

飞机装配指令语义元素智能挖掘方法

Intelligent Mining Methods of Semantic Elements of Aircraft Assembly Instructions

西北工业大学 俞琳 赵磊 李原 张杰

[摘要] 为了更好地实现对飞机装配指令典型语义元素的提取、组织、重用,提出了飞机装配指令典型语义元素智能挖掘方法。对飞机装配指令数据元素进行预处理,实现自然语言层面的语句结构标准化划分,通过对语义粒度的分析,将装配指令划分为典型指令语义元素、标准化装配指令语句、典型指令模板;采用语块分析理念对装配指令进行深层次的语义概念剖析,基于隐马尔可夫模型(Hidden Markov Model, HMM),实现对装配指令的语块标注。进一步采用 Apriori 方法,完成对装配指令典型操作语义元素的智能挖掘。

关键词: 装配指令 语义元素 智能挖掘 Apriori

[ABSTRACT] Intelligent mining method of typical semantic elements of aircraft assembly instruction is proposed to better achieve the extraction, organization and reuse of the elements. Standardized partition of statement structure is realized at the natural language level through preprocessing the data elements of aircraft assembly instructions and these instructions are divided into semantic elements of typical instructions, standardized assembly instruction statements and typical instruction templates by analyzing semantic granularities. Based on Hidden Markov model, chunk labeling of assembly instructions is realized via a deep semantic concept analysis of these instructions under the guidance of the idea of chunk parsing. The intelligent mining of typical operational semantic elements of assembly instructions is completed by adopting the method of Apriori.

Keywords: Assembly instruction Semantic element Intelligent mining Apriori

飞机装配指令(Assembly Order, AO)设计是一项复杂而繁琐的工作,经验在规划过程中占据了重要地位。如何整理和提炼典型装配指令语义元素,对实现知识重用、规范装配指令规划过程和提升效率存在着重要意义。

现阶段, AO 设计已经累积了大量工艺经验和设计方法,装配操作语义也已被形式化为与特定工程活动相

对应的抽象知识集合,并在一定程度上实现了指令语义元素的存储、表示及重用。但是设计语言表达的多样化,语句结构的离散化,往往使得设计意图传递引起多重歧义,从而进一步需要对 AO 指令表述语义进行粒度化划分,对指令语句进行结构化、连续化处理,对指令语义元素进行形式化表述,更有利于对不同粒度的装配指令元素进行表示、存储、重组与继承。

目前,对产品设计知识表达已有多方面的研究。沈斌等^[1]采用面向对象技术对知识对象进行划分,建立产品设计的多类型、多层次知识表达模型;刘培奇等^[2]将产生式规则的前提、结论和处理分别用概念图表示,提出一种扩展式规则知识表示方法;魏丽等^[3]采用关键特性表示粗粒度、非完备的产品信息。但这些都过于粗略或过于详细地描述工艺知识,不适合从一个全局的角度来管理工艺知识。同时,对于知识的获取方法,国内外学者也做了大量研究。凌卫青等^[4]以功能驱动实现基于实例的产品设计知识获取方法;高伟等^[5]定义了工艺决策数据表,以 ID3 算法构建工艺决策树,提出一种基于概念学习系统(Concept Learning System, CLS)的工艺决策树学习算法;高伟等^[6]又从工艺重用的角度,利用聚类算法提取工艺知识术语,实现计算机辅助工艺规划(Computer Aided Process Planning, CAPP)智能化;薛俊芳等^[7]提出了基于粗糙集的零件合并专家知识获取方法等。上述研究都采用了不同的获取算法对知识进行发现和挖掘。然而,考虑到工艺知识的非规范性和不确定性,从知识粒度的角度对工艺知识进行形式化分析,规范在工艺规程中使用的知识,有利于对不平衡数据进行知识挖掘。同时在工艺知识粒(Process Knowledge Granule)分类的基础上,依照不同的使用需求对工艺知识粒进行操作,方便了工艺知识的自动获取,不仅满足工艺规程的生成,还适合工艺知识的复杂性特点。在一定程度上,粒计算是面向知识的计算,不同于数字计算,研究粒计算的主要问题是颗粒构造和描述问题^[8]。

通过以上分析可以发现,基于实例以及知识工程的装配指令设计方法,为装配指令典型语义元素的挖掘、表示、存储、调用提供了手段;但是,针对飞机装配指令语言描述结构化特点,缺乏对装配指令语义描述粒度的划分,不利于装配指令典型语义元素的规范化管理及聚

类分析,为装配指令描述语言的形式化、标准化设置了障碍;同时,针对飞机装配指令设计,缺乏装配指令典型操作序列的关联关系挖掘,不利于装配指令的快速设计。

因此,本文提出基于装配指令语块分析的概念,从飞机装配指令语义描述粒度着手,实现装配指令典型语义元素及装配指令操作行为关联关系的智能挖掘。

1 装配指令语义粒度划分

1.1 粒度的相关概念

(1)定义1——粒度。

粒度是指数据仓库的数据单位中保存数据的细化或综合程度的级别,数据细化的程度越高,粒度就越小,细化程度越低,粒度就越大。

(2)定义2——装配指令粒度。

装配指令粒度是语义对描述内容的细化程度,具有相似结构的语义描述对象的数据集合,能够反映装配指令语义关系的组织结构,是装配指令语义相似度分析与关联度分析的基础。

飞机装配工艺知识粒度可以从2方面理解和划分:其一,可以理解为装配工艺的设计过程中粒度的细化过程;其二,可以理解为装配指令包含工艺指令元素的粒度细化过程。飞机装配工艺设计过程中其粒度的划分过程如图1所示。飞机装配指令粒度是含有一定语义的装配指令信息的集合体,可以是一组参数化的指令语句,也可以是用来对单个指令语义元素内容进行概括的无歧义的装配指令术语。

1.2 语义粒度划分

飞机装配指令设计,装配任务设计过程及其粒度都进一步向精细化方向发展,需要对现有的“工艺组件划分→装配顺序规划→AO→工序→工步”的装配过程分解模式实施拓展,使其分解粒度既能满足一般装配过程的层进式表述需求,又能够适应仿真环境下装配操作和动作的细分需求,从而支撑飞机装配指令的可视化、标准化设计。对工艺知识的信息描述情况是随着对工艺信息描述的详细、规范程度的深入而逐步加深的,其对应的描述粒度是逐渐缩小的。

(1)典型指令语义元素。

典型指令语义元素(PK_c)是飞机装配指令中涉及的粒度最小的语义元素,是可以由简短工艺词组对单个工艺语句内容进行概括的、满足一定支持度要求和无歧义的工艺术语,是企业长期实践中固定下来的工艺设计中的固定用语。工艺术语来源于各种规范、标准、工艺手册等数据资源,根据企业的自身情况和专业特点,为提高工艺文件的质量和工艺语言的专业化,通过对工艺术语进行统一和标准化,形成了典型指令语义元素。

飞机装配指令中典型指令语义元素可表示为:

$$PK_c = \{PK_{c1}, PK_{c2}, \dots, PK_{cn}\} \quad (1)$$

(2)典型装配指令语句。

典型装配指令语句(PK_s)是装配指令(AO)文件中抽取出的一条完整的、可以说明装配指令内容的单个语句,从装配任务角度讲,可能是可执行的一条装配工序或工步操作,并涉及到装配操作、装配对象、装配资源等典型装配工艺元素。典型装配指令语句是在飞机装配工艺设计环节中常用的、相对固定的、具有较高支持度且经过标准化和规整化的工艺语句。标准化装配指令语句知识粒可表示为:

$$PK_s = \{PK_{s1}, PK_{s2}, \dots, PK_{sm}\} \quad (2)$$

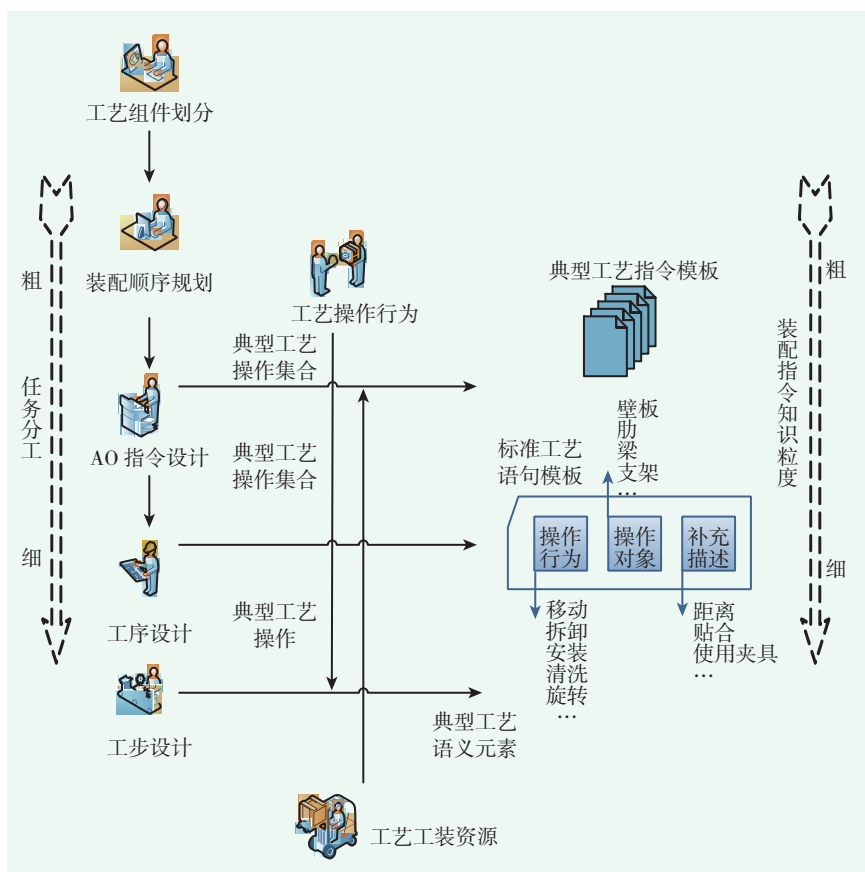


图1 飞机工艺知识粒度划分过程

Fig.1 Division of aircraft process knowledge granularities

(3) 典型指令模板。

典型指令模板(PK_i)是飞机装配指令中具有经验化、实例化的前后序操作的段落式工艺指令,即通过大量的工艺文件分析整理,可以明确发现在某些典型装配操作任务之前或之后,必然进行下一步相关工序/工步的操作。在装配指令或装配工艺规程中形成的固化的典型装配指令实例,从语义元素角度分析,可能是若干典型装配工艺元素或标准化装配指令语句的集合。通过典型指令模板的执行,能够完成一定的装配任务。典型指令模板可表示为:

$$PK_i = \{PK_{i1}, PK_{i2}, \dots, PK_{in}\} \quad (3)$$

装配指令典型语义元素、典型装配指令语句、典型指令模板三者之间有着密切的关联关系,根据装配任务分工由装配宏规划(工艺组件划分,装配顺序规划)到装配工艺微规划(AO指令设计、工序/工步设计),其典型工艺知识粒度也由粗变细。不仅在知识粒度层面三者之间划分越来越细,同时还存在一定的组成关系,经过标准化形式化处理的典型指令语义元素,可以组成典型的装配指令语句或典型装配指令模板,典型装配指令语句的合集也构成了典型装配指令模板;反之,典型装配指令语句、典型工艺模板通过合理的语句结构划分,可以拆分为典型指令语义元素。亦可表示为:

$$PK_c \in PK_s \in PK_i \quad (4)$$

2 装配指令语块分析

2.1 语法结构分析

针对飞机装配指令设计语句,功能描述语言的主语一般隐性表示实时操作的装配工人。宾语一般表示零部件或装配资源。谓语一般分为2个层面:(1)相对主谓之间表示装配操作行为;(2)相对2个装配对象之间表示两者之间的装配约束关系。介词短语表示方法或位置关系。试以“打开卡板”这样一个词语为例。其中有2个基本单元“打开”和“卡板”,“打开”为动词,“卡板”为名词,因此从句法结构看是V+N的动宾结构;从语义关系结构看,前者是“行为动作”,后者是“受事”。

典型装配指令语句结构划分实例如图2所示。

2.2 语块标注

在基于统计的语块识别方法中,文本语块识别被看成是一个序列的预估问题。本文采用IOB2的标注集合来进行语块标注。在IOB2中,如果用 X 表示某个短语的话,那么,一个句子中的语块标识可能是下面3种符

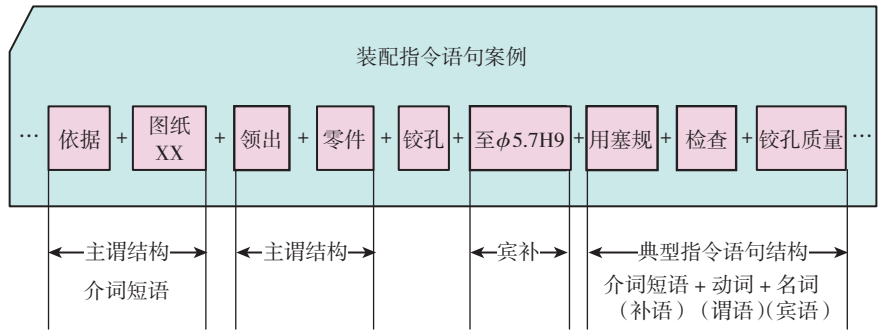


图2 典型装配指令语句结构划分实例

Fig.2 Structure division case of typical assembly instruction statements

号中的一种。

$B-X$ 为短语 X 的第一个词; $I-X$ 为短语 X 的非初始词; O 为不属于任何短语的词,即语块外部词。

在对语料库进行标注时,严格依照语块的定义及相关的标注规范进行。通过标注集合进行标注,可以在一定程度上反映出不同的词在句子中的组合关系,如句子中的2个相邻的词 $wb_i = \langle w_i, b_i \rangle$ 和 $wb_{i+1} = \langle w_{i+1}, b_{i+1} \rangle$,其中 w_i 表示语块对应的词语, b_i 表示语块的标注。

针对飞机装配指令分析的词块:

(1) 名词语块。

名词语块可以是一个名词或定中结构。定中结构包含一个中心名词和一个前置修饰成分。前置修饰成分包括:形容词、副词、名词等。

(2) 动词语块。

针对飞机装配指令设计语言,其动词语块主要指对装配行为加以副词修饰的装配操作,以及对装配操作行为的修饰说明。前置修饰部分主要是副词,后置修饰部分主要是名词、数词等。

(3) 介词语块。

通常,介词语块只包含介词本身,但对于一些固定搭配的介词短语如“在...中”,“在...上”,也被划分为一个介词语块。

本文将采用隐马尔可夫模型(Hidden Markov Model, HMM)对语义语块进行标记。HMM是一种统计技术,它起源于语音识别,可以用于词性标注,根据对训练数据的训练,得出一些概率值,根据这些概率值对相应的数据进行识别,对于一个装配指令操作序列:

$$W_j = w_i / ps_i, w_{i+1} / ps_{i+1}, \dots, w_j / ps_{j+1}$$

其中, ps_i 表示词 w_i 的词性,选择合适的语块标注序列:

$$B_j = b_i, b_{i+1}, \dots, b_j$$

使得 $P(B_{ij} | W_{ij})$ 达到最大,HMM作为一种统计分析模型,隐马尔可夫模型可由一个五元组构成:

$$HMM \in \{S, O, \Pi, A, B\} \quad (5)$$

其中, $S = \{s_1, s_2, \dots, s_N\}$ 为状态有限集合,其状态数量为

N ; $O=\{o_1, o_2, \dots, o_M\}$ 为状态输出符号集合,其输出的符号数为 M ; $\Pi=\{\pi_1, \pi_2, \dots, \pi_N\}$ 为初始状态概率矩阵,其中 π_i 是状态 s_i 作为初始状态的概率; $A=\{a_{i,j}\}$ 为状态转移概率矩阵,其中, $a_{i,j}$ 是从状态 s_i 转移到状态 s_j 的概率, $1 \leq i \leq N, 1 \leq j \leq N$; $B=\{b_{i,k}\}$ 为符号输出概率矩阵,其中, $b_{i,k}$ 是从状态 s_i 输出 O_k 的概率。 $1 \leq i \leq M, 1 \leq k \leq M$ 。

一个隐马尔可夫模型是一组有限的状态,其中的某一个状态可以以一定的概率转移到另外的状态,而且在转移时产生有限的输出,输出也将以一定的概率发生。在马尔可夫模型中,每一个状态都代表一个可观察的事件,而在隐马尔可夫模型中,状态本身是不能直接观察的,故称之为“隐马尔可夫模型”。

将语块分析过程看成是 Chunk 标注过程,就可以采用类似基于隐马尔可夫模型的词性标注方法来标注 Chunk。令 O 和 I 分别代表输出序列和输入序列,则语块标注问题可视为计算以下条件概率的极大值

$$\begin{aligned} \hat{O} &= \arg \max P(O|I) \\ &= \arg \max \frac{P(I|O)P(O)}{P(I)}, \end{aligned} \quad (6)$$

式中, $P(O|I)$ 是一致输入序列 I 的情况下,出现输出序列 O 的条件概率。式中符号 $\arg \max$ 表示通过考察不同的候选输出序列 O 来寻找使条件概率取得最大值的那个输出序列 \hat{O} 。

在进行语块标注的过程中,输入序列 I 为词序列 w_i 和词性序列 s_i 。输出序列 O 为语块的标注序列。由于语块标注的输入是词和词性序列的组合,使得模型的输入集合较大;而输出方面语块标记则过少,导致模型的输出集合不够丰富。因此,根据这样的输入输出训练得到的标注模型准确率不高。为了使得模型更准确,可以通过降低模型的输入规模,导入部分上下文信息到模型的输出部分,来增大输出的标记集合。

$$\arg \max \prod_{i=1, \dots, n} P(w_i, s_i | s_{i-1}, c_{i-1}) P(s_i, c_i | s_{i-1}, c_{i-1}), \quad (7)$$

其中,输入的信息为:词 w_i , 词性 s_i (i 为词的位置), 输出信息从单一的语块标记 c_i 变为 s_{i-1}, c_{i-1} , 使得输出集合增大。条件概率:

$$P(w_i, s_i | s_{i-1}, c_{i-1}) = N_1 / N_2, \quad (8)$$

其中, N_1 表示训练语料中单词 w_i 的词性为 s_i 时,被标记为 c_i 的次数; N_2 表示为词性为 s_i 被标记为 c_i 的次数。转移概率:

$$P(s_i, c_i | s_{i-1}, c_{i-1}) = N_3 / N_4, \quad (9)$$

其中, N_3 表示训练语料中词性 s_{i-1} , 语块标注为 c_{i-1} 与词性 s_i , 语块标注为 c_i 的连续次数; N_4 表示词性为 s_{i-1} , 语块标注为 c_{i-1} 的出现次数。

在基于 HMM 的语块标注中,可以采用 Viterbi 算法运用动态规划来搜索最优状态序列,其相关步骤如下。

STEP 1: 输入工艺语义(完成分词、词性标注等步骤),假定观察序列为 $O=\{o_1, o_2, \dots, o_i\}$ 。初始化数 ϕ 组,其中 $\phi_t(j)$ 存放到时刻 t 时 j 状态最优解的前一状态。

STEP 2: 初始化 $\delta_1(i) = \pi_i b_i(o_1)$, $\phi_1(i) = 0$, $1 \leq i \leq N$ 。

STEP 3: 递归。

$$\begin{aligned} \delta_t(j) &= \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}] b_{j, o_t}, 2 \leq t \leq T, 1 \leq j \leq N_t(i); \\ \phi_t(j) &= \arg \max_{1 \leq i \leq N} [\delta_{t-1}(i) a_{ij}], 2 \leq t \leq T, 1 \leq j \leq N_t. \end{aligned}$$

STEP 4: 终结。

$$P^* = \max_{1 \leq i \leq N} [\delta_T(i)],$$

$$s_T^* = \arg \max_{1 \leq i \leq N} [\delta_T(i)].$$

STEP 5: 最佳序列。

$$s_t^* = \phi_{t+1}(s_{t+1}^*), t = T-1, T-2, \dots, 1.$$

得到的序列 s_t^* 便是该句的词和词性条件下每个位置的最佳语块标记序列。至此,借助隐马尔可夫理论,就可以完成对飞机装配指令语块标注的任务。

3 装配指令语义智能挖掘

通过对飞机装配指令集进行语块分析标注,实现了装配指令语义元素的划分,可以进一步采用 Apriori 方法,最终完成对飞机装配指令典型工艺元素的数据挖掘。

Apriori 算法的基本思想:算法需要对数据集进行多步处理。首先,简单统计所有含一个元素项目集的出现频率,并找出那些不小于最小支持度的项目集,即一维频繁项目集,进一步循环处理直到再没有频繁项目集生成。循环过程是,第 k 步中,根据第 $k-1$ 步生成的 $(k-1)$ 维频繁项目集产生 k 维候选项目集,然后对数据库进行搜索,得到候选项目集的支持度,与最小支持度比较,从而找到 k 项频繁项目集。算法当候选 k 项集的集合 C_k 为空,即不能找到频繁 k 项集时结束。

对于飞机装配指令语义挖掘,给定一个事务集,即装配指令集:

$$AO = \{ao_1, ao_2, \dots, ao_n\}, \quad (10)$$

其中, ao_i ($1 \leq i \leq n$) 是每个事务的数据,称为数据项,装配指令是数据项;数据项集 $I = \{i_1, i_2, \dots, i_m\}$ 是 AO 中所有数据的集合,即 I 是所有工艺指令中曾经出现过的工序的集合, I 中的元素 i_j ($1 \leq j \leq m$) 代表工序。关联规则模型中每个事务含有 I 的一个子集,即每个工艺路线中包含有工序序列,这个工序序列是所有工序集合 I 的子集。关联规则是形如 $A \Rightarrow B$ 的逻辑蕴涵式,其中 $A \subset I, B \subset I$, 且 $A \cap B = F$ 。项集 A 在事务集中出现

的次占事务集中总事务的百分比称为项集的支持度。支持度 $\text{support}(A \Rightarrow B)$, 即 AO 中包含 $A \cup B$ 的百分比, 支持度是对关联规则重要性(或适用范围)的衡量, 因此可以以支持计数衡量支持度; 置信度 $\text{Confidence}(A \Rightarrow B)$, 即 AO 中包含 A 事务的同时也包含 B 事务的百分比, 置信度是对关联规则准确性的衡量。如果支持度和置信度都满足各自的阈值, 则 $A \Rightarrow B$ 可以看成是一个有意义的关联规则。获取典型工序序列就是从大量工艺指令数据中, 获得支持度和置信度分别大于最小支持度和最小置信度的关联性强度的工序序列。

采用 Apriori 算法对飞机装配指令典型工艺语义进行挖掘, 其挖掘步骤如下。

STEP 1: 生成频繁 1 项集 L_1 。

通过单趟扫描数据事务集 AO , 计算出各个 1 项集的支持度, 即扫描装配指令数据集, 计算出各个工序的支持度, 从而得到频繁 1 项集构成的集合 L_1 , L_1 就是频繁出现的工序的集合。

STEP 2: 连接步——生成潜在频繁工序序列集 C_k 。

潜在频繁集 C_k 是包含 k 个工序的候选工序序列集合, 频繁 k 项集 L_k 是包含 k 个工序的频繁出现的工序序列集合。为了产生频繁 k 项集构成的集合 L_k , 需要预先生成一个潜在频繁集 C_k , C_k 由连接运算 $L_{k-1} * L_{k-1}$ 得到, 运算 “*” 定义为:

$$C_k = L_{k-1} * L_{k-1} = \{X \cup Y \mid X, Y \in L_{k-1}, |X \cap Y| = k-2\} \quad (11)$$

式中, X, Y 是频繁 $(k-1)$ 项集 L_{k-1} 中的元素, 是包含了 $k-1$ 个工序的工序序列; $|X \cap Y|$ 描述 X 和 Y 中相同工序的个数; $X \cup Y$ 是同时包含 X 和 Y 的工序序列。

由于工序序列是有序关联的, 应作进一步规定, 若满足:

$$\begin{cases} p, q \in L_{k-1} \\ p = \{p_1, p_2, \dots, p_{k-2}, p_{k-1}\} \\ q = \{q_1, q_2, \dots, q_{k-2}, q_{k-1}\} \\ p_{i+1} = q_i \quad (1 \leq i \leq k-1) \end{cases} \quad (12)$$

则 $p \cup q = \{p_1, p_2, \dots, p_{k-2}, p_{k-1}\}$ 是潜在频繁 k 项集集合 C_k 的元素。这里的潜在频繁 k 项集的集合 C_k 是指有可能成为频繁 k 项集的项集组成的集合。

STEP 3: 剪枝步——减少 C_k 中的工序序列数据项。

由于 C_k 是 L_k 的超集, 可能有些元素不是频繁的。 C_k 很庞大, 将会带来巨大的计算量, 为减少 C_k 的规模, Apriori 遵从性质“任何非频繁的 $(k-1)$ 项集必定不是频繁 k 项集的子集”。所以, 当前在 k 项集的某个 $(k-1)$ 子集不是 L_{k-1} 的成员时, 则该潜在频繁项集不可能是频繁的, 可以从 C_k 中移去, 这就是 Apriori 的剪枝思想。

STEP 4: 计算 C_k 中各个工序序列的支持度。

通过单趟扫描数据事务集 AO , 计算出各个 C_k 中各个项集的支持度, 即通过扫描工艺路线数据集, 计算出每个工序序列的支持度。为了减少计算候选项集支持度所涉及的记录总数, 在每次计算 C_k 支持度的过程中, 给不包含 C_k 中的任何项集的事务打上删除标记, 在以后数据事务集 T 的扫描计数中不予考虑。这样, 计算候选项集支持度所涉及的记录数目将小于事务数据库中的实际记录数, 并且随着 k 值的增大, 这一差值也不断增大, 有效地降低了候选项集的计数速度, 提高了算法的效率。

STEP 5: 生成频繁 k 项集。

将 C_k 中不满足最小支持度的项剔除, 形成由频繁 k 项集构成的集合 L_k , 即将所有包含 k 个工序的工序序列集合中不满足最小支持度的工序序列剔除, 形成频繁出现的工序序列的集合。

STEP 6: 算法终止条件。

通过循环迭代, 重复 STEP 2~STEP 5, 直到不能产生频繁项集的集合(非空集合)为止, Apriori 算法求出所有满足最小支持度的频繁项集, 最后获得典型工序序列。

至此, 完成了飞机装配指令的典型工艺语义挖掘过程。首先, 进行飞机装配指令语料库的准备; 接着, 实现飞机装配指令语料库语义数据预处理, 主要包括语义粒度的划分, 语块分析, 基于隐马尔可夫的语块标注; 然后基于 Apriori 实现飞机装配指令典型工艺元素的智能挖掘。

4 应用验证

本文提出的飞机装配指令典型语义智能挖掘方法, 以某型机机翼组件装配指令为例, 进行实例分析, 对装配指令(AO)进行基于隐马尔可夫的语块化标注。其中, 该机型机翼组件装配指令手册包含字符 9034 个字符(包含标点符号, 不计空格), 其中有效字符 6272 个, 统计语块 1322 对。通过统计分析可知, 表示数量型修饰的数量词语块(MP)以及修饰性介词短语语块(PP)能达到更为准确的统计与标注, 在名词语块(NP)及动词语块(VP)方面, 其标注统计遗漏或错误之处基本表现为对专业术语语块的再划分, 如 [组合 /n 件 /n], [锉 /n 修 /v 壁板 /n] 等, 相对于飞机装配指令语句较为规范的固化语句模板, 其出错率在 5% 以下, 统计出的标注语块, 能够满足后期对典型指令语义元素做进一步的语义相似度及关联度分析。语块标识统计结果如表 1 所示。

为了进一步对装配指令操作语义元素进行关联数据挖掘, 提取典型装配操作指令, 如表 2 所示。

表1 语块分类标注结果

语块类别	统计数量 / 对	覆盖率 /%	正确率 /%
动词语块 (VP)	329	99.65	96.34
名词语块 (NP)	396	100	98.87
介词语块 (PP)	264	98.52	98.61
数量词语块 (MP)	198	100	100
连词语块 (CP)	135	100	100

表2 典型装配指令操作序实例

指令编号	操作对象	工步 1	工步 2	工步 3	工步 4	工步 5	工步 6
AO ₁	6 肋孔	扩孔	铰孔	检验	镗窝	检验	
AO ₂	其余孔	铰孔	检验	镗窝	检验		
AO ₃	5 肋孔	扩孔	铰孔	检验	镗窝	检验	
AO ₄	其余孔	铰孔	检验	镗窝	检验		
AO ₅	上壁板	钻孔	镗窝	清洗	镟头	涂料	
AO ₆	口盖孔	划孔	钻孔	扩孔	铰孔	检验	镗窝
AO ₇	口盖孔	划孔	钻孔	扩孔	铰孔	检验	镗窝
AO ₈	密封托板	钻孔	镗窝	清洗	铆接	涂料	

通过对 8 条典型装配指令操作序列分析,即事务集 $|AO|=8$, 设最小事务支持计数 $\text{Support}_{\min}=4$, 最小置信度 $\text{Confidence}_{\min}=60\%$, 则计算 L_2 频繁 2 项集频繁置信度为:

- 扩孔→铰孔: $\text{Confidence}=4/4=100\%$;
- 铰孔→检验: $\text{Confidence}=6/6=100\%$;
- 检验→镗窝: $\text{Confidence}=6/10=60\%$;
- 镗窝→检验: $\text{Confidence}=6/8=75\%$ 。

Apriori 方法循环类推, 计算 L_3 频繁 3 项集频繁置信度:

扩孔→铰孔→检验: $\text{Confidence}=(100\%+66.67\%)/2=83.34\%$; 铰孔→检验→镗窝: $\text{Confidence}=100\%$ 。

计算 L_4 频繁 4 项集频繁置信度为:

扩孔→铰孔→检验→镗窝: $\text{Confidence}=(100\%+83.34\%)/2=91.67\%$; 铰孔→检验→镗窝→检验: $\text{Confidence}=4/6=66.67\%$ 。

计算 L_5 频繁 5 项集频繁置信度为:

扩孔→铰孔→检验→镗窝→检验: $\text{Confidence}=(91.67\%+66.67\%)/2=79.17\%$ 。典型装配指令操作语义工步关联分析过程如表 3 所示。

5 结束语

针对装配指令语义粒度分析, 实现飞机装配指令语义描述粒度层面的定义及划分, 包含典型指令语义元素、典型指令语句、典型指令模板等, 进而实现飞机装配指令语义挖掘的数据预处理任务; 根据装配指令语义元素类别划分的需求, 提出了飞机装配指令语块式结构分

表3 典型装配指令操作语义工步关联分析过程

工步	扩孔	铰孔	检验	镗窝	钻孔	清洗	镟头	划孔	铆接	涂料
计数	4	6	10	8	4	2	1	2	1	2

↓

工步	扩孔	铰孔	检验	镗窝	钻孔
计数	4	6	10	8	4

↓

工步	扩孔→铰孔	铰孔→检验	检验→镗窝	镗窝→检验
计数	4	6	6	6

↓

工步	扩孔→铰孔→检验	铰孔→检验→镗窝
计数	4	6

↓

工步	扩孔→铰孔→检验→镗窝	铰孔→检验→镗窝→检验
计数	4	6

↓

工步	扩孔→铰孔→检验→镗窝→检验
计数	4

析方法, 并基于隐马尔可夫(HMM)理论, 实现飞机装配指令的语块标注; 基于 Apriori 方法, 实现飞机装配指令典型工序的语义挖掘。以装配指令语块标注数据为基础, 构造飞机装配指令集, 通过支持度分析及连接步, 剪枝步处理, 生成飞机装配指令典型操作序; 以某型飞机中外翼组件装配指令语料为例, 举例分析装配指令语块标注过程及典型工序语义挖掘过程, 验证了该方法的正确性。

参考文献

- [1] 沈斌, 官大. 产品设计知识重用研究. 计算机工程, 2006, 18(36): 186-188.
- [2] 刘培奇, 李增智, 赵银亮. 扩展产生式规则知识表示方法. 西安交通大学学报, 2004, 38(6): 587-590.
- [3] 魏丽, 郑联语. 概要工艺规划中关键特性的识别过程及方法. 计算机集成制造系统, 2007, 13(1): 147-152.
- [4] 凌卫青, 赵艾萍, 谢友柏. 基于实例的产品设计知识获取方法及实现. 计算机辅助设计与图形学学报, 2002, 14(11): 1014-1019.
- [5] 高伟, 胡晓兵. 基于 CLS 的决策型工艺知识发现算法. 四川大学学报(工程科学版), 2006, 38(2): 79-83.
- [6] 高伟, 殷国富, 戈鹏. 基于知识发现原理的制造工艺重用方法研究. 中国机械工程, 2003, 14(19): 1667-1680.
- [7] 薛俊芳, 向东, 邱长华. 基于粗糙集的零件合并专家知识获取方法. 计算机集成制造系统, 2007, 13(8): 1658-1664.
- [8] BARGIELA A, PEDRYCZ W. Granular computing: an introduction. Boston: Kluwer Academic Publishers, 2003.

(责编 圭一)