖掘算法适合数据水平划分、基于任务内并行的挖掘算法。

- ✎ 云计算技术可以认为是物联网应用的一块**基石**，能够保证分布式并行数据挖掘，高效实时挖掘；
- ✎ 云服务模式是数据挖掘的普适模式，可以保证挖掘技术的共享，降低数据挖掘的应用门槛，满足海量挖掘的要求。

一、物联网的数据挖掘

❗【问题】随着访问压力的增加，读写操作不断增加，数据库的压力绝对越来越大，可能接近极限，这时可能人们想到增加从服务器，做什么集群之类的，可是问题又来了，数据量也快速增长，怎么办？

❗【回答】可以考虑对读写操作进行分离，按照业务把不同的数据放到不同的库中。其实在一个大型且臃肿的数据库中表和表之间的数据很多是没有关系的，或者更加不需要 (join) 操作，理论上就应该把他们分别放到不同的服务器。例如用户的收藏夹的数据和博客的数据库就可以放到两个独立的服务器。这个就叫**垂直划分**。



当博客或者收藏夹的数据不断增加后，应该怎么办，这就引出了另外一个做法，叫**水平划分**。

一、物联网的数据挖掘

水平划分



- ✘ 手机之家的架构是：水平切分—对数据进行水平分割。
- ① 最好分到同一个数据库。
- ② 一种已经证明是切实可行的方案：主表 + 辅表。
- ③ 有3种类型：主表不打散、主表打散无辅表、主表打散有辅表。
- ④ 但对程序员来说，TA看到的只是一张表，不妨称之为虚表(逻辑表)?，这张虚表实际上可能是由N张实表(物理表)组成的。
- ⑤ 可以按照业务来和环境，把数据分到不同的数据库中。
- ⑥ 另外，大型数据库，还可以做读写分离等。

- ✘ 则把一个表的数据划分到不同的数据库，两个数据库的表结构一样。怎么划分，应该根据一定的规则，可以根据数据的产生者来做引导，上面的数据是由人产生的，可以根据人的id来划分数据库。然后再根据一定的规则，先获知数据在哪个数据库。
- ✘ 其实很多大型网站都经历了数据库垂直划分和水平的划分的阶段。其实这个可以根据经验来确定。
- ✘ 以刚才的博客为例，数据可以根据userid的奇偶来确定数据的划分。把id为基数的放到A库，为偶数的放B库。
- ✘ 这样通过userId就可以知道用户的博客的数据在哪个数据库。其实可以根据 $userId \% 10$ 来处理。还可以根据著名的HASH算法来处理。

二、数据挖掘技术在物联网中的需求

4. 应用领域

✎ 物联网数据挖掘分析应用通常都可以归纳为预测和寻证分析两大类。

① **预测 (Forecasting)** : 主要用在 (完全或部分) 了解现状的情况下, 推测系统在近期或者中远期的状态。

✎ 例如:

- ① 在**智能电网**中, 预测近期扰动的可能性和发生的地点;
- ② 在**智能交通系统**中, 预测拥阻和事故在特定时间和地点可能发生的概率;
- ③ 在**环保体系**中, 根据不同地点的废物排放, 预测将来发生生物化学反应产生污染的可能性。

② **寻证分析 (Provenance Analysis)** : 当系统出现问题或者达不到预期效果时, 分析它在运行过程中哪个环节出现了问题。

✎ 例如:

- ① 在**食品安全**应用中, 一旦发生质量问题, 需要在食品供应链中寻找相应证据, 明确原因和责任;
- ② 在**环境监控**中, 当污染物水平超标时, 需要在记录中寻找分析原因。

✎ **数据挖掘技术在物联网中的需求**

- ① 精准农业;
- ① 市场营销
- ① 智能家居
- ① ...

- ① 金融安全
- ① 产品制造和质量监控
- ① 互联网用户行为分析

二、数据挖掘技术在物联网中的需求

4. 应用领域

(1) 精准农业

- ✎ 通过植入土壤或暴露在空气中的传感器监控土壤性状和环境状况。
- ✎ 数据通过物联网传输到远程控制中心，可及时查清当前农作物的生长环境现状和变化趋势，确定农作物的生产目标。
- ✎ 通过数据挖掘的方法，可以知道：环境温度湿度和土壤各项参数等因素是如何影响农作物产量的，如何调节它们才能够最大限度地提高农作物产量



二、数据挖掘技术在物联网中的需求

4. 应用领域

(2) 市场营销

- ✎ 利用数据挖掘技术通过对用户数据的分析，可以得到关于顾客购物取向和兴趣的信息，从而为商业决策提供依据
- ✎ 数据库营销(Database Marketing)
 - ① 通过交互式查询、数据分割和模型预测等方法来选择潜在的顾客以便向它们推销产品
 - ① 预测采用何种销售渠道和优惠条件，使得用户最有可能被打动
- ✎ 货篮分析(Basket Analysis)
 - ① 通过分析市场销售数据(例如POS数据库)来发现顾客的购买行为模式



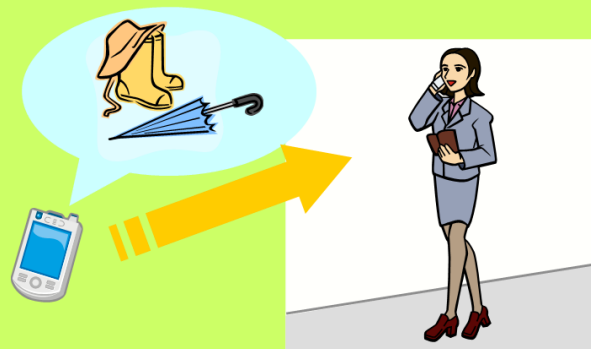
二、数据挖掘技术在物联网中的需求

4. 应用领域

(3) 智能家居

✎ 以获取天气信息为例：一方面，智能设备随时关注气象信息，并针对雨天发出报警提醒；另一方面，另外一些智能终端会随时跟踪主人的行踪，并通过数据挖掘方法由主人的历史行动特征数据预测他的去向

✎ 一旦预测到主人要出门，那么就在合适的时候由相应的智能终端提醒他不要忘记带雨伞。例如，如果主人在门口，就将由安装在门上的智能设备向他发出提醒，如果在车内，则由车载计算机发出提醒



二、数据挖掘技术在物联网中的需求

4. 应用领域

(4) 金融安全

- ✎ 由于金融投资的风险很大，所以在进行投资决策时，需要通过对各种投资方向的数据进行分析，以选择最佳的投资方向。数据挖掘可以通过对已有数据的处理，找到数据对象之间的关系，然后利用学习得到的模式进行合理的预测；
- ✎ 金融欺诈识别主要是通过分析正常行为和诈骗行为的数据和模式，得到诈骗行为的一些特性，这样当某项业务记录符合这样的特征时，识别系统可以向决策人员提出警告。



二、数据挖掘技术在物联网中的需求

4. 应用领域

(5) 产品制造和质量监控

- ✎ 随着科技进步，制造业已不是简单的手工劳动，而是集成了多种先进科技的流水作业。在产品的生产制造过程中常常伴随有大量的数据，如产品的各种加工条件或控制参数(如时间、温度等)。通过各种监控仪器收集的这些数据反映了每个生产环节的状态，对生产的顺利进行起着这关重要的作用。
- ✎ 通过数据挖掘对数据进行分析，可以得到产品质量与这些参数之间的关系，从而能获得针对性很强的建议以改进产品质量，而且有可能发现新的更高效节约的控制模式，为厂家带来丰厚的回报



二、数据挖掘技术在物联网中的需求

4. 应用领域

(6) 互联网用户行为分析

- 随着中国互联网用户数的激增，用户在互联网上的行为分析逐渐引起了关注。例如，用户在上网的时候通常需要不停地从一个网页通过HTTP链接跳转到另一个网页
- 获取互联网用户访问模式带来的好处很多，首先可以辅助改善分布式网络系统的性能，例如在有高度相关的站点间提供快速有效的访问通道。其次，能够帮助更好地组织和设计网页，以及帮助改善市场营销策略(例如把广告放在合适的网页上)以更好地吸引客户的注意



第9章 物联网智能决策--9.4 物联网的其它智能化技术

一、物联网的其它智能化技术

✎ 针对物联网数据的智能处理，需要研究以下内容：

(1) 以融合和决策为目的海量数据的实时挖掘

基于物联网服务的需求，物联网中的数据挖掘应分为两个方面：

- ① 辅助常规决策的数据挖掘；
- ② 辅助数据融合的数据挖掘。

(2) 以情境感知为目的的不确定性建模和推理

- ① 针对数据本身的不确定性；
- ② 针对语义映射的不确定性；
- ③ 针对查询分析的不确定性。

✎ 物联网的智能化技术是将智能技术的研究成果应用到物联网中，实现物联网的智能化。

✎ 比如物联网可以结合智能化技术如人工智能等，应用到物联网中。

✎ 物联网的目标是：实现一个智慧化的世界，它不仅仅感知世界，关键在于影响世界，智能化的控制世界。物联网根据具体应用结合人工智能，可以实现智能控制和决策。

第9章 物联网智能决策--9.4 物联网的其它智能化技术

一、物联网的其它智能化技术

1. 物联网的智能化技术

✎ 人工智能一般有两种不同的方式：

- (1) 一种采用传统的编程技术，使系统呈现智能的效果，而不考虑所用方法是否与人或动物机体所用的方法相同，这种方法叫**工程学方法** (Engineering approach)；
- (2) 另一种是**模拟法** (Modeling approach)，不仅要看效果，还要求实现方法也和人类或生物机体所用的方法相同或相类似。

① 采用**工程学方法**，需要人工详细规定程序逻辑，在已有的实践中被多次采用。

从不同的数据源（就包含物联网的感知信息）收集的数据中提取有用的数据，对数据进行滤除以保证数据的质量，将数据经转换、重构后存入数据仓库或数据集市，然后寻找合适的查询、报告和分析工具和数据挖掘工具对信息进行处理，最后转变为决策。

② **模拟法**应用于物联网的一个方向是**专家系统**：是一种模拟人类专家解决领域问题的计算机程序系统，不但采用**基于规则的推理方法**，而且采用诸如**人工神经网络的方法与技术**。根据专家系统处理的问题的类型，把专家系统分为**解释型、诊断型、调试型、维修型、教育型、预测型、规划型、设计型和控制型**等类型。

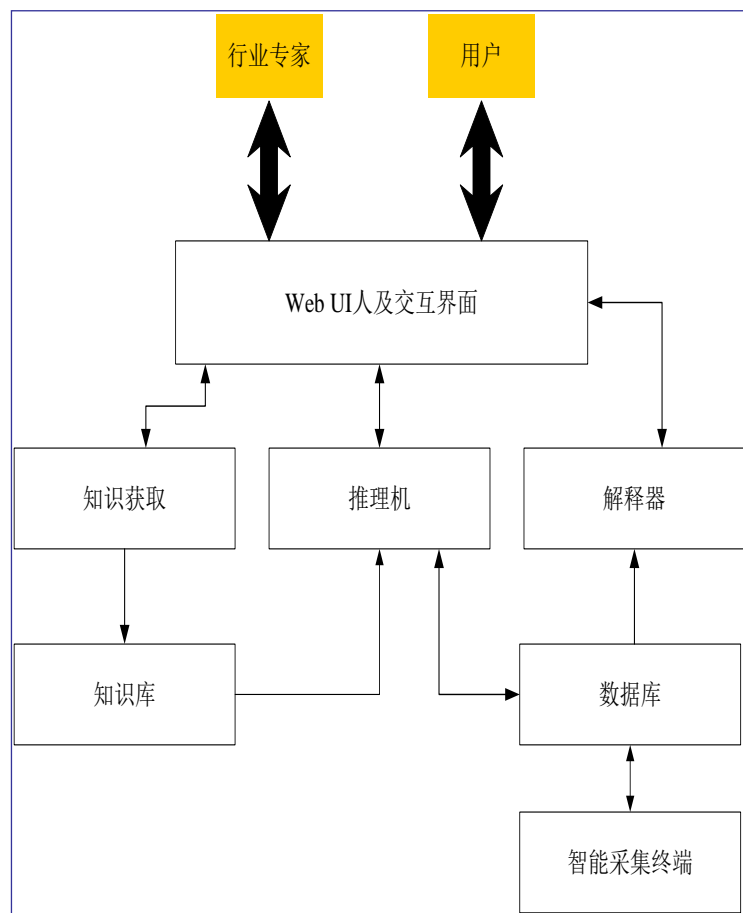
③ 另外一个方向为**模式识别**，通过计算机用数学技术方法来研究模式的自动处理和判读，如用计算机实现模式（文字、声音、人物、物体等）的自动识别。计算机识别的显著特点是速度快、准确性和效率高，识别过程与人类的学习过程相似，可使物联网在**“识别端”——信息处理过程的起点就具有智能性，保证物联网上的每个非人类的智能物体有类似人类的“自觉行为”。**

第9章 物联网智能决策--9.4 物联网的其它智能化技术

一、物联网的其它智能化技术

2. 物联网专家系统

✎ 物联网专家系统是指在物联网上的一类具有专门知识和经验的计算机智能程序系统或智能机器设备，**通过网络化部署专家系统来实现物联网数据的基本智能处理**，对用户**提供智能化的专家服务功能**。



① 物联网专家系统可实现对多用户的专家服务，其决策数据来源于物联网智能终端的采集数据。

② **智能终端采集**的数据提交到物联网应用数据库，数据反映了当前问题求解状态的集合。

③ **推理机**是实施问题求解的核心执行机构，是对知识进行解释的程序，它按照一定策略对找到的知识进行解释执行，并把结果记录到数据库中。

④ **解释器**用于对求解过程作出说明，并回答问题的提问。

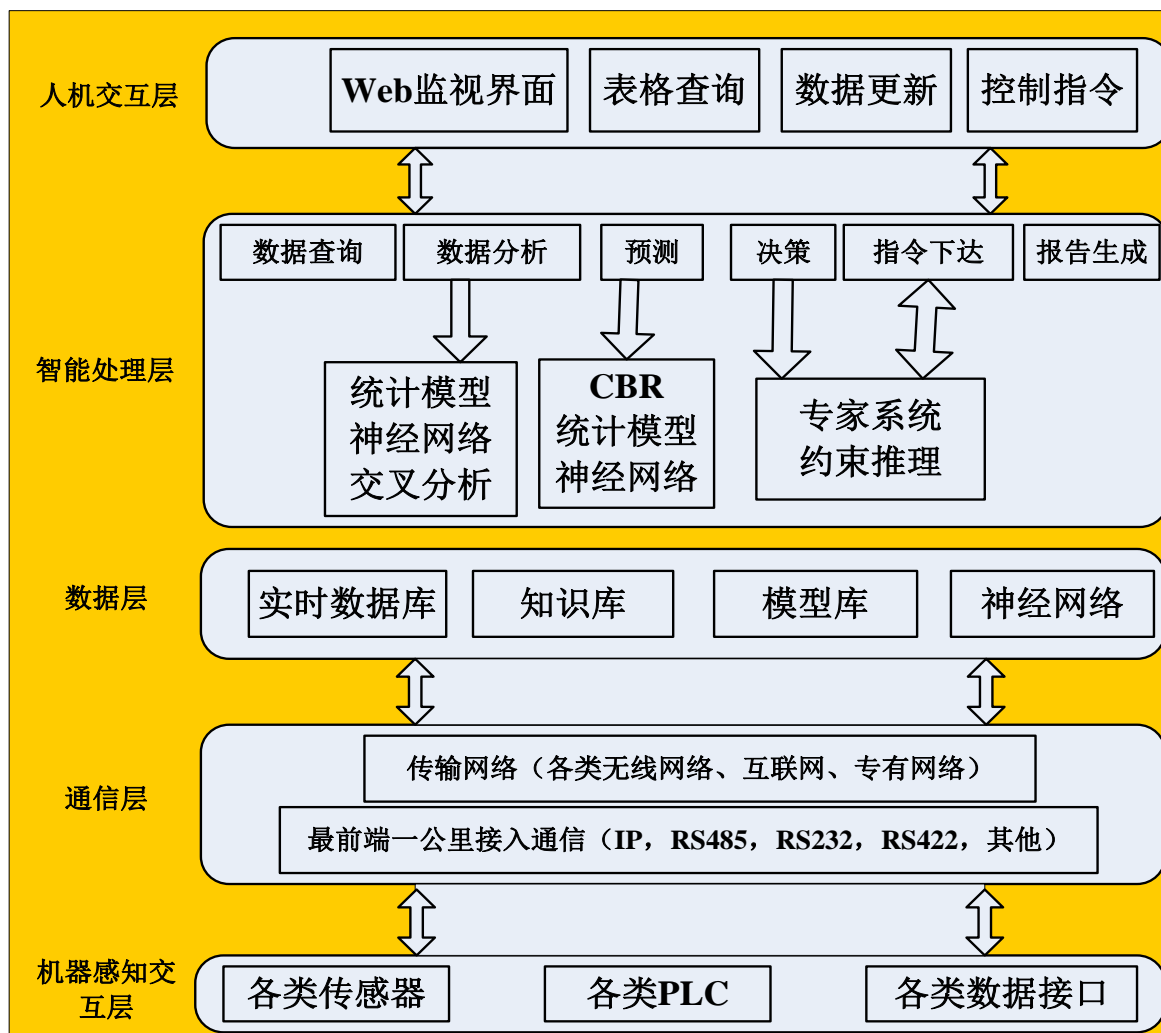
⑤ **知识库**是问题求解所需要的行业领域知识的集合，包括基本事实、规则等信息。

⑥ **知识获取**负责建立、修改和扩充知识库，是专家系统中把问题求解的各种专门知识从专家头脑中或者其它知识源转换到知识库的重要机构。在物联网中引入专家系统使物联网对其接入的数据具有分析判断并提供决策依据的能力，使物联网实现初步的智能化。

第9章 物联网智能决策--9.4 物联网的其它智能化技术

一、物联网的其它智能化技术

3. 物联网的智能化处理



❗ 物联网的智能化主要体现在**数据层**和**智能处理层**。

- ❶ **数据层**主要包括实时数据库、知识库、模型库、神经网络、历史数据库。
- ❷ **智能处理层**包括数据查询、数据分析、预测、决策、指令下达、报告生成等信息处理。

❗ 其中数据分析需要统计模型、神经网络、交叉分析等智能分析工具。预测需要统计学模型、神经网络等手段支持。

第9章 物联网智能决策--9.4 物联网的其它智能化技术

一、物联网的其它智能化技术

4. 物联网的含义

✎ 物联网主要解决物品与物品(Thing to Thing,T2T)，人与物品 (Human to Thing,H2T)，人与人(Human to Human,H2H)之间的互连。但是与传统互联网不同的是,H2T是指人利用通用装置与物品之间的连接,从而使得物品连接更加的简化，而H2H是指人之间不依赖于PC而进行的互连。因为互联网并没有考虑到对于任何物品连接的问题，故我们使用物联网来解决这个传统意义上的问题。物联网顾名思义就是连接物品的网络,许多学者讨论物联网中，经常会引入一个M2M的概念,可以解释成为人到人(Man to Man)、人到机器(Man to Machine)、机器到机器从本质上而言，在人与机器、机器与机器的交互，大部分是为了实现人与人之间的信息交互。——国际电信联盟(ITU)的定义