

基于网评与专利数据的产品概念设计知识服务研究

张桂海, 余军合*, 战洪飞, 朱琳

(宁波大学 机械工程与力学学院, 浙江 宁波 315211)

摘要: 概念设计对产品的创新性和研发周期具有决定性的影响, 为了提高产品概念设计的效率和水平, 提出一种以网络评论和专利数据为核心的产品概念设计知识服务方法. 在该方法中, 构建了基于网络产品评论数据和专利数据的产品概念设计知识服务模型, 应用文本挖掘技术从产品评论数据中挖掘用户需求信息, 构建专利技术功效矩阵, 挖掘专利的技术和功效信息, 进而向设计人员提供用户需求及相应的专利技术方案等设计知识. 最后, 以多家企业作为实例, 验证该方法的有效性.

关键词: 数据驱动; 概念设计; 知识服务

中图分类号: TP391

文献标志码: A

文章编号: 1001-5132 (2018) 05-0024-07

随着用户需求个性化和多样化不断增强, 企业必须加快产品的创新和优化升级. 概念设计是用户需求融入设计方案和产品创新概念形成的重要阶段, 需要使用大量的设计知识. 向设计人员提供相应的辅助知识服务, 将会提高概念设计的效率和水平.

国内外学者对产品设计的知识服务进行了大量研究. 在信息检索方面, Degemmis 等^[1]提出了一种采用包含用户兴趣的语义信息文档代替关键词来搜索的方法, 实现了基于内容的知识服务. 对于设计知识的建模和表达, Chang 等^[2]从本体角度出发, 对设计数据进行集成, 构建了一个设计知识服务框架, 为设计人员提供知识推送服务; Zhang 等^[3]提出了一种基于本体框架表达设计元意图的方法. 知识重用方面, 郝佳等^[4]研究了产品设计知识的重用模式, 提出了一种包含导航知识、知识组件和知识地图的产品设计知识组织方法; Liu 等^[5]提出了一种面向设计过程的知识重用方法, 将设计活动分为结构设计和参数设计, 对不同的设计活动提供相应的设计知识. 在推送方法上, 王生发等^[6]采

用知识管理的方式, 以工作流程为驱动, 提出了一种产品设计知识主动推送方法. 另外, 有的学者对产品的需求、功能、结构及其关系等进行了分析. 杨艳华等^[7]针对如何准确获得满足需求的结构方案问题, 提出了需求/结构映射的灰关联相似性匹配方法. 杨波等^[8]在分析功能描述方法的基础上, 提出了一种对功能扩展建模的方法, 避免产品的结构形式被过早限制.

本文从产品评论和专利等数据出发, 提出一种对产品概念设计提供知识服务的方法. 通过挖掘用户需求信息和专利的功效与技术信息, 构建产品概念设计辅助知识模型, 实现需求分析-技术方案的辅助知识服务过程, 提高设计效率和水平.

1 基于产品评论与专利数据的设计知识获取

产品概念设计是设计人员根据设计目标在各种设计约束条件下, 利用多种类型的知识对设计问题求解, 得到满足设计要求的方案的活动和过程^[9]. 产品概念设计涉及用户需求状况、产品功能、实现技术(原理和结构)等知识.

收稿日期: 2017-12-12.

宁波大学学报(理工版)网址: <http://journal1g.nbu.edu.cn/>

基金项目: 国家自然科学基金(71671097); 浙江省自然科学基金(LY16G010004); 浙江省公益技术应用研究计划(2016C31047); 浙江省公益技术研究计划(LGG18E050002).

第一作者: 张桂海(1989-), 男, 广西桂平人, 在读硕士研究生, 主要研究方向: 数据挖掘、企业信息化. E-mail: 18892614899@163.com

*通信作者: 余军合(1971-), 男, 湖北天门人, 博士/副教授, 主要研究方向: 产品数据管理、制造系统工程. E-mail: yujunhe@nbu.edu.cn

1.1 基于产品评论的用户需求分析

产品评论包含了用户的满意状况、产品优缺点以及用户期望,挖掘产品评论能够明确产品在优化升级时的改进方向。

产品评论为非结构化的文本数据,需要利用文本挖掘技术和自然语言处理技术进行分析,主要涉及评论数据的获取、预处理、产品特征提取、情感分析和结果可视化等内容。

利用 Python 语言编写网络爬虫程序,在电商平台采集用户的评论数据,构建产品评论语料库。文本数据需要进行预处理才能进一步分析,主要包括数据清理、分句分词、词性标注和停用词删除等。在此基础上,提取评论中的产品特征词。产品特征为用户评论和关注的对象,往往出现在一些特定的语法模式中。采用语法模式^[10]的方法,提取符合特定语法模式的产品特征,包括产品功能、属性、组件、组件属性以及跟产品相关的实体等^[11]。为了了解用户满意度,需要计算用户对产品的情感倾向,基于情感词典和副词词典,利用 SO-PMI 算法^[12]计算用户评论中的情感词与基准词在语料库中的共现概率,计算用户对每一项产品特征的情感褒贬义倾向及强度。程度副词可以表达情感极性的强度,对不同的程度副词赋予不同的权重,挖掘其对情感的增强、减弱作用,以提高情感分析的精度。按照程度的不同,将程度副词划分为4个等级并赋相应权重:极其(1.5)、非常(1.2)、一般(1)、稍微(0.8)。否定副词对情感极性具有反转的作用,应根据否定的次数来确定其权重系数。每条产品评论中某项产品特征的情感得分 S 计算公式为:

$$S = K_1 K_2 K_3, \quad (1)$$

式中: K_1 为观点词极性系数,极性为正时, K_1 取值 1, 极性为负, K_1 取值 -1; K_2 为程度副词的权重; K_3 为否定词的权重系数,修饰观点词的否定词个数为奇数时, K_3 取值 -1, 否定词个数为偶数时, K_3 取值 1。若观点词无程度副词或否定词修饰,则相应的系数取值为 1。

最后将产品特征及其情感倾向进行可视化表达,分析用户需求情况。用户评论分析过程如图 1 所示。语法模式集为: $G = \{g_1, g_2, \dots, g_{14}\} = \{(n+adj), (n+v), (n+adv+adj), (n+adv+v), (v+n), (n+adv+adv+adj), (n+adv+adv+v), (n+n), (adv+v+n), (adj+de+n),$

$(n+n+n), (v+adj), (v+adv+adj), (v+adv+adv+adj)\}$ 。

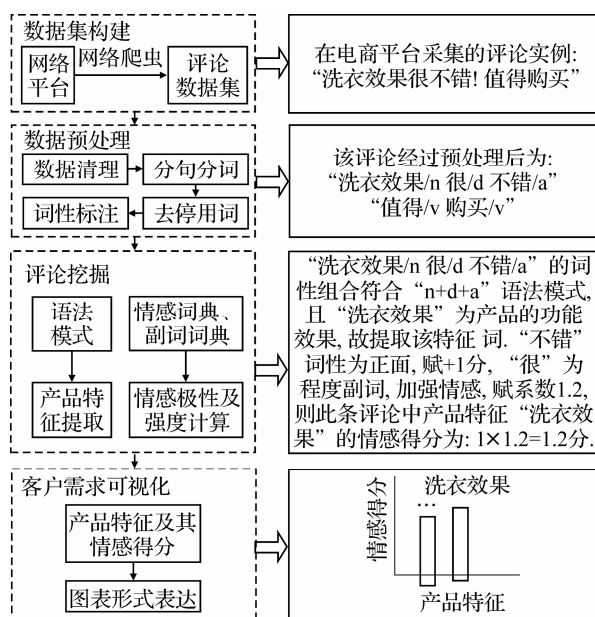


图 1 基于产品评论的用户需求分析过程

为进一步挖掘用户需求信息,可在预处理后的评论语句中提取同一语句中出现的 product 特征词,以挖掘产品属性对应的主体词以及主体词对应的属性或功能等信息。

1.2 基于专利数据的技术功效矩阵构建

专利是新技术、新工艺等创新成果的集中体现,蕴含着大量新颖的原创性求解方案^[13]。专利文献数量巨大,为了能快速检索到相关的专利,可通过专利的主题实现专利文献的快速定位。基于此,本文采用专利技术功效矩阵分析方法。

专利技术功效矩阵能够表达专利文献中的技术主题与功能主题及其内在联系,一方面可以了解有哪些专利技术来实现某一种功能效果;另一方面可以直观地看出某一种专利技术有多少种功能效果。

构建专利技术功效矩阵,先选定专利的技术和功效分类架构。利用文本主题聚类算法,挖掘专利数据的主题信息,并根据相关背景知识,对主题信息进行解读,划分出特定领域的专利技术分类和功效分类,并构建对应的技术-功效列表。然后将每项专利与对应的技术-功效列表进行关联映射,进行分类标引以确定该项专利在对应的技术和功效分类架构中的位置。最后根据专利与技术-功效列表的对应关系,完成技术功效矩阵图的制作。

1.2.1 专利技术和功效分类架构的确定

采用 LDA (Latent Dirichlet Allocation) 主题模型将专利集聚成若干主题类, LDA 主题模型是从大量文本语料中挖掘主题的典型模型. 采用 Dirichlet 分布作为主题分布信息的先验知识, 对文档的生产过程进行刻画^[14]. 文档中主题(T)可由在语义上与该主题相关的一系列词(w_i)及其权重 $p(w_i|T)$ 构成的向量表示, 词语与主题越紧密, 则该词语的条件概率越大, 反之越小^[15]. 主题的向量表示为:

$$T = ((w_1, p(w_1|T)), (w_2, p(w_2|T)), \dots, (w_n, p(w_n|T))). \quad (2)$$

专利的主题可分为技术主题类和功效主题类, 利用 LDA 主题聚类算法得到的每个主题中, 主题词涉及的主题情况有 3 种: (1) 主要为功效主题词; (2) 主要为技术主题词; (3) 包含功效主题词和技术主题词. 对主题词的解读, 需要具备相关领域的知识背景, 从主题词中抽象概括出相应的主题. LDA 主题聚类结果可能会出现某些主题类的遗漏或者不合理之处. 因此, 在 LDA 主题聚类的基础上, 再结合研究对象的特点, 对这些主题类别进行完善, 得到较为全面的主题类.

1.2.2 专利技术词和功效词的提取

技术词和功效词的提取是构建专利技术功效矩阵的关键之一. 专利技术词主要是指专利所涉及的原理方法、结构、工具、材料、组件、技术流程等词汇, 功效词是指表示性能、效果、用途、目标等的词汇^[16]. 专利摘要能够较全面地对专利的原理方法、材料结构、功能用途、性能效果等方面进行阐述. 因此, 可以从专利摘要中提取技术词和功效词^[17]. 通常中文专利的技术词前有较为明显的引导词, 如: “利用…手段” “采用…技术” 等, 通过定义一些规则库, 抽取这些含有技术词的句子, 并抽取其中的名词短语作为技术词候选词; 功效词通常由动宾结构($v+n$ 、 $v+n+n$)或者主谓结构($n+v/a$ 、 $n+n+v/a$)的短语构成, 将专利文本进行预处理后, 将符合动宾结构和主谓结构的短语提取出来, 构成功效词候选词, 并过滤候选词中的干扰项, 进而构建专利的技术词和功效词词典.

本文将专利的摘要分别与技术词和功效词词典进行匹配, 提取每项专利的技术词和功效词. 对于专利新语料中没有提取到的技术词和功效词,

则进行人工添加到相应的词典, 从而不断完善该领域的专利技术词词典与功效词词典. 结合相关的专利文献信息和产品特点, 适当地将每项专利的技术词和功效词划分到技术类和功效类的分类框架里, 并在每项专利中标记其所属的功效技术类号, 统计每个技术类和功效类对应的专利数量, 并完成技术功效矩阵的建立. 专利技术功效矩阵构建流程如图 2 所示.

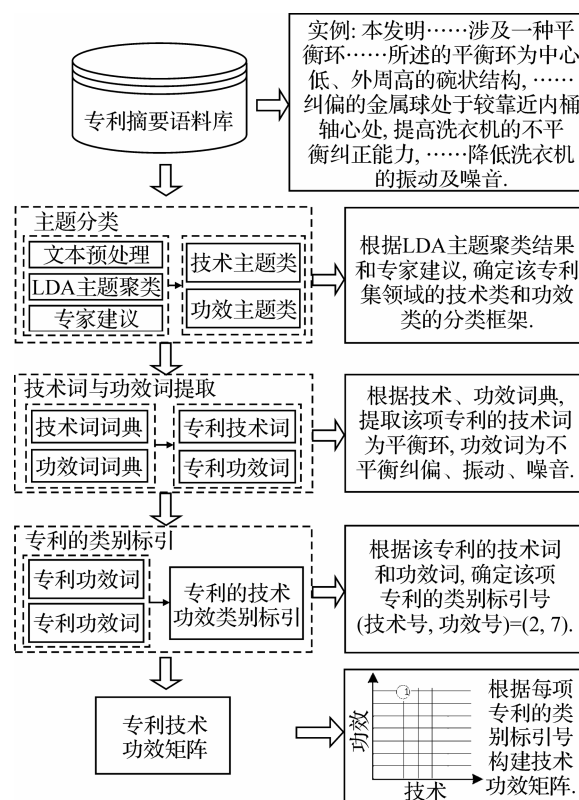


图2 专利技术功效矩阵构建流程

2 基于产品评论与专利的产品概念设计知识服务模式

本文构建的产品概念设计知识服务模式主要包括用户需求模型、需求-功效-技术的映射过程、专利技术功效矩阵等内容, 如图 3 所示. 从产品评论中挖掘得到的用户需求经过分解转换, 得到一系列的子需求, 结合产品原有的树形结构用户需求模型, 实现用户需求模型的修改完善, 比如需求的增加、删除、修改和替换等, 从而将用户需求融入新产品概念设计中. 需求的变动引起产品功能效果相应的变化, 为了能快速准确地获取实现该部分功能效果的技术方案, 可从专利中提取专利的功

效主题和技术主题, 确定每项专利的技术和功效, 并通过需求-功效-技术的映射关系, 得到相应的技术方案解。

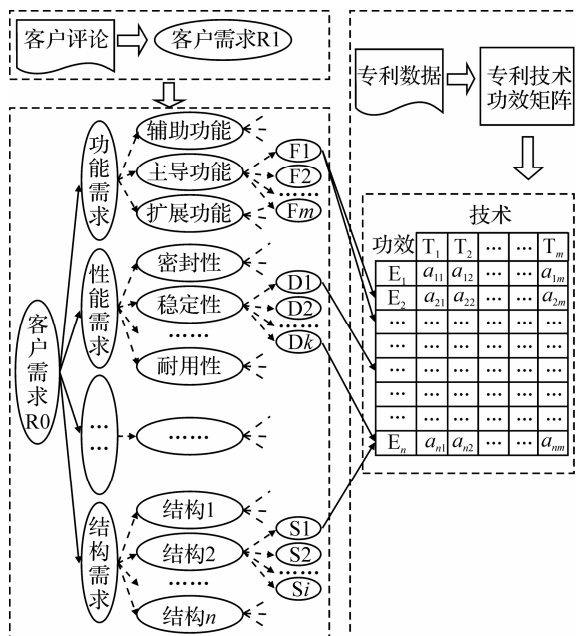


图3 数据驱动的产品概念设计辅助知识服务模型

2.1 用户需求模型树的构建

对用户需求进行分类整理并分解细化, 以促进需求信息的重用和变更。本文将用户需求信息分为7大类: 功能需求、性能需求、经济性需求、安全可靠需求、外观需求、拆卸维修需求和结构需求, 如图3中用户需求R0部分。

用户需求的模糊性、个性化和多样化导致了用户需求的动态性。为此, 在分析产品评论的用户需求信息R1时, 应按照用户需求模型R0的分类方式将R1进行分类和细化, 并将其融入到用户需求模型R0中, 进而指导设计人员确定产品功效方案的制定。

2.2 需求-功效-技术的映射分析

为了将用户需求融入到产品概念设计中, 需要将用户需求映射到产品的功能效果上。本文采用领域专家打分以及使用三角模糊数处理方法来实现用户需求到功效的映射。根据专家的打分和三角模糊数处理后, 构建用户需求与产品功效的关联矩阵(见表1), k_{ij} 表示强度值。用户需求与产品功效的关系强度见表2^[18], 其余的强度值代表关系强度介于表中相邻的等级之间。

表1 用户需求与产品功效的关联映射

用户	功效			
需求	E_1	E_2	$\dots$	E_n
C_1	k_{11}	k_{12}	$\dots$	k_{1n}
C_2	k_{21}	k_{22}	$\dots$	k_{2n}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
C_m	k_{m1}	k_{m2}	$\dots$	k_{mn}

表2 用户需求与产品功效的对应关系

强度值	关系强度
1	关系很强
0.6	关系较强
0.3	关系较弱
0	没有关系

用户需求的变化引起产品功效的变更, 需要相应的技术方案来实现这部分功效。功效-技术表达形式可以直观地表达专利的功效-技术的相互关系, 因此, 可以基于专利的技术功效矩阵来构建功效-技术的映射关系, 如图4所示。图中 a 的值可以根据相应的专利数量来确定。

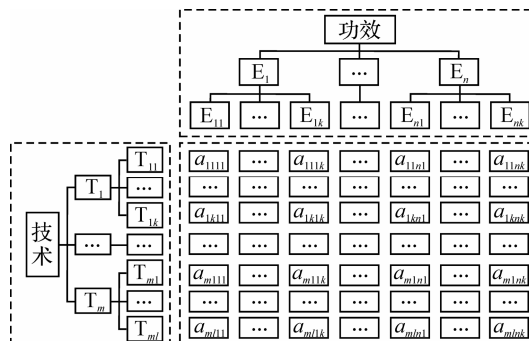


图4 基于专利的产品功效-技术的映射模型

3 实例验证

利用网络爬虫技术在电商平台采集8家企业相同价位多款型号洗衣机的评论数据, 每家8000条评论数据, 构建洗衣机产品评论语料库。在专利检索网站采集洗衣机领域的专利数据(共11730项), 专利类型为发明和实用新型。

3.1 洗衣机产品的用户需求分析

对评论数据的处理分析, 计算每一家企业的洗衣机产品的每项产品特征对应的情感得分。每家企业产品特征的情感得分计算方法如下:

$$S_{ik} = \sum_{j=1}^{m_k} (s_{ip} + s_{im}) / M_k, \quad (3)$$

式中: S_{ik} 为企业 k 的第 i 项产品特征的情感得分; s_{ip} 和 s_{im} 分别为每条含有产品特征 i 的评论中, 产品特征 i 的正面情感得分和负面情感得分; m_k 为第 k 家企业产品评论含有该项产品特征的评论总数; M_k 为第 k 家企业的总评论数. 根据公式(1)来计算 s_{ip} 和 s_{im} 的值: 当公式(1)的 $S>0$ 时, $s_{ip}=S$, 当 $S<0$ 时, $s_{im}=S$.

根据公式(3), 可得到 8 家企业产品评论中用户关注的产品特征及其情感得分情况, 如图 5 所示. 图中的数值为正时, 说明用户对该项产品特征的总体态度为满意; 数值为负时, 说明用户对该项产品特征的总体态度为不满意. 数值的绝对值越大, 说明该项产品特征的情感倾向越明显, 用户关注度越高; 数值的绝对值越小, 说明用户对该项产品

特征的正负面评价数量较为接近, 用户的满意度较低, 且其关注度也越小.

以 C 企业为例, 从图 5 中可以看出, 对于降低噪音这项需求, 用户的关注度很高, C(0.102)企业的总体得分较低, 与其他企业相比处于劣势, 因此在分析噪音需求时, 应加强降低噪音方面的改进.

3.2 洗衣机专利的技术功效构建分析

在专利语料库中随机抽取 5000 条专利摘要, 用于 LDA 主题聚类的语料集. 通过反复试验确定, 主题数 k 值取 18 时聚类效果最好. 在 LDA 主题聚类基础上, 结合相关专业知识和洗衣机产品的功能、结构等内容, 进一步完善主题类的划分, 最后得到结果如图 6 所示的技术主题类和功效主题类.

在专利的技术和功效分类框架确定后, 提取

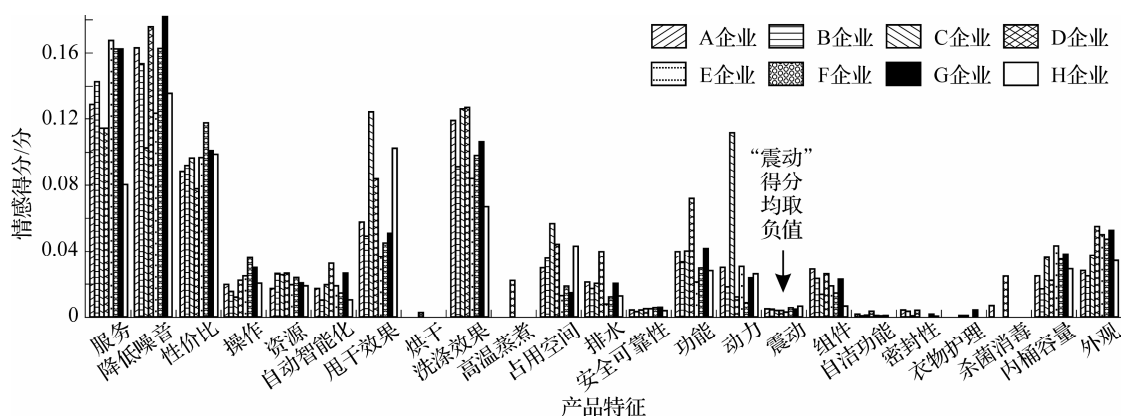


图 5 8 家企业洗衣机的产品特征评价分析结果

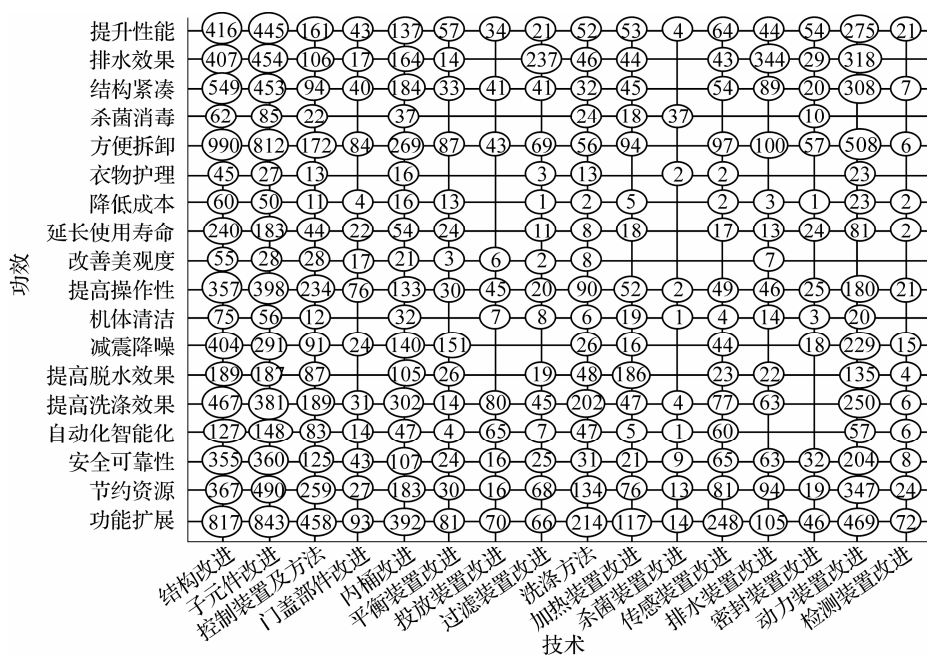


图 6 洗衣机专利的技术功效矩阵图

每项专利的技术词和功效词, 然后根据洗衣机的功能、结构等方面因素和专利涉及的功能效果、技术原理和结构等因素, 将提取到的技术词和功效词进行归类, 确定每项专利在分类框架中的位置. 并统计每一类功效和技术相应的专利数量, 以矩阵图的方式表达. 洗衣机专利的技术功效矩阵图如图6所示.

3.3 面向用户降噪需求的专利技术方案推荐

以降低噪音的需求为例, 降低噪音的需求-功效-技术的映射和技术方案的推荐如图7所示.

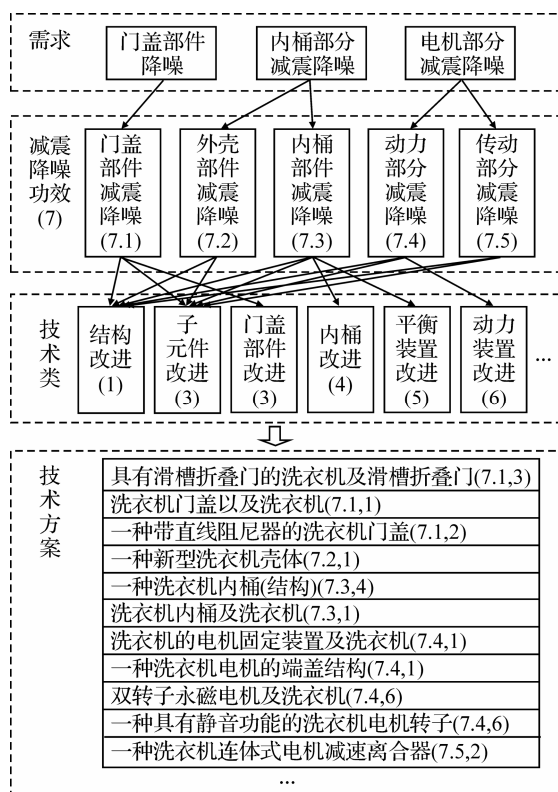


图7 降低噪音需求、专利功效-技术映射与技术方案推荐

根据产品评论可知噪音涉及3方面: 洗衣机门盖产生的噪音、内桶产生的噪音和电机产生的噪音. 专家对“降低噪音”需求与“减震降噪”功效的关系度打分为(0.8, 0.9, 1), 得其关系强度值为1, 说明“降低噪音”需求与“减震降噪”功效具有很强的关系, 所以将“降低噪音”需求映射到“减震降噪”功效. 为了提高获取专利方案的效率, 需要对用户的需求和相应的减震降噪功效的细分项进行相应的技术方案推荐. 比如, 某企业的产品评论中“噪音”的主体词主要分布在“门盖”上, 则向其推荐洗衣机门盖方面的技术方案. 通过对专利功

效类细分项的分类标引, 可实现专利的针对性推荐. 如某一项专利涉及在洗衣机门盖处增设阻尼器(属于子元件改进, 技术分类标号为1), 控制了门盖闭合时的下降速度, 从而降低了门盖碰撞时产生的噪音(属于减震降噪类(分类标号为7.1)), 则该专利的分类标引号为(7.1, 1).

4 总结

本文基于产品评论以及专利数据, 构建了数据驱动的产品设计辅助知识服务模型, 主要内容有以下3个方面:

(1)通过用户需求分析和构建专利技术功效矩阵, 提取了产品概念设计的辅助知识;

(2)构建用户需求模型以及用户需求到产品功能-原理-结构的映射过程, 结合技术功效矩阵, 找到相关的技术方案;

(3)在概念设计阶段, 向设计人员提供用户需求和专利技术方案等辅助知识, 提高产品设计的效率与水平.

参考文献:

- [1] Degemmis M, Lops P, Semeraro G. A content-collaborative recommender that exploits WordNet-based user profiles for neighborhood formation[J]. User Modeling and User-Adapted Interaction, 2007, 17(3):217-255.
- [2] Chang X, Terpenney J. Ontology-based data integration and decision support for product E-design[J]. Robotics and Computer-Integrated Manufacturing, 2009, 25(6):863-870.
- [3] Zhang Y Z, Luo X F. Design meta-intent representation[J]. Computer Integrated Manufacturing Systems, 2010, 16(7):1345-1353.
- [4] 郝佳, 阎艳, 王国新, 等. 面向用户的非几何设计知识组织方法[J]. 计算机集成制造系统, 2014, 20(6):1300-1307.
- [5] Liu X, David R, Yu Z. Ontology based knowledge modeling and reuse approach in product redesign[C]. IEEE International Conference on Information Reuse & Integration, 2010:270-273.
- [6] 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.
- [7] 杨艳华, 姚立纲, 朱祖平, 等. 需求/结构映射的灰关联相似性匹配与求解[J]. 计算机集成制造系统, 2012, 4(18):697-704.

- [8] 杨波, 高长青, 尹晓玲, 等. 面向概念化设计的扩展功能建模方法[J]. 机械工程学报, 2013, 15(49):153-162.
- [9] 吉祥, 顾新建, 代风, 等. 基于本体和粗糙集的产品设计知识推送技术[J]. 计算机集成制造系统, 2013, 19(1): 7-20.
- [10] 单晓红, 杨柳. 网络产品评论挖掘研究[J]. 计算机系统应用, 2014, 23(2):1-6.
- [11] 徐凯. 基于产品特征的用户评论情感倾向分析研究[D]. 合肥: 合肥工业大学, 2015.
- [12] Turney P D, Littman M L. Measuring praise and criticism: Inference of semantic orientation from association[J]. ACM Transactions on Information System, 2003, 21(4): 315-346.
- [13] 陈颖, 张晓林. 专利技术功效矩阵构建词汇模型研究[J]. 情报科学, 2012, 30(11):1704-1709.
- [14] Griffiths T L, Steyvers M. Finding scientific topics[C]. Proceedings of the National Academy of Science, 2004: 5228-5235.
- [15] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation [J]. Journal of Machine Learning Research, 2003, 3:993-1022.
- [16] 胡菊香, 吕学强, 刘秀磊, 等. 专利技术功效短语获取研究[J]. 科学技术与工程, 2016, 16(14):228-235.
- [17] 章伟华. 基于需求建模的产品配置实施关键技术及其在电梯产品中的应用研究[D]. 杭州: 浙江大学, 2010.
- [18] 单鸿波, 葛滨, 于海燕, 等. 面向概念设计的功能需求获取机制研究[J]. 中国机械工程, 2013, 24(12):1605-1610.

Knowledge service for product conceptual design based on online review and patent

ZHANG Gui-hai, YU Jun-he^{*}, ZHAN Hong-fei, ZHU Lin

(Faculty of Mechanical Engineering & Mechanics, Ningbo University, Ningbo 315211, China)

Abstract: Conceptual design has a decisive impact on product innovation and R & D cycle. In order to improve the efficiency and quality of product conceptual design, a knowledge service method of conceptual design based on online review and patent is proposed. In the method, a design knowledge service model for product conceptual design is built on the basis of product review and patent. Customer demand information is mined from product review data with text mining technology. Patent technology-efficiency matrix is built to dig patent technology and function information. Then designers are provided with customer needs and the corresponding design knowledge such as patent and technology programs. Finally, a number of enterprises are taken as an example to verify the effectiveness of the method.

Key words: data-driven; conceptual design; knowledge service

(责任编辑 韩 超)