

流程工厂设计中的规则挖掘与智能预测

温蕊¹⁾, 唐卫清^{1,2)}, 苏智勇³⁾

¹⁾(南京理工大学计算机科学与工程学院 南京 210094)

²⁾(中国科学院计算技术研究所 北京 100190)

³⁾(南京理工大学自动化学院 南京 210094)

(wenrui@163.com)

摘要: 流程工厂设计涉及不同的工程应用背景及专业知识, 大量隐含的设计规则尚未提炼, 因此建模过程存在效率低、模型质量不高等问题. 结合流程工厂模型的领域特性, 提出一种用于挖掘工厂设计规则的频繁类型树模型. 该模型以大量实例工厂模型中各构件的属性及拓扑连接关系为依据构造频繁类型树; 在后续模型的构建过程中, 通过对该树的查询为当前工厂设计提供实时的智能预测. 实验结果表明, 文中模型能够挖掘到不同工程应用背景下的隐含建模知识, 在未来模型的构建过程中, 利用这些知识所提供的智能预测可以大幅缩小构件选择范围, 有效地提升流程工厂模型的构建效率及设计质量.

关键词: 流程工厂建模; 结构分析; 关联挖掘; 构件预测

中图分类号: TP391

Rule Mining and Intelligent Prediction in Process Plant Design

Wen Rui¹⁾, Tang Weiqing^{1,2)}, and Su Zhiyong³⁾

¹⁾(School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094)

²⁾(Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100190)

³⁾(School of Automation, Nanjing University of Science and Technology, Nanjing 210094)

Abstract: Process plant design involves a variety of engineering backgrounds and specialized knowledge. With abundant latent design rules not being extracted, the existing modeling methods present the disadvantages of low efficiency and deficient model quality. In that sense, a Frequent-Type-Tree model for mining process plant design rules is proposed. The model constructs a Frequent Type Tree by analyzing component attributes and topological correlations in plant models and provides real-time intelligent predictions by querying the tree for future design. Experimental results show that the proposed model can mine valuable latent modeling knowledge in various application backgrounds. The mined knowledge is then used to shrink ranges of component selection by a large margin while future modeling efficiency and design quality is improved significantly.

Key words: process plant modeling; structural analysis; association mining; component prediction

流程工厂模型的典型构建方式是设计人员在设计系统平台所展示的三维空间中根据用户需求

及实际工程需要, 从工程数据库中选择并逐一添加构件, 所以流程工厂模型最终是由设计人员来

收稿日期: 2014-04-20; 修回日期: 2014-07-21. 基金项目: 国家自然科学基金(61300160). 温蕊(1989—), 女, 博士研究生, 主要研究方向为 CAD&CG; 唐卫清(1965—), 男, 博士, 研究员, 博士生导师, CCF 高级会员, 主要研究方向为制造业信息化、CAD&CG、计算机支持的协同设计; 苏智勇(1981—), 男, 博士, 副教授, 硕士生导师, CCF 会员, 主要研究方向为计算机视觉、图像视频处理、计算机图形学.

决定和改进的. 这种设计模式存在 2 个弊端:

1) 模型复杂度的不断增大将导致建模过程中的疏忽与错误愈发难以发现, 这直接影响了最终模型的质量. 假设模型构件数量为 n , 各构件平均与 m 个其他构件拓扑相连, 由于校验单个构件的设计正确性需通过其拓扑连接关系来加以判断, 故校验整个模型的复杂度为 $O(nm)$. 在实际应用中, 工厂模型动辄包含数以万计的构件, 所以任何细小的错误都可能需花费大量的时间和精力去校验.

2) 工程应用背景广泛, 建模过程涉及不同的专业知识和大量的工程约束, 因此这种设计模式完全依赖于用户的具体需求和设计人员的专业知识及工作经验. 工程应用背景的不同导致工厂设计规则亦有所不同. 由于实际工程应用中包含大量虽存在但尚未提炼成文的设计规则, 所以设计人员专业知识及工作经验的缺乏会导致常识性错误时有发生. 可见, 人机交互的实时性和正确性是影响设计人员工作效率及模型设计质量的重要因素^[1].

智能预测是提高人机交互的实时性和正确性的方法之一. 以使用搜索引擎查询信息为例, 根据用户输入的关键词, 搜索引擎会实时地预测可能符合用户搜索意图的词条, 这种智能预测在一定程度上可以加快用户的查询速率, 同时保证输入的正确性. 在流程工厂领域, 倘若对历史模型中各类型构件之间隐含的关联关系及工厂设计规则进行分析与提炼, 则在后续的建模过程中, 当设计人员添加某类型构件时, 系统就可以利用已提炼出的规则实时地提供合适的候选构件类型. 这样不仅缩小了设计人员在工程数据库中选择目标构件的范围, 提高了设计效率; 同时, 由于历史模型是经过人工校验的, 其质量可以得到保证, 因此借鉴历史建模经验所提供的设计方案一方面可以使当前模型与历史模型保持一致的设计风格, 另一方面对于缺乏经验的设计人员可以有效避免常识性错误的发生, 保证了构建模型的质量.

数据挖掘技术可以很好地为智能预测提供支持, 其中关联挖掘的任务是发现隐藏在数据项之间的关联模式^[2], 挖掘到的规则可以用于预测未知的关联关系. 流程工厂中很多构件经常以组合的形式成对出现, 因此可以将关联挖掘的相关思想应用于流程工厂设计中的智能预测^[3].

本文将数据挖掘与流程工厂建模相结合, 通过关联挖掘的基本思想及策略, 结合流程工厂的领域特性, 从模型的拓扑结构入手, 提出一种用于

挖掘工厂设计规则的频繁类型树(frequent-type tree, FT-Tree)模型. 首先分析历史模型中各构件的属性及拓扑连接信息, 并以此为依据构造 FT-Tree, 获取工厂设计规则; 然后通过衡量各构件类型和关联规则的支持度与可信度来对 FT-Tree 进行剪枝, 确定最终的 FT-Tree; 最后系统通过对 FT-Tree 的实时查询选择合适的构件推荐给设计人员, 实现后续工厂建模中的智能预测, 有效地提高模型的构建效率及设计质量.

1 关联挖掘

Agrawal 等^[4]于 1993 年首先提出了挖掘顾客交易数据库中项集间的关联规则问题, 之后诸多研究人员对关联规则的挖掘问题进行了大量研究^[5-8], 主要集中于对原有算法的效率优化.

关联规则定义如下^[9]: 设 $I=\{i_1, i_2, \dots, i_m\}$ 是二进制文字的集合, 其中的元素称为项; 记 D 为交易 T 的集合, T 为项的集合, 且 $T \subseteq I$, 关联规则表示为 $X \rightarrow Y$, 其中 $X \subseteq I, Y \subseteq I$, 且 $X \cap Y = \emptyset$. 关联规则的支持度(support)和置信度(confidence)是规则兴趣度的 2 种度量^[10-11], 它们分别反映所发现规则的有效性和确定性. 假设关联规则 $X \rightarrow Y$ 有支持度 s 和置信度 c , 则表示在交易 D 中, 有 $s\%$ 的交易包含 $X \cup Y$, $c\%$ 的交易包含 $X \cap Y$. 一般情况下, 如果支持度 s 和置信度 c 满足最小支持度和最小置信度, 则认为该关联规则是有趣的^[12]. 所有支持度大于最小支持度的项集称为频繁项集.

Apriori 算法^[4]基于频繁项集的 2 个基本性质: 1) 如果一个集合是频繁项集, 则它的所有子集都是频繁项集; 2) 如果一个集合不是频繁项集, 则它的所有超集都不是频繁项集; 通过反复扫描数据库, 逐层搜索频繁项集. FP-Growth 算法^[5]通过构造一个扩展的前缀树结构来压缩数据记录, 使得挖掘频繁项集只需要扫描 2 次数据记录, 而且不依赖于生成候选集合, 所以相比 Apriori 算法效率较高. 可以看出, 关联挖掘是通过项集的共现来发现项与项之间隐藏的关联规则.

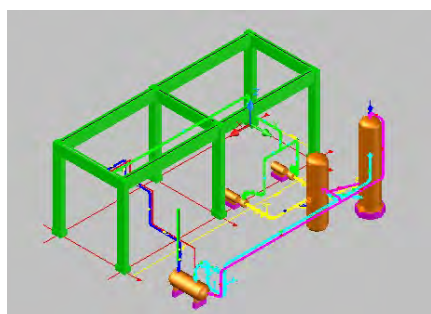
2 流程工厂模型

2.1 基本元素

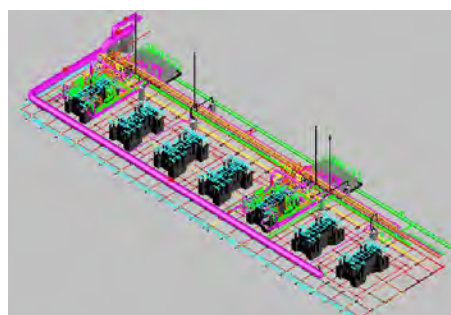
流程工厂是用来制造化学或物理制成品的反应容器、管线及其支撑的集合, 主要由设备、管子、

管件、阀门及仪表等基本构件组成^[13]. 工厂 CAD 模型中的基本构件通常都已经标准化、系列化, 但构件数量巨大、种类繁多. 除上述基本构件外, 还有针对不同工程应用背景的专业构件类型, 如电器、土建等. 因此很多工程 CAD 软件系统都配有专用的工程数据库管理系统, 不同的实际工程应

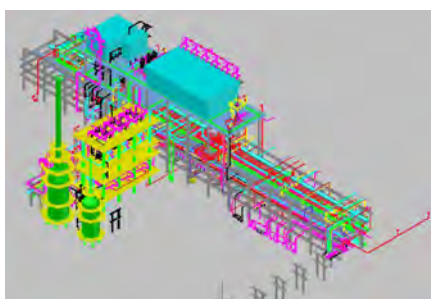
用背景使用其专业的工程数据库. 图 1 给出了 4 个具有代表性的流程工厂模型示意图. 建立与修改三维模型的过程不仅为工程人员提供了一种直观的三维设计手段, 而且其中记录的信息是生成各种工程图表的基础, 精确的工程图表可以帮助设计人员更好地修正错误、指导施工.



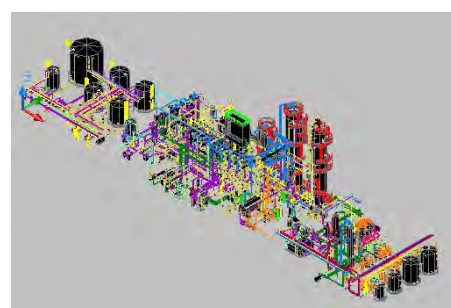
a 一个简单流程工厂模型图



b 某化工尿素生产线模型图



c 某汽油醚化项目模型图



d 某化工过氧化钠生产线模型图

图 1 流程工厂模型示例

2.2 拓扑结构

流程工厂设计的重点在于设计对象的结构和拓扑描述, 要求准确表达基本构件在三维空间中的位置及相互关系, 因此拓扑信息是流程工厂模型的核心所在^[14]. 表 1 统计了图 1 中所示各模型的元件与拓扑连接关系的数量, 可见流程工厂模型所包含的元件数量巨大, 整体拓扑结构复杂. 如果不能正确地表达管网模型的拓扑连接关系, 将会给模型的后续处理(如图纸的生成)造成不必要的困难. 工程设计人员在建立软模型的同时将大量的设计约束隐含于基本构件的连接与匹配之中, 这些约束是自动生成工厂图表的基础, 因此基本构件之间拓扑关系的建立与维护是流程工厂建模的关键问题.

由于设计系统的差异, 流程工厂拓扑结构的表示方法也不尽相同. 国内外使用较多的 2 种流程工厂拓扑结构表示方法为基于对偶点和智能线的表示方法和基于分支结构的表示方法. 基于对偶

点和智能线的工厂模型拓扑结构表示方法中, 对偶点用于描述管件之间的拓扑连接; 智能线描述通过将管件、管子抽象为三维线段或圆弧从而建立管件、管子之间的关系约束^[14]. 一个简单的对偶点与智能线的示例如图 2 所示, 其中黑色圆点表示对偶点, 虚线表示智能线.

表 1 模型数据统计

模型名称	元件数量	拓扑连接关系总数
图 1a	308	602
图 1b	9 795	19 044
图 1c	22 596	45 373
图 1d	41 569	83 359

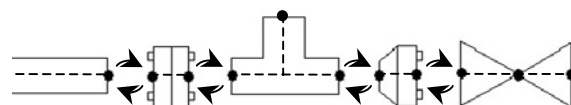


图 2 基于对偶点和智能线的管线示例

基于分支结构的工厂模型拓扑结构表示方法赋予了每个管道多个分支, 分支下面是具体的管件, 如图 3 所示. 分支与管道的不同在于分支只有 2 个端点, 而管道的端点个数由其分支数量决定. 所有的分支必须有起点和终点; 分支的方向必须是管道的流向; 分支中的管件顺序也同样重要, 它决定了管道的最终走向.

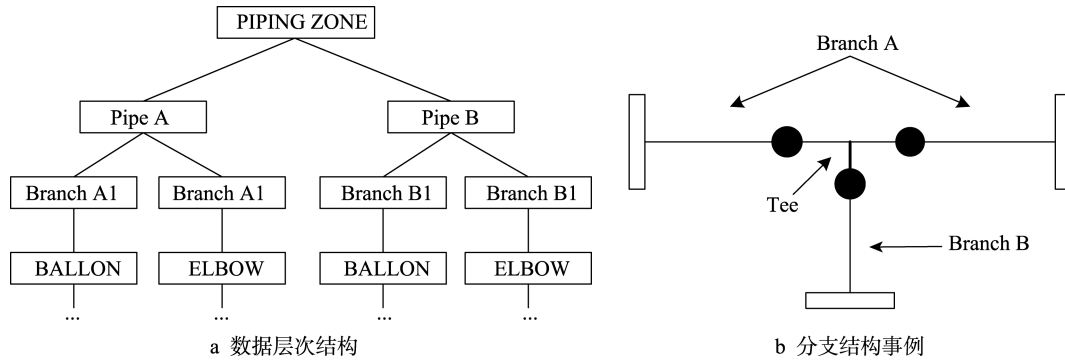


图 3 基于分支结构的管线示例

3 基于 FT-Tree 的流程工厂设计规则挖掘

为了提高设计人员的工作效率及所构建模型的质量, 本文以关联关系挖掘的相关思想为基础, 结合流程工厂模型的领域特性, 提出一种 FT-Tree 模型, 用于流程工厂设计的智能预测.

3.1 概念定义

为了更直观地阐述工厂模型中构件类型之间关联规则的挖掘问题, 本文给出如下定义:

定义 1. 设 M 为当前工程应用背景下的历史工厂模型, 将其表示为如下四元组: $M = (T, C, f, \Sigma_C)$. 其中, T 表示构件类型 t 的集合. C 表示构件 c 的集合. f 表示 M 中构件类型 t 与构件 c 的对应关系, 有 $f: c \rightarrow t$. Σ_C 表示 M 中各构件之间的拓扑连接关系, 有 $\Sigma_C = \{(c_i, c_j, A_{ij}) \mid c_i \text{ and } c_j \text{ is linked, } c_i \in C, c_j \in C, 1 \leq i, j \leq n\}$, 其中, A_{ij} 为构件 c_i 和 c_j 之间的连接属性, 如流向、材料、管路等级约束等; $n = |C|$, 为 M 中 C 的数量.

定义 2. 若有类型 $t \in T$, 与 t 有关联关系的集合 Q 为

$$Q(t) = \{(c_i, A_i) \mid \forall (c, c_i, A_i) \in \Sigma_C, f(c) = t, 1 \leq i \leq n\}.$$

定义 3. 若有类型 $t \in T$, $P(t)$ 为类型 t 在 M 中出现的次数, 有

$$P(t) = \{c_i \mid \forall c_i \in C, f(c_i) = t, 1 \leq i \leq n\}.$$

不同的工厂 CAD 系统对于流程工厂拓扑结构的表示方法虽不同, 但其本质都是相同的, 即流程工厂模型是结构化的数据模型, 模型中不同类型的构件按照一定的拓扑结构及关联属性组合在一起表达一个固有的、符合实际工程应用的功能. 所以, 可以根据历史工厂模型中构件之间的拓扑连接情况, 对各构件类型之间的关联关系进行分析.

定义 4. 若有类型 $t_i, t_j \in T$, $P(t_i, t_j, A_{ij})$ 为 M 中类型 t_i 与 t_j 以属性 A_{ij} 拓扑相连的次数, 有

$$P(t_i, t_j, A_{ij}) = |\{(c_i, c_j, A_{ij}) \mid \forall (c_i, c_j, A_{ij}) \in \Sigma_C, f(c_i) = t_i, f(c_j) = t_j\}|$$

定义 5. 若有类型 $t_i, t_j \in T$, 关联规则 $(t_i, A_{ij}) \rightarrow t_j$ 表示: 若有构件 $c_i \in C$, $f(c_i) = t_i$ 且属性为 A_{ij} , 则 c_i 一定与 c_j 以属性 A_{ij} 拓扑相连, 其中 $f(c_j) = t_j$, $c_j \in C$. 有如下公式^[8-9]:

$$S(t_i) = \frac{P(t_i)}{n} \tag{1}$$

$$R((t_i, A_{ij}) \rightarrow t_j) = \frac{P(t_i, t_j, A_{ij})}{|Q(t_i)|} \tag{2}$$

其中, $S(t_i)$ 表示在 M 中类型 t_i 的支持度, $R((t_i, A_{ij}) \rightarrow t_j)$ 表示规则 $(t_i, A_{ij}) \rightarrow t_j$ 的可信度.

3.2 FT-Tree 模型

在工程 CAD 系统中不同的工程应用背景使用其专业的工程数据库, 且数据库中的类型名称字段是设计人员在建模过程中用于判别所选构件是否为目标类型构件的重要依据^[15], 如图 4 所示. 所以本文提出的 FT-Tree 以构件的类型名称作为树中节点的主要标志; 有父子关系的两树节点表示它们所标志的类型之间存在拓扑连接关系, 且连接属性包括流向、材料、管路等级约束等; 除根节点外树中各节点均有独立的计数器, 用以记录该节点

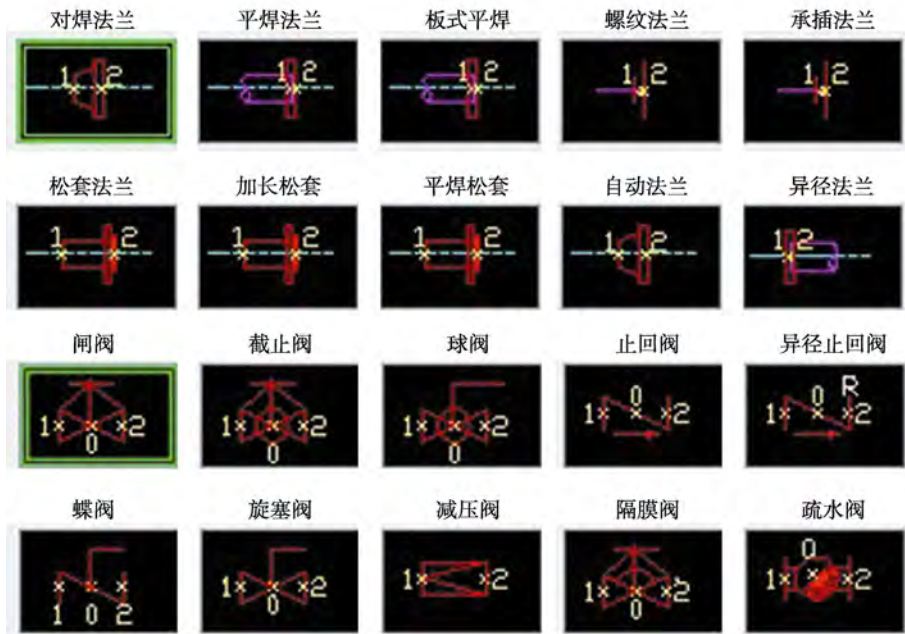


图 4 构件及对应类型名称示例

在当前关联关系下出现的频率，从而可以通过计算构件类型的支持度和构件之间连接关系的可信度，来判断该拓扑连接关系在当前工程应用背景下的可利用性。

3.2.1 构造

设 FT-Tree 的根节点为 r ，相应工程应用背景下的历史工厂模型中各构件的集合为 C ，构件之间的连接关系集合为 Σ_C ，构件 $c_i, c_j \in C$ ， $L(c_i)$ 是表示所有与 c_i 拓扑相连的构件 c_j 及它们之间的连接属性 A_{ij} 的集合。

FT-Tree 的具体构造过程如下：

Step1. 获取未遍历构件 $c_i \in C$ 在相应工程数据库下的所属类型名称 t_i 。

Step1.1. 若 t_i 存在于 r 的子节点中，那么将 r 的子节点中标志为 t_i 的节点的计数器增 1，执行 Step2；

Step1.2. 若 t_i 不存在于 r 的子节点中，那么添加一个以 r 为父节点、标志为 t_i 的节点，并将该新增节点的计数器初始化为 1，执行 Step2。

Step2. 对 $L(c_i)$ 中所有与 c_i 相连接的构件 c_j 及它们之间的连接属性 A_{ij} 执行如下操作：

Step2.1. 若 c_j 的类型 t_j 存在于标志为 t_i 的节点子节点中且 t_i 与 t_j 之间的连接属性为 A_{ij} ，表明已经出现过当前情况下的拓扑连接关系，那么将这个标志为 t_j 的子节点的计数器增 1；若 $L(c_i)$ 的遍历完成则执行 Step3，否则执行 Step2。

Step2.2. 若 c_j 的类型 t_j 不存在于标志为 t_i 的节点子节点中，表明还未出现过当前情况下的拓扑连接关系，那么为标志为 t_i 的节点添加一个标志为 t_j 的子节点，

同时该父子节点之间的连接属性标志为 A_{ij} ，并将 t_j 的计数器初始化为 1；若 $L(c_i)$ 的遍历完成则执行 Step3，否则执行 Step2。

Step2.3. 若 c_j 的类型 t_j 存在于标志为 t_i 的节点子节点中，但该父子节点之间的连接属性标志不为 A_{ij} ，表明 t_i 与 t_j 之间还未出现过连接属性为 A_{ij} 的拓扑连接关系，那么为标志为 t_i 的节点添加一个标志为 t_j 的子节点，同时该父子节点之间的连接属性标志为 A_{ij} ，并将 t_j 的计数器初始化为 1；若 $L(c_i)$ 的遍历完成执行 Step3，否则执行 Step2。

Step3. 若对 C 中所有构件的遍历完成，则终止；否则，转 Step1，继续遍历。

构造的 FT-Tree 如图 5 所示，其中虚线表示两节点属性相连。FT-Tree 的根节点 r 为所有构件类型共同的父节点；第二层为类型层 (type layer)，包含了历史模型中所有的构件类型；第三层为频繁类型层 (frequent type layer, FT layer)，包含了 type layer

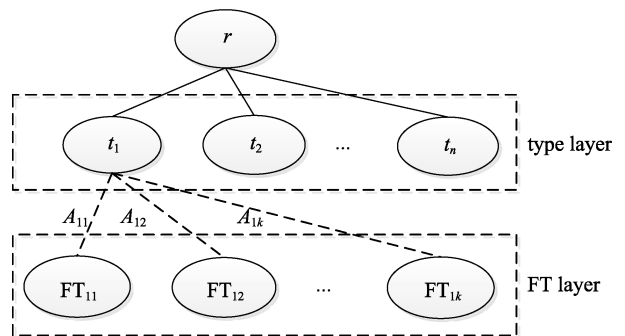


图 5 FT-Tree 示意图

中各节点的频繁类型(frequent type, FT)子节点. type layer 中的节点 t_i , t_i 在 FT layer 中的子节点 FT_{ij} 及其连接属性 A_{ij} 构成了关联规则: $(t_i, A_{ij}) \rightarrow FT_{ij}$. 图 6 所示为 FT-Tree 的一个简单实例.

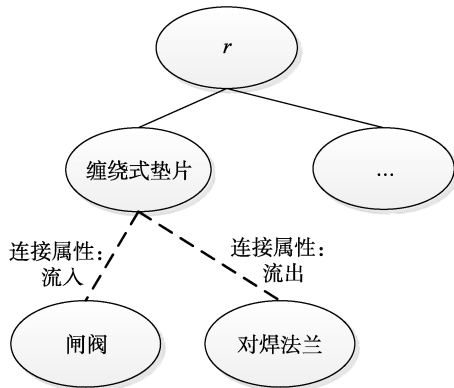


图 6 FT-Tree 简单实例

3.2.2 剪枝

随着应用范围的扩大及历史模型数量的增加, 构件类型的数量也会随之增多, FT-Tree 会由于更新而导致其规模不断增大, 所以需要适当地对 FT-Tree 进行剪枝, 以保证在模型构建时的智能预测中对 FT-Tree 的实时查询是高效快捷的.

通过衡量各构件类型的支持度(式(1)), 可以判断各类型在该工程应用背景下的重要程度: 支持度越高, 表明该构件类型越重要; 通过衡量频繁类型 FT 的可信度(式(2)), 可以判断该构件类型的 FT 的可信程度: 可信度越高, 表明该 FT 的可信性越强, 当前规则的可利用性越高.

本文采用式(2)将 FT layer 中各节点的 FT 按可信度的值从大到小进行排序, 各取前 $k(k>0)$ 个 FT 作为最终频繁类型. 若 n 表示历史流程工厂模型中所包含的构件类型的数量, 那么构造 FT-Tree 的空间复杂度为 $O(kn)$, 查询 FT-Tree 的复杂度为 $O(n+k)$.

4 智能预测

智能预测是基于以下基本思想: 对历史流程工厂模型 $M = (T, C, f, \Sigma_C)$, 若其构成的 FT-Tree 中存在关联规则: $(t_x, A) \rightarrow t_y$, 且 $t_x, t_y \in T$, 则在构建流程工厂模型时, 当添加任意构件 x 时, 若有 $f(x) = t_x$, 且关联属性为 A , 则根据上述关联规则, 预测有构件 y 与构件 x 拓扑相连, 且 $f(x) = t_y$.

在工厂设计进行之前, 设计人员首先设置阈

值 $i(i > k)$, 表示期望预测的构件类型个数. 在之后的实时预测中, 以构件 c 作为输入, 首先获取 c 的类型名称 t 及关联属性 A ; 然后在相应 FT-Tree 的 type layer 中查找是否存在标志为 t 的节点, 若存在, 则筛选出拓扑连接属性为 A 的所有子节点; 最后在筛选的子节点中选择 i 个可信度值最大的节点所标志的类型名称的构件作为候选类型, 将其推荐给设计人员. 智能预测的流程如图 7 所示.

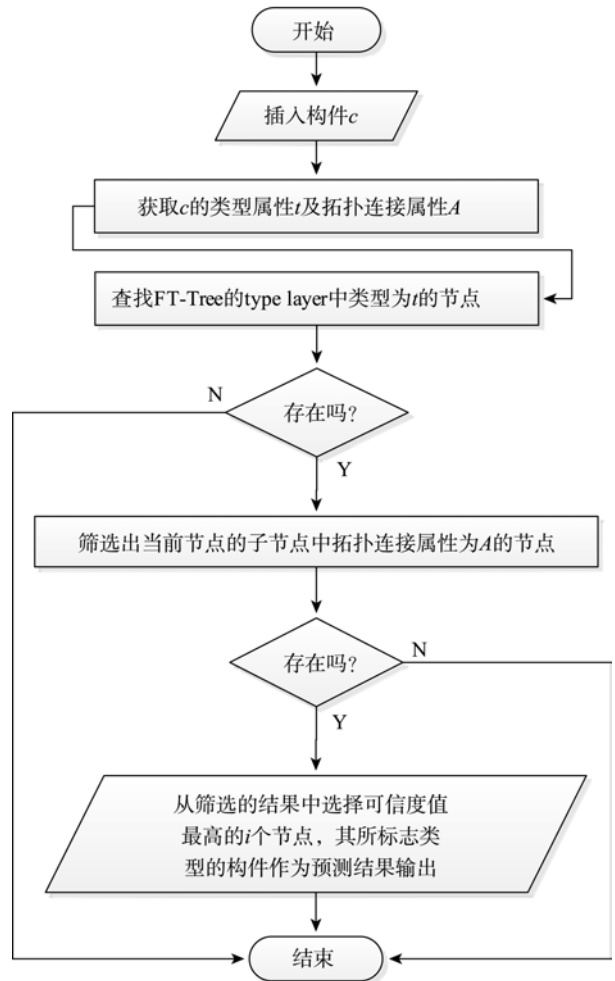


图 7 智能预测流程图

数据库中的类型名称字段是设计人员在建模中选择元件的重要依据, 而 FT-Tree 是对历史模型中各类型构件间拓扑连接关系进行学习并提炼的结果, 对树的实时查询可以将符合当前输入的相关联构件提供给设计人员进行选择. 这样不仅缩小了建模过程中构件的选择范围, 有助于提高设计人员的工作效率, 同时由于借鉴了历史设计经验, 使缺乏经验的设计人员可以避免常识性错误的发生, 使有经验的设计人员可以在多种

设计方案中选择最优方案进行设计,保证了模型的质量.

5 实验及结果分析

5.1 预处理

为了有效地构造 FT-Tree,需要对历史模型进行预处理.模型的预处理及各构件的表示遵循如下原则:1)所有历史模型中的构件类型均有唯一的标志;2)每个模型中各构件均有唯一的标志;3)每个构件均对应唯一的类型;4)每个构件最多有一个拓扑连接关系构件集合,且各连接关系有属性标志,若构件无关联关系、独立的存在,则无关联关系构件集合.

实验以 PDSOFT[®] 3DPiping 与 Visual Studio 2008 作为实验平台,对图 1 所示的不同工程数据库背景下的流程工厂模型进行如上所述的预处理.之后,按照第 3 节的 FT-Tree 构造算法分别为图 1 中各模型建立 FT-Tree.表 2 分别统计了预处理后各模型的构件类型和关联关系集合的数量,以及其 FT-Tree 的 type layer 层节点的数量,可以看出,FT-Tree 的 type layer 节点的数量少于相应模型的类型数量,原因在于 FT-Tree 剔除了模型中无拓扑

连接关系的构件类型.

表 2 预处理结果及 FT-Tree 信息统计

模型名称	类型数量	关联关系集合数量	type layer 节点数量
图 1a	22	307	21
图 1b	137	9 638	132
图 1c	187	22 474	181
图 1d	774	41 137	770

5.2 实验结果

本文实验选择管道的流向作为拓扑连接属性 A ,有如下情况: $A=0$ 时表示流向未定; $A=1$ 时表示流入; $A=2$ 时表示流出; $A=3$ 时表示顺流, $A=4$ 时表示逆流.在实际工厂模型的构建过程中,利用所属工程应用背景下历史工厂模型所构造的 FT-Tree,根据当前添加的构件类型及拓扑连接属性对下一步将要添加的构件的类型进行智能预测,按照第 4 节的智能预测方法进行预测实验,预测结果及所涉及的关联规则如表 3~5 所示.限于篇幅限制,表 3~5 分别展示了图 1b~1d 模型所属工程应用背景下使用频率(即支持度)最高的 4 种构件类型,并依据其 FT-Tree 中的支持度和可信度由大到小依次展示各自可信度最高的 5 种预测结果(即频繁类型 FT)

表 3 图 1b Top 5 预测结果及关联规则展示

规则 $(t_i, A_{ij}) \rightarrow t_j$		FT-Tree		
输入 t_i	属性 A_{ij}	预测结果 t_j /共现频率	$S(t_i)$	$R((t_i, A_{ij}) \rightarrow t_j)$ (估算预测正确率)
无缝钢管	3	无缝钢管/2	0.213 783	1.0
	4	无缝钢管/2		1.0
	2	截止阀/211		0.111 404
	1	截止阀/133		0.070 557
	0	支架/15		0.062 762
截止阀	1	无缝钢管/211	0.033 180	0.653 251
	2	无缝钢管/133		0.411 765
	2	仪表泡/80		0.247 678
	2	石棉橡胶垫片/51		0.157 895
	1	石棉橡胶垫片/50		0.154 799
高压无缝钢管	1	高压螺纹法兰/44	0.026 953	0.171 875
	2	高压螺纹法兰/43		0.166 023
	0	支架/3		0.034 483
	1	止回阀/4		0.015 625
	2	止回阀/4		0.015 625

存在以下特殊情况:某类构件的频繁类型 FT 的数量小于用户设置的预测候选构件个数

表 4 图 1c Top 5 预测结果及关联规则展示

规则 $(t_i, A_{ij}) \rightarrow t_j$			FT-Tree	
输入 t_i	属性 A_{ij}	预测结果 t_j /共现频率	$S(t_i)$	$R((t_i, A_{ij}) \rightarrow t_j)$ (估算预测正确率)
无缝钢管	3	无缝钢管/12	0.213 489	1.0
	4	无缝钢管/12		1.0
	0	支管座/644		0.578 097
	2	承插焊法兰/605		0.131 123
	2	弯头/587		0.127 221
缠绕式垫片	1	闸阀/951	0.121 836	0.345 567
	2	闸阀/951		0.345 316
	2	对焊法兰/693		0.251 634
	1	对焊法兰/673		0.244 549
	2	承插焊法兰/636		0.230 937
弯头	1	无缝钢管/587	0.089 219	0.291 315
	2	无缝钢管/572		0.287 582
	0	弹簧支吊架/18		0.088 67
	2	弯头/10		0.005 028
	1	弯头/10		0.004 963
对焊法兰	1	缠绕式垫片/693	0.064 790	0.480 916
	2	缠绕式垫片/673		0.466 389
	2	仪表泡/44		0.030 492
	1	柔性石墨板垫片/21		0.014 573
	2	柔性石墨板垫片/21		0.014 573

表 5 图 1d Top 5 预测结果及关联规则展示

规则 $(t_i, A_{ij}) \rightarrow t_j$			FT-Tree	
输入 t_i	属性 A_{ij}	预测结果 t_j /共现频率	$S(t_i)$	$R((t_i, A_{ij}) \rightarrow t_j)$ (估算预测正确率)
无缝钢管	4	无缝不锈钢管/2	0.155 861	1.0
	0	对焊接管台/5539		0.837 607
	3	无缝不锈钢管/1		0.5
	3	无缝碳钢管/1		0.5
	2	90° 承插焊弯头/534		0.086 478
非金属垫片	0	三通调节阀/6	0.094 566	0.5
	0	对焊法兰/6		0.5
	2	对焊法兰/1724		0.439 348
	1	对焊法兰/1724		0.398 981
	2	球阀/1288		0.328 236
对焊法兰	0	非金属垫片/6	0.086 314	1.0
	1	非金属垫片/1724		0.484 406
	2	非金属垫片/1556		0.437 447
	2	柔性石墨复合垫片/ 85		0.023 897
	1	柔性石墨复合垫片/ 76		0.021 354
90° 弯头(R=1.5D)	2	仪表泡/7	0.075 008	0.002 285
	2	螺纹管帽/1		0.000 326
	1	不定长金属管/1		0.000 322

及所涉及的关联规则。其中,估算预测正确率表示智能预测结果为正确结果的概率,从可信度的定义(式(2))可以看出,可信度实际上给出了模型预测的正确率。

5.3 可行性分析

通过分析图 1b~1d 所示的工厂模型,同时结合实际的工程应用发现,表 3~5 所示的预测结果是有效且符合实际工程要求的。例如表 3 中结果显示:在一定工程应用背景下,当添加截止阀且连接属性为 1 和 2 时,无缝钢管与其拓扑连接的概率分别为 0.653 251 和 0.411 765,表明截止阀与无缝钢管所构成规则可用性较高,而在实际的工程应用中,由于工艺流程的需要,无缝钢管与截止阀构成的组合正是流程工厂管道设计中重要的组成部分。又如表 5 中所示,规则“(对焊法兰, 0)→非金属垫片”,其可信度为 1.0,表明相应工程应用背景下的模型中,若有对焊法兰且流向未定(即为 0)时,必定存在非金属垫片与其拓扑相连。可见,由于借鉴了历史建模经验,可以避免一些常识性错误的发生,有效地提高构建模型的质量。

同时,本文所提出的 FT-Tree 模型较适合于流程工厂设计中的智能预测。通用的关联挖掘算法只关注项集中各项之间的共现频率,而流程工厂模型中构件之间的关联关系需要满足一定的工程属性约束(如流向等),因此理想的预测结果除了满足与当前构件有较高的共现频率外,还需符合一定的连接属性:如表 4 所示,当 t_i 为无缝钢管、 A_{ij} 为 3 时,与无缝钢管的共现频率为 12;当 A_{ij} 为 2 时,与承插焊法兰的共现频率为 605。由于连接属性的不同,使得无缝钢管与无缝钢管的共现频率虽低于无缝钢管与承插焊法兰,但是其估算预测正确率却高于后者。原因在于历史模型中,当 t_i 为无缝钢管、 A_{ij} 为 3 时,与其相连的构件只有无缝钢管,因此其估算预测率为 1.0;而当 A_{ij} 为 2 时,与无缝钢管相连的构件除承插焊法兰外,还有弯头等其他构件存在。因此在后续的智能预测中,当 t_i 为无缝钢管、 A_{ij} 为 3 时,由于连接属性的限制,预测构件为该条件下可信度最高的无缝钢管,而非共现频率最高的承插焊法兰;当 A_{ij} 为 2,预测则为承插焊法兰等符合当前条件的构件。所以构件之间的连接属性是工厂设计预测中不可忽略的部分,也是 FT-Tree 不同于其他关联挖掘模型的重要因素。FT-Tree 模型利用流程工厂模型的特点,将构件的

出现频率与连接属性相结合来对工厂设计规则进行挖掘与提炼,使得基于 FT-Tree 的智能预测在候选构件的正确性及合理性方面具有优越性。

采用本文提出的模型构建方式可以有效地提高设计人员的工作效率。假设当前工程数据库中构件的类型数量为 t ,在现有的建模方式下,构建模型的复杂度为 $O(tm)$;采用本文方法,通过对 FT-Tree 的实时查询来辅助设计人员构建模型,若将智能预测的构件类型个数置为 k ,则构建模型的复杂度为 $O(kn)$ 。而在实际应用中,构件的类型数量 t 成百上千,而本文所设置的 k 值在一般情况下只需一位数,由于缩小了构件选择的范围,所以一定程度上提高了设计人员的工作效率。

6 结 语

本文分析了当前流程工厂设计模式存在的不足,首次进行了针对流程工厂设计中智能预测的相关研究,并将数据挖掘思想与流程工厂设计相结合,利用流程工厂模型的特点提出了一种频繁类型树模型 FT-Tree。该模型通过遍历相同工程应用背景下历史工厂模型中的构件,根据各自属性及构件之间的拓扑连接关系构造 FT-Tree,将支持度和可信度作为关联关系重要性的衡量准则,以此对 FT-Tree 进行剪枝,保证对树的检索效率;在后续的工厂建模中,根据当前添加构件的属性实时查询 FT-Tree,从而对将要添加的构件类型进行预测。FT-Tree 模型通过对历史建模知识的学习将不同工程应用背景下隐含的设计规则进行分析与提炼,可以帮助设计人员及施工人员更好地了解模型;同时使得计算机辅助的智能预测缩小了模型构建过程中用户选择构件的范围,操作较现有建模方法更加便捷快速,可以有效地提高模型的构建效率及设计质量。

在未来的工作中,可以通过挖掘工厂模型中更多的隐含设计知识,更加全面细致地为工厂模型的构建提供智能预测,不断提高人机交互的工作效率。

参考文献(References):

- [1] Jia Hongwei, Tang Weiqing. Human-computer interaction resource model for collaborative design[J]. Computer Integrated

- Manufacturing Systems, 2007, 13(12): 2351-2357(in Chinese)
(贾红妮, 唐卫清. 协同设计人机交互资源模型的研究[J]. 计算机集成制造系统, 2007, 13(12): 2351-2357)
- [2] Cai Weijie, Zhang Xiaohui, Zhu Jianqiu, *et al.* Survey of association rule generation[J]. Computer Engineering, 2001, 27(5): 31-33(in Chinese)
(蔡伟杰, 张晓辉, 朱建秋, 等. 关联规则挖掘综述[J]. 计算机工程, 2001, 27(5): 31-33)
- [3] Padhy N, Mishra P, Panigrahi R. The survey of data mining applications and feature scope[J]. International Journal of Computer Science, Engineering and Information Technology, 2012, 2(3): 43-58
- [4] Agrawal R, Imieliński T, Swami A. Mining association rules between sets of items in large databases[C]//Proceedings of the ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 1993: 207-216
- [5] Han J W, Pei J, Yin Y W. Mining frequent patterns without candidate generation[C]//Proceedings of the ACM SIGMOD International Conference on Management of Data. New York: ACM Press, 2000: 1-12
- [6] Geng L Q, Hamilton H J. Interestingness measures for data mining: a survey[J]. ACM Computing Surveys, 2006, 38(3): Article No.9
- [7] Ceglar A, Roddick J F. Association mining[J]. ACM Computing Surveys, 2006, 38(2): Article No.5
- [8] Agrawal R, Srikant R. Fast algorithms for mining association rules in large databases[C]//Proceedings of the 20th International Conference on Very Large Data Bases. San Francisco: Morgan Kaufmann, 1994: 487-499
- [9] Sun J P. Mining attribute association in query predicates for access path generation[C]//Proceedings of the ACM Research in Applied Computation Symposium. New York: ACM Press, 2012: 106-111
- [10] Lin X K, Mukherji A, Rundensteiner E A, *et al.* PARAS: a parameter space framework for online association mining[J]. Proceedings of the VLDB Endowment, 2013, 6(3): 193-204
- [11] Zaki M J. Scalable algorithms for association mining[J]. IEEE Transactions on Knowledge and Data Engineering, 2000, 12(3): 372-390
- [12] Kubat M, Hafez A, Raghavan V V, *et al.* Itemset trees for targeted association querying[J]. IEEE Transactions on Knowledge and Data Engineering, 2003, 15(6): 1522-1534
- [13] Huang Xiaojian. Research on fast rendering techniques in engineering CAD[D]. Beijing: Institute of Computing Technology, Chinese Academy of Sciences, 2002(in Chinese)
(黄晓剑. 工程 CAD 中的快速绘制技术研究[D]. 北京:中国科学院计算技术研究所, 2002)
- [14] Dai Xiaofeng. The research on AEC CAD modeling technology based on enhanced graph and multi-state model[D]. Beijing: Institute of Computing Technology, Chinese Academy of Sciences, 2000(in Chinese)
(戴肖锋. 基于扩展图与多态模型的工程 CAD 建模技术研究[D]. 北京:中国科学院计算技术研究所, 2000)
- [15] Hao Jing, He Tao, Li Shicai, *et al.* Data management method in cooperative plant design system[J]. Journal of Computer-Aided Design & Computer Graphics, 2006, 18(4): 563-567(in Chinese)
(郝 静, 何 涛, 李士才, 等. 协同式工厂设计系统中的数据管理方法[J]. 计算机辅助设计与图形学学报, 2006, 18(4): 563-567)