

产品设计过程中质量知识的匹配方法*

茅 健¹ 曹乃亮² 曹衍龙³ 杨将新³

(1. 上海工程技术大学机械工程学院, 上海 201620; 2. 中国科学院长春光学精密机械与物理研究所, 长春 130033;
3. 浙江大学现代制造工程研究所, 杭州 310027)

【摘要】 提出了一种产品设计过程中基于语义距离的设计质量知识的表示及匹配方法,将产品设计过程划分为5个阶段,构建产品设计质量信息模型。依照设计质量信息模型,对设计任务、设计人员和设计质量信息进行分析,以任务元、对象元、信息元来构建三者的形式化描述体系,并建立其数据结构。利用HNC(概念层次网络)知识库,基于语境框架文本相似度计算,并结合语义距离计算,实现了文本间相似度的精确量化和比较。以注塑机方案设计过程为例,实现了设计质量信息匹配,验证了该方法的有效性。

关键词: 产品设计 质量控制 知识匹配 文本相似度 概念层次网络 语义距离
中图分类号: TP391.1 **文献标识码:** A **文章编号:** 1000-1298(2012)01-0197-05

Matching Method for Quality Knowledge in Product Designing Process

Mao Jian¹ Cao Nailiang² Cao Yanlong³ Yang Jiangxin³

(1. College of Mechanical Engineering, Shanghai University of Engineering Science, Shanghai 201620, China
2. Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, Changchun 130033, China
3. Institute of Modern Manufacturing Engineering, Zhejiang University, Hangzhou 310027, China)

Abstract

The design quality information and matching method based on semantic distance was presented in product designing process. Product design process was divided into five phases, and then the model of product design quality information was built. According to the model, the design task, designer and design quality information were analyzed. The task meta, object meta and information meta were described and the data structure was built. Using hierarchical network of concepts (HNC) and combining with semantic distance calculation, the quantification and comparing of text similarity can be realized based on framework of context text similarity. Finally, the conceptual design of injection molding machine was taken as an example, the result showed the effectiveness of the proposed method.

Key words Product design, Quality control, Knowledge matching, Text similarity, Hierarchical network of concepts, Semantic distance

引言

据统计,在产品质量问题中属于设计问题的约占75%^[1],产品设计质量的控制和提高已成为制造商竞争焦点。在产品设计过程中,产品设计人员获得的相关设计质量信息越多、越准确,越能设计出符合需求的高质量产品。

通过设计质量知识的匹配,可以根据设计人员的设计任务将相关的设计质量信息检索出来,这对产品设计质量的提升具有重要的意义。许多学者在此方面做了大量研究,Caldwell^[2]等提出了一种将正确的设计质量知识传递给正确的设计人员的方法;Vallet等^[3]基于本体驱动领域表达,考虑用户偏好进行文档的动态检索;王生发等^[4]提出了基于知

收稿日期: 2011-01-19 修回日期: 2011-03-14

* 国家自然科学基金资助项目(50835008)

作者简介: 茅健,副教授,博士后,主要从事先进制造技术研究,E-mail: jmiao@sues.edu.cn

通讯作者: 曹衍龙,副研究员,博士后,主要从事先进制造技术、质量工程等研究,E-mail: sdcaoyl@zju.edu.cn

识管理的、以 workflow 驱动的、产品设计的知识主动推送体系结构;孙伟等^[5]提出了基于混合语义模型的产品知识文档检索方法。

文本相似度计算是实现设计质量信息匹配的有效方法,Agirre 等^[6]使用 WordNet 进行相似度计算;金博等^[7]使用知网知识结构进行相似度计算,并将其应用范围扩大到段落。冯少荣等^[8]从语义上分析文本之间的相似值,使结果接近人的主观衡量。本文提出一种产品设计过程中设计质量知识的表示及匹配方法,实现设计质量信息的智能化匹配,并用实例验证此方法的有效性。

1 产品设计质量过程表示

有效组织产品设计过程中的质量数据,是实现设计质量知识匹配的关键之一。为了实现设计质量知识的匹配与传递,以任务元、对象元、信息元来构建设计任务、设计人员和设计质量知识的形式化描述体系。

(1)设计人员

设计人员是以产品设计为核心展开活动所涉及的人员。按照设计任务组织者分配的任务,在相关知识的帮助下完成相应的设计任务。设计人员信息可用对象元 $D = \{D_n, D_r, D_t, D_f\}$ 来表示。其中, D_n 表示设计人员的名称(或编号); D_r 表示设计人员的角色; D_t 表示分配给设计人员的设计任务; D_f 表示设计人员具有的功能。

(2)设计任务

产品设计任务是按照设计任务书的要求,在产品设计各个阶段完成规定的设计内容。在以 workflow 驱动的产品设计过程中,设计任务依附于 workflow,设计组织者依照产品设计需求,并综合考察设计人员的人员状况和知识后,将任务分配给设计人员, workflow 指导设计人员按照定义好的任务、角色和规则来执行和控制业务过程。

设计任务可表示为: $T = \{T_n, T_r, T_o, T_p, T_d\}$ 。其中, T_n 表示该设计任务的名称; T_r 表示该设计任务对应的设计任务层次树中的位置; T_o 表示设计任务对应的产品开发流程的节点; T_p 表示对设计任务的属性描述; T_d 表示该设计任务对应的设计人员。

(3)设计质量知识

设计质量知识是产品设计所需要的各种知识的集合,包括规范文件、设计原理、决策信息、失败案例、功能与结构的映射关系、结构的分解关系、工艺性文件等。为了使设计人员在设计过程中实现推动式而不是拉动式的设计,需要为这些设计质量知识定义统一的表示结构,从而使相似度匹配计算成为

可能。

设计质量知识可表示为 $I = \{I_n, I_e, I_c, I_a, I_o, I_p\}$ 。其中, I_n 表示设计质量知识对应的唯一编号; I_e 表示设计质量知识所属对象或者产品的名称; I_c 表示设计质量知识的类别,如设计原理、设计规则、设计方案、设计图纸等; I_a 表示该设计质量知识的存储路径; I_o 表示对设计质量知识的总体描述,摘要性地概括其内容; I_p 表示设计质量知识的属性信息,包括创建人、创建阶段、生成日期等。

2 基于语义距离的文本相似度计算

基于语义理解的文本相似度是对词语、句子与段落之间相似度的计算分析,与基于统计学的相似度计算方法相比,不需要大规模语义库的支持,也不需要长时间的训练,因此在信息检索、数据挖掘等领域有着广泛的应用。

2.1 HNC 理论与语义距离

HNC(概念层次网络)理论以语义表达为基础,对语义的表达进行概念化、层次化、网络化,取得了自然语言表述和处理模式的新突破^[9]。基于 HNC 理论建立的概念知识库和语言知识库,为语义距离计算提供了必要的数据库支持。所谓语义距离是指不同概念之间关联性强弱的定量描述。语义距离计算通过语义块之间的联想,以最小语义距离作为“多义选一”的基本判据,实现了自然语言在语义层的确切表达和比较。在 HNC 体系下,两语义块之间的距离等于它们之间所有可能路径中的最短路径。

2.2 文本相似度的精确量化和比较

语境框架是 HNC 理论体系下提出的文本内容的形式化结构,覃张华等^[10]基于 HNC 理论实现了

文本语境框架的提取。设文本为 $T_o = \sum_{i=1}^m t_{oi} (m \geq 1)$,

$T_p = \sum_{j=1}^n t_{pj} (n \geq 1)$,其语义结构分别为

$$\begin{cases} \text{SemFrame}(T_o) = \sum_{i=1}^m J(t_{oi}) \\ \text{SemFrame}(T_p) = \sum_{j=1}^n J(t_{pj}) \end{cases} \quad (1)$$

其中 $\text{SemFrame}()$ 表示文本的语义结构, $J(t_{oi}) (i = 1, 2, \dots, m)$ 、 $J(t_{pj}) (j = 1, 2, \dots, n)$ 分别为 T_o 、 T_p 语义结构的语境框架。

文本 T_o 和 T_p 的相似度可以用其语境框架的相似度计算得到,其计算公式为

$$\text{Sim}(T_o, T_p) = \prod_{i=1}^m \sum_{j=1}^n \text{Sim}(J(t_{oi}), J(t_{pj})) \quad (2)$$

其中 $\text{Sim}(J(t_{oi}), J(t_{pj}))$ 表示 $J(t_{oi})$ 和 $J(t_{pj})$ 之间的

相似度,只有当其达到设定的阈值时,被匹配文本和匹配文本认定为匹配成功,否则匹配失败。

晋耀红^[11]在语境框架的基础上,设计实现了领域、情景、背景相似度计算算法,但是在语义层的多义取一问题上没有实现根本解决。本文利用 HNC 知识库,基于语境框架文本相似度和语义距离的分析计算,实现文本相似度的精确量化和比较。基于语境框架的文本相似度计算把文本内容抽象成领域、情景、背景 3 个侧面,其相似度计算公式为

$$\begin{aligned} \text{Sim}(J(t_{oi}), J(t_{pj})) &= \text{SimDom}(J(t_{oi}), \\ &J(t_{pj})) \text{SimSit}(J(t_{oi}), \\ &J(t_{pj})) \text{SimBac}(J(t_{oi}), J(t_{pj})) \end{aligned} \quad (3)$$

其中 $\text{SimDom}()$ 、 $\text{SimSit}()$ 、 $\text{SimBac}()$ 分别表示领域相似度、情景相似度、背景相似度。

不失一般性,此处以情景相似度为例展开分析。在语境框架的情景提取中,总可以把情景提取为情景行为的主体、情景行为的受体、情景行为 3 部分,分别对应一个语义块,语义块是语句的下一级语义单元构成单位,它是句类的函数(HNC 理论将句类分为 7 种^[9])。设 $J(t_{oi})$ 、 $J(t_{pj})$ 的情景信息语义块分别为 $\{e_{t_{oi}1}, e_{t_{oi}2}, e_{t_{oi}3}\}$ 、 $\{e_{t_{pj}1}, e_{t_{pj}2}, e_{t_{pj}3}\}$, 则情景相似度为

$$\text{SimSit}(J(t_{oi}), J(t_{pj})) = \sum_{k=1}^3 \text{Similar}(e_{t_{oi}k}, e_{t_{pj}k}) \quad (4)$$

其中 $\text{Similar}(e_{t_{oi}k}, e_{t_{pj}k})$ 是语义块 $e_{t_{oi}k}$ 和 $e_{t_{pj}k}$ 之间的语义相似度。

在情景层次网络体系下, $e_{t_{oi}k}$ 、 $e_{t_{pj}k}$ 等语义块分别对应其中的一个节点,若此体系为 n 层,则两个语义块的距离介于 $0 \sim 2n$, 当 $e_{t_{oi}k} = e_{t_{pj}k}$ 时, $\text{Similar}(e_{t_{oi}k}, e_{t_{pj}k}) = 1$, 当 $|e_{t_{oi}k} - e_{t_{pj}k}| = 2n$ 时, $\text{Similar}(e_{t_{oi}k}, e_{t_{pj}k}) = 0$, 随着语义块 $e_{t_{oi}k}$ 和 $e_{t_{pj}k}$ 之间差距的增大,语义相似度的值越来越小,因此构造语义相似度的公式

$$\text{Similar}(e_{t_{oi}k}, e_{t_{pj}k}) = 1 - \frac{E}{2n} \quad (\text{Similar} \in [0, 1]) \quad (5)$$

其中 E 为 $e_{t_{oi}k}$ 和 $e_{t_{pj}k}$ 语义块之间的距离。

在情景层次网络体系中影响语义距离的因素主要有语义块重合度和语义深度。在相同的语义块距离条件下,实际相似度值随语义块重合度的增大而增大。语义块重合度是指包含相同上位节点的个数,引入重合度修正系数

$$\alpha = \frac{\text{Coin}(e_{t_{oi}k}, e_{t_{pj}k})}{n} \quad (6)$$

式中 $\text{Coin}(e_{t_{oi}k}, e_{t_{pj}k})$ ——上位节点重合数
 n ——层数

语义块深度是指语义块在体系下的层次,距离相等的两个语义块的相似度随着它们所在层次深度的增加而增加,所以在计算概念之间语义距离的时候,必须考虑语义块深度的影响,其修正系数为

$$\beta = \frac{D(e_{t_{oi}k}) + D(e_{t_{pj}k})}{2n} \quad (7)$$

式中 $D(e_{t_{oi}k})$ 、 $D(e_{t_{pj}k})$ ——语义块的深度
 则修正后的语义相似度计算公式为

$$\text{Similar}(e_{t_{oi}k}, e_{t_{pj}k}) = \alpha\beta \left(1 - \frac{E}{2n}\right) \quad (8)$$

将式(8)代入式(4)即得情景相似度的计算值,同样求得领域相似度和背景相似度后代入式(3)可得各个语义框架之间的相似度值,最终实现式(2)的求解。

3 设计质量知识匹配的实现

任务元、对象元、信息元构成了设计质量信息模型的形式化描述体系。通过文本的相似度计算,可以实现设计质量信息的实时有效匹配。将上节的文本相似度计算方法应用到设计领域,设计质量信息 I 包括历史设计质量及当前设计质量信息,以此作为待匹配文本,设计任务 T 和设计人员 D 为匹配文本,匹配的结果即为当前设计人员在当前任务下的相关设计质量知识信息。

基于语义距离的匹配过程首先需要对原有的质量信息数据进行处理,处理过程如图 1 所示。其中文本系统存储了所有与设计质量相关的文本信息,通过数据库中的设计质量信息表,提取生成质量信息文本的元信息,并通过合理性分析、归并、集成等操作生成信息元表。语境框架生成单元提取信息元表中的信息生成其领域、情景、背景信息,分别存储在领域表、情景表、背景表中。

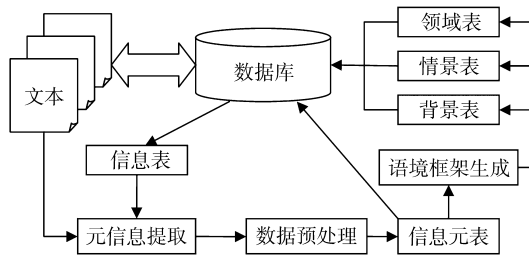


图 1 数据预处理过程

Fig. 1 Data preprocessing process

因此设计任务与设计质量信息的匹配可以简化为寻求 T 和 I 文本的最大相似度。

已知: $T = \{T_n, T_r, T_o, T_p, T_d\}$ 为匹配文本; $I = \{I_n, I_e, I_c, I_a, I_o, I_p\}$ 是存储在数据库中的信息元,以 I 为待匹配文本,匹配过程为已知的 T 与数据库

中的各条信息元一一计算。

把 T, I 代入式(2)中,得

$$\text{Sim}(T, I) = \prod_{j=n,r,o,p,d} \left(\sum_{i=n,e,c,a,o,p} \text{Sim}(J(T_j), J(I_i)) \right) \quad (9)$$

在匹配过程中,提取 T 和 I 中的语境框架信息,由式(3)~(8)计算得到各语境框架之间的相似度,代入式(9),若 $\text{Sim}(T, I) \neq 0$,则此条 I 即为满足设计任务要求的信息元,并存储在表 I' 中,依此计算可获得符合设计任务要求的 I' 。

同理,已知: $D = \{D_n, D_r, D_i, D_f\}$,把 D 和 I' 每条数据分别代入式(2)中计算,把满足要求的信息元存入另一数据表 I'' 中,获得最终供设计人员参考的信息元数据表 I'' 。

信息元与信息实体存在一一对应的关系,通过 $I'' \rightarrow I_{\text{path}}$ 中存储的路径信息,设计人员就可以获得设

计质量历史信息,以供设计参考。由于在匹配过程中,与设计质量信息匹配的过程不是直接与质量信息内容进行匹配,而是与这些信息元及其语境框架进行匹配,因此避免了由于直接与数据库中相应的数据进行匹配而产生的数据交换量大、效率低的缺点。

4 实例

以某注塑机的方案设计为例,设计过程在产品数据管理(PDM)平台下进行。如表 1 所示,列出了方案设计阶段的某一设计人员对应的任务元 and 对象元。从 PDM 系统中选取 200 份设计阶段的质量文档对应的信息元,采用基于语义距离的文本相似度方法对文档检索匹配。采用查全率、查准率作为测试指标,并与基于语义理解的设计质量文档检索方法对比。

表 1 匹配过程的任务元、对象元和信息元列表
Tab. 1 List of task meta, object meta and information meta in matching process

对象元	D_n 设计人员编号为 Hz09102
	D_r 具有总工程师审核角色,主要做注塑机的总体结构设计
	D_i 2 400 t 全液压式注塑机的方案设计
	D_f 负责对各个设计方案进行评审,选出一个最优方案或者在总结这些方案的基础上提出一个新方案
任务元	T_n 方案设计
	T_r 方案设计阶段的方案设计
	T_o 按照项目立项的要求开始方案设计,为项目创建作准备
	T_p 选取注塑机的液压系统、控制系统、锁模结构、注射装置和驱动系统
	T_d Hz09102 负责方案评审, Hz09101 负责方案设计…
信息元	200 份设计阶段的质量文档对应的信息元表

设 R 为满足匹配条件的文档数量, A 为自动筛选返回的文档数量,由于 R 不易求得,因此实际计算过程中,由设计人员本人根据设计主题,筛选出符合条件的质量文档,记为 R' ,并以此作为 R 的近似来求解,则查全率、查准率计算公式为

$$\begin{cases} F = \frac{R' \cap A}{R'} \times 100\% \\ P = \frac{R' \cap A}{A} \times 100\% \end{cases} \quad (10)$$

由方案设计人员从 200 份质量文档中筛选出满

足要求的文档,根据返回文档数计算查全率和查准率,如果相似度大于 0.5,就算满足要求的一份文档。以此类推,得出最后结果如表 2 所示,通过比较可知,基于语义距离的文本相似度匹配检索方法在查全率和查准率方面都有明显的提高。这是因为基于 HNC 的语境框架模型能够比传统的语义理解更加准确地反映设计人员的真实需求,提高了信息的利用率,从而节省了设计者花费在信息检索上的时间。

表 2 基于语义距离和语义理解的检索比较
Tab. 2 Comparison between semantic distance and semantic understand

方法	文档总数	满足条件的文档数 R'	返回文档数 A	误查数	查全率 $F/\%$	查准率 $P/\%$
基于语义理解	200	100	70	8	62	88.6
基于语义距离	200	100	92	5	87	94.6

5 结束语

基于文本相似度的智能化匹配方法可有效避免设计质量信息的遗漏,减少设计人员花费在设计质量信息检索上的时间,使设计人员集中在设计任务上。本文依照设计质量信息模型,以任务元、对象元、信息元来构建设计任务、设计人员和设计质量信

息三者以及三者之间的形式化描述体系,并建立其数据结构。在基于语境框架文本相似度计算的基础上,结合语义距离计算,实现了设计质量信息的准确匹配,模型、数据体系结构是基础,文本相似度计算是手段,设计质量信息的准确匹配是目的。设计人员通过获得准确的信息,保证了决策的合理性,防止了设计质量问题的重复发生。

参 考 文 献

- 1 夷嫵霖. 设计质量控制系统的研究与开发[D]. 重庆:重庆大学,2005.
- 2 Caldwell N H M, Clarkson P J. Web-based knowledge management for distributed design [J]. IEEE Intelligent System, 2000, 15(3):40~47.
- 3 Vallet D, Castells P, Fernandez M, et al. Personalized content retrieval in context using ontological knowledge [J]. IEEE Transactions on Circuits and Systems for Video Technology, 2007, 17(3):336~346.
- 4 王生发,顾新建,郭剑锋,等. 面向产品设计的知识主动推送研究[J]. 计算机集成制造系统, 2007,13(2): 234~239. Wang Shengfa, Gu Xinjian, Guo Jianfeng, et al. Knowledge active push for product design [J]. Computer Integrated Manufacturing Systems, 2007, 13(2):234~239. (in Chinese)
- 5 孙伟,马沁怡,郭莉,等. 混合语义模型的产品知识文档检索[J]. 重庆大学学报:自然科学版,2008, 31(10): 1 198~1 208. Sun Wei, Ma Qinyi, Guo Li, et al. Product knowledge documents retrieval based on hybrid semantic model [J]. Journal of Chongqing University: Natural Science, 2008, 31(10):1 198~1 208. (in Chinese)
- 6 Eneko Agirre, German Rigau. Word sense disambiguation using conceptual distance [C] // Proceedings of the 16th Conference on Computational Linguistics, 1995: 16~22.
- 7 金博,史彦军,滕弘飞. 基于语义理解的文本相似度算法[J]. 大连理工大学学报,2005,45(2):291~297. Jin Bo, Shi Yanjun, Teng Hongfei. Experimental research on premixed gases explosion in overpressure [J]. Journal of Dalian University of Technology, 2005, 45(2): 291~297. (in Chinese)
- 8 冯少荣,肖文俊. 基于语义距离的高效文本聚类算法[J]. 华南理工大学学报:自然科学版, 2008, 36(5):30~37. Feng Shaorong, Xiao Wenjun. High-efficiency text clustering algorithm based on semantic distance [J]. Journal of South China University of Technology: Natural Science Edition, 2008, 36(5):30~37. (in Chinese)
- 9 黄曾阳. 概念层次网络理论[M]. 北京:清华大学出版社,1998.
- 10 覃张华,王景中. 基于 HNC 理论的短文本语境框架提取实现[J]. 北京工商大学学报:自然科学版,2007,25(5): 49~52. Qin Zhanghua, Wang Jingzhong. Short text context framework extracting implementation based on HNC theory [J]. Journal of Beijing Technology and Business University: Natural Science Edition, 2007, 25(5): 49~52. (in Chinese)
- 11 晋耀红. 基于语境框架的文本相似度计算[J]. 计算机工程与应用,2004,40(16):30~39. Jin Yaohong. Text similarity computing baseol on context framework model[J]. Computer Engineering and Applications, 2004,40(16):30~39. (in Chinese)
- 12 包志炎,肖刚,高飞,等. 基于实例推理的个性化产品评价方法[J]. 农业机械学报,2011,42(9):197~201. Bao Zhiyan, Xiao Gang, Gao Fei, et al. Evaluation method of individualized products based on case reasoning [J]. Transactions of the Chinese Society for Agricultural Machinery, 2011,42(9):197~201. (in Chinese)
- 13 胡伟,胡国清,魏昕,等. 基于图论的产品实例检索结构相似度分析[J]. 农业机械学报,2011,42(8):184~188. Hu Wei, Hu Guoqing, Wei Xin, et al. Structural similarity analysis in case retrieval using graph theory[J]. Transactions of the Chinese Society for Agricultural Machinery, 2011,42(8):184~188. (in Chinese)
- 14 陈亮,张邵平,王伟. 产品多学科协同设计中基于语义推理的动态协调方法[J]. 农业机械学报,2011,42(5):208~213. Chen Liang, Zhang Shaoping, Wang Wei. Dynamic coordination principle and method of multidisciplinary collaborative product design based on semantics reasoning[J]. Transactions of the Chinese Society for Agricultural Machinery, 2011, 42(5):208~213. (in Chinese)
- 15 张太华,顾新建,白福友. 基于产品知识模块本体的产品知识集成[J]. 农业机械学报,2011,42(5):214~221. Zhang Taihua, Gu Xinjian, Bai Fuyou. Integration of product knowledge based on modular ontologies of product knowledge [J]. Transactions of the Chinese Society for Agricultural Machinery, 2011,42(5):214~221. (in Chinese)