

基于情境的知识推送技术研究

王士凯^{1,2}, 王力^{1,2}, 江萍^{1,2}, 李怀英^{1,2}, 尹道明^{1,2}

(1. 合肥工业大学管理学院, 安徽合肥 230009;
2. 教育部过程优化与智能决策重点实验室, 安徽合肥 230009)

摘要: 基于情境的知识推送技术是解决当前集团企业在知识管理中的知识有效利用的最直接手段, 情境与知识的有效结合使知识推送过程更加智能化, 提高了知识服务推送的准确性。本研究采用本体描述语言 (Ontology Web Language, OWL), 完成了领域知识与情境知识本体模型的建立、情境-知识关系模型建立以及情境相似度算法设计。构建出仿真过程知识领域本体模型, 运用 protégé 建模工具进行本体实现。最后在跨学科制造领域知识管理工具中进行知识推送的初步应用及验证。

关键词: 云制造; 知识情境; 知识建模; 本体; 知识推送

中图分类号: TP31

文献标识码: A

文章编号: 1673-629X(2013)02-0131-04

doi:10.3969/j.issn.1673-629X.2013.02.033

Knowledge Pushing Technology Research Based on Knowledge Context

WANG Shi-kai^{1,2}, WANG Li^{1,2}, JIANG Ping^{1,2}, LI Huai-ying^{1,2}, YIN Dao-ming^{1,2}

(1. School of Management, Hefei University of Technology, Hefei 230009, China;

2. Key Laboratory of Process Optimization and Intelligent Decision-making
of Ministry of Education, Hefei 230009, China)

Abstract: The knowledge push technology based on context is the most direct means to solve the effective utilization of the current knowledge of the group companies in knowledge management, effective combination of context and knowledge makes the process of knowledge pushing more intelligent, and improve the accuracy of the knowledge-based services pushing. It uses the OWL (Ontology Web Language) to complete the establishment of domain and context knowledge model, the establishment of knowledge-context relation model, and the design of similar matching algorithm. Build the domain knowledge ontology model of missile simulation process which can be realized by using protégé modeling tools. Finally, the initial application and validation of knowledge push will be finished under the knowledge management tools of interdisciplinary manufacturing field.

Key words: cloud manufacturing; knowledge context; knowledge modeling; ontology; knowledge push

0 引言

随着企业信息化的普及, 知识管理已经成为企业管理中的一个重要技术手段。信息时代的到来, 知识资源呈现出爆炸式增长, 企业的知识管理方式亟需更新和改进。如何有效地对领域知识进行描述, 并针对知识用户个性化特征, 实现领域知识的智能化推送, 以期解决知识拉取所造成的“知识过剩”问题, 是当前知识管理研究的热点。合理的知识推送过程, 能极大地

提高企业知识有效利用, 发挥出知识管理在现代化企业运作中的强大作用。

知识推送是由信息推送的概念演变而来, 最早由美国《信息周刊》的主编 RICADELA 在介绍微软公司的知识推送时提出, 并将其定义为一种网络化的知识服务的方式^[1]。通过近几年的发展, 知识推送课题的研究引起了国内外学者的广泛关注, 文献[2]以本体论为依托, 研究了知识管理系统的知识推送技术; 文献[3]研究了一种“以顾客为中心”的知识推送服务系统的构建方法; 在知识本体表示方面, 文献[4]研究了基于描述逻辑的学科知识本体建模; 在知识的相似度匹配方面, 文献[5]对模糊知识匹配方法 IDM 进行了改进; 企业进行知识管理的目的是用最少的学习时间利用现有的知识以创造最大的价值, 而知识推送的主动性特征使得它成为知识管理向实用化发展的一种非

收稿日期: 2012-06-06; 修回日期: 2012-09-10

基金项目: 国家“863”云制造主题项目(2011AA040501); 国家自然科学基金资助项目(70871033); 安徽省教育自然科学基金重点项目(KJ2011A006)

作者简介: 王士凯(1988-), 男, 山东临沂人, 硕士生, 研究方向为人工智能、知识管理。

常关键、重要的技术。知识推送是一个值得关注的、崭新的课题。文中侧重研究基于情境的知识推送问题,通过分析知识的情境信息建立基于本体的情境化知识表示,收集、分析用户信息建立用户情境本体。在此基础上,提出个性化知识服务推送模型,并给出了跨学科制造领域知识管理工具中知识推送模块的实现,使得集团企业能有效实现知识用户与知识服务之间的传递、共享和重用,充分满足了知识用户对知识服务的个性化需求。

1 相关工作

本体(Ontology)最早是一个哲学的范畴,研究物质世界的存在的本质,后来随着人工智能的发展,被人工智能界予以新的定义。目前在领域知识描述方面本体论被广泛研究和应用,它是对领域知识的概念化、形式化、明确化的描述。基于情境的知识表示采用本体表示较其它表示方法有着许多优势,如可重用性、共享、推理功能、标准化等等。本体技术借鉴了语义网技术,语义 Web 标准使得 Web 上的信息具有计算机可以理解的含义,满足智能软件代理(Agent)对异构和分布信息的有效访问和检索。通过 OWL 语言,可以将领域知识和情境知识的相关概念和属性抽象为本体的类和关系,而将具体的知识项和情境实体表现为本体实例,OWL 的很多机制可以明晰地表达语义。

知识管理工作采用本体技术的第一步就是进行知识表示,但不同于术语本体(如 WordNet),本研究采用了复杂公理化的描述逻辑(Description Logic, DL)。DL 代表了一类基于逻辑的知识表示语言,它是 OWL 子语言中的一个。DL 主要由三部分组成,分别是 TBox(Terminological Box)、ABox(Assertional Box)和 RBox(Role Box)。AL(Attribute Language)描述语言是所有描述逻辑的基础。对 AL 描述语言进行扩展,就可以得到描述能力更强的描述逻辑。其中 Schmidt-Schau 和 Smolka 提出的 ALC(Attribute Language with Complements)能够满足本研究中的建模需求。OWL 建立在资源描述框架(Resource Description Framework, RDF)基础上,所有的 RDF 子句都可以表示成三元组的形式,因此可以把知识本体的类、属性和实例都表示为三元组的形式。

2 本体构建

2.1 基于情境的本体建模

2.1.1 本体模型

知识只有在特定的环境下才有应用价值,为了将情境集成到知识中,将知识分为两部分,领域知识和情境知识。领域知识包括知识内容和知识载体两个部

分。知识内容是指能对应用实践有指导意义的信息的描述,包括了知识载体的索引信息,知识载体指导知识用户方便地找到知识具体所在。情境知识是知识情境进行描述的知识,知识情境指知识产生和应用的具体背景和环境。领域知识和情境知识分别对知识的内容和环境进行了描述,是一个统一体,缺一不可。文献[6,7]系统地研究了集成情境的知识模型构建技术,提出了 KMIC 方法。文献[8]研究了基于情境的知识本体表示方法,并简单做了一个帮助医生根据病人病历信息进行诊断的系统。

基于情境的知识推送必须包括用户信息,因此需要建立用户本体模型。用户模型的构成主要由两部分,一部分包含了用户的基本信息,另一部分包含了用户的使用记录。相对应的用户情境本体的获取也有两种方式,前者可以通过用户注册信息以及用户知识服务订阅内容中获取,是一种显性的获取方式;后者通过收集和分析用户的网络行为得出用户的知识需求偏好,是一种隐性的获取方式。

在本研究中,知识的本体表示记为 $O = \langle K_R, K_T, K_A \rangle$ 。其中 K_T (TBox)是对领域知识术语或应用相关背景知识以及用户情境的描述,如概念和属性及其相互关系等,定义了元知识或者是称为定义了术语,是概念公理的集合。对于两个概念 C 和 D ,有如下的定义:概念的所属关系($C \subseteq D$),概念的交、并、否定($C \cup D$ 、 $C \cap D$ 、 $\neg C$),概念间通过属性相连,包括全局限定($\forall R.D$)和存在限定($\exists R.D$),如 $C \subseteq \exists R.D$,表示概念 C 通过属性 R 与个体 D 相连,此外还有顶层概念和底层概念(T 、 \perp)。 K_A (ABox)为知识的断言集合,包括了概念断言和关系断言,概念断言即个体所属类别的断言,如 $C(a)$ 表示 a 是概念 C 的个体,关系断言可以理解为个体之间关系的断言,如 $R(a,b)$ 表示二元组 (a,b) 是属性 R 的一个个体。

根据以上对基本概念的了解,为了方便表述,将领域知识本体、用户情境本体、情境知识本体分别记为 O^D 、 O^U 、 O^C 。

2.1.2 情境-知识关系模型

企业知识分为领域知识和情境知识,将两者有效地结合起来可以准确地描述企业实际应用过程中的知识,通过下面的情境-知识关系模型进行知识与情境的关联。

定义 1: 设 $C_i \in O^C$, $D_j \in O^D$, C_i 和 D_j 分别为 K_T^C 和 K_T^D 中的概念,知识 C_i 发生时的情境为 D_j ,则情境 D_j 和知识 C_i 的关系将被表示为 $P_{i,j} = \{(C_i, D_j)\}$ 。

情境-知识关系模型用本体记为 $O^R = \langle K_T^R, K_A^R \rangle$ 。

(1) 由定义显然 $K_A^R = \Phi$;

(2) 记 R_1 是情境 C_i 的属性, R_2 是领域 D_j 的属性, 则情境-知识的关系在本体中可记为, $P_{i,j} = \exists R_1. C_i \cap \exists R_2. D_j$ 。

2.2 情境相似度算法

情境相似度计算算法主要实现用户情境本体与知识情境本体的相似度计算,它是知识推送服务引擎得以实现的最重要一步,算法的优劣关系到知识推送服务质量的高低。目前主要存在的相似度计算策略有:基于特征匹配的模型、基于语义距离的模型和基于信息内容的模型以及混合模型^[9-11]。由于不同本体之间的层次结构和概率模型是不同的,因此前三种相似度计算策略相对适用于领域内本体概念相似度计算。计算不同本体间的概念相似度大多采用混合模型,它利用了集合理论、本体层次结构和属性结构等信息内容,适合于计算实体类型之间的语义相似度。

本研究提出的用户-知识情境概念相似度算法是一种综合方法,为了方便表示,记为 UCSim(User-Context Similarity)。算法第一步针对特定知识用户的情境本体提取出知识情境作为参加相似度计算的候选概念集,可以有效地减少算法计算量,提高算法运行效率。UCSim 算法通过计算概念名称、属性、实例和关系的相似度,并最终加权得出最终情境相似度。

2.2.1 概念名称相似度

由于情境本体涵盖了领域内所有应用环境概念与术语,而对选定知识使用者的用户情境仅体现用户自身的环境,因此有必要在情境相似度匹配之前先进行情境的候选。概念名称的相似度的计算是字符串的相似度的计算,对于用户本体 O^U 中的某一用户情境的概念 U 与知识情境本体 O^C 中的概念 C 的相似度计算可表示为:

$$S_{\text{name}}(U, C) = \frac{1}{n} \sum_{i=1}^n \text{MaxSim}(u_i, c_j)$$

其中 $u_i (1 \leq i \leq n)$ 是概念 u 的字符串语义分词结果, $c_j (1 \leq j \leq m)$ 是概念 c 的字符串语义分词结果。

2.2.2 属性相似度

属性相似度计算的理论依据是,两个本体概念的属性相似度越高,这两个本体概念的相似度就越高。Tversky 曾提出了一种基于属性的相似度计算方法^[12],计算公式如下:

$$S_{\text{instance}}(U, C) = \frac{f(U_a \cap C_a)}{f(U_a \cap C_a) + \lambda f(U_a - C_a) + \mu f(C_a - U_a)}$$

其中 U_a 和 C_a 表示概念的属性集, f 为一个非负度量函数, $U_a \cap C_a$ 表示两个概念拥有的相同属性的集合, $U_a - C_a$ 表示用户本体概念中含有而情境本体概念里没有的属性集,同理 $C_a - U_a$ 表示情境本体概念里有

而用户本体概念里没有的属性集。 λ, μ 可以看作是权重调节系数,且满足 $\lambda, \mu \geq 0$ 。公式的返回值在区域 $[0, 1]$ 内。

2.2.3 概念实例相似度

概念的实例是构成本体必不可少的部分,如果两个概念下面的实例是完全相同的,则可以判定这两个概念也是相同的,如果两个概念下面的实例大多相同,则可以说这两个概念是相似概念,因此计算概念实例相似度公式如下:

$$S_{\text{instance}}(U, C) = \frac{P(U, C)}{P(U, C) + P(U, \bar{C}) + P(\bar{U}, C)}$$

其中 $P(U, C)$ 表示从实例空间随机抽取一个实例同时从属于概念 U 和 C 的概率。约定将子概念的实例也看作是它的父概念的一个实例。

2.2.4 概念关系相似度

本体关系相似度的计算是一个很繁琐的过程,为简单起见,文中只考虑层次语义关系,选择了三个在所有关系中比例最大的关系作为相似度影响因素,分别为同义关系、继承关系和包含关系。在三种关系的对比中,发现同义关系两边的节点表示的是同一个含义,因此同义关系的权重应当大于其它两种关系的权重。三种关系的权重分别设为 $\text{value}(t)$ 、 $\text{value}(j)$ 、 $\text{value}(b)$,且 $\text{value}(t) > \text{value}(j) > \text{value}(b)$ 。

概念关系的相似度计算利用语义邻居的概念,在情境本体 U 和用户本体 C 中以实体为中心向周围取这三种关系内的实体作为概念的语义邻居,通过计算两个语义邻居集合中本体概念关系相似度大小,选择出与要计算的某个用户最相似的情境。概念关系相似度计算过程类似于概念名称相似度计算,记为 $S_{\text{relation}}(U, C)$ 。

2.2.5 情境综合相似度

基于上面的规则,可以得出用户情境本体和知识情境本体的最终相似度值,计算公式如下:

$$\text{Sim}(U, C) = \alpha S_{\text{name}}(U, C) + \beta S_{\text{attribute}}(U, C) + \lambda S_{\text{instance}}(U, C) + \mu S_{\text{relation}}(U, C)$$

其中 $\alpha, \beta, \lambda, \mu \in (0, 1)$ 且满足 $\alpha + \beta + \lambda + \mu = 1$, 分别代表了概念名称、属性、实例和关系在相似度计算中的权重。 $\text{Sim}(U, C)$ 的取值范围是 $[0, 1]$ 之间的实数,即满足 $\text{Sim}(U, C) \in [0, 1]$, 相似度值越高表明概念越相似,通过预先设定相似度阈值 $\text{Sim}_{\text{set}}(U, C)$, 将相似度值 $\text{Sim}(U, C) > \text{Sim}_{\text{set}}(U, C)$ 的情境概念选取出来做接下来的知识服务匹配。

2.3 知识-情境-用户匹配模型

定义2: 设知识用户情境为 U_k , 则推送的知识服务集合可表示为 $K_s, \{K_s \in K_T^D \mid (U_k \approx C_i) \wedge (P_{i,j} \in K_T^R) \wedge (K_s \subseteq D_j)\}$ 。

通过情境-知识关系模型的建立,基于情境的知识推送步骤表示如下:

步骤1 对特定知识使用者检索用户情境得出用户情境,用户情境 U_m 包含情境 u , $\{U_m \subseteq U_T \mid u \subseteq U_m\}$ 。

步骤2 利用2.2节 UCSim 算法计算出与用户情境相似的知识情境并进行排序,得相似情境集: $\{C_n \subseteq C_T \mid (U_m \approx C_u) \wedge (C_u \subseteq C_n)\}$, 其中 $U_k \approx C_i$ 表示对于某一知识用户,其用户情境为 U_k , 情境 C_i 与之相似。

步骤3 检索 O^R , 得出含情境 C_n 的知识-情境关系对 $P_{i,j}$, $\{P_{i,j} \subseteq P_T \mid (P_{i,j} \subseteq \exists R_1. C_i) \wedge (C_n \subseteq C_i)\}$ 。

步骤4 检索 O^D , 得出与 $P_{i,j}$ 有关联的领域知识集, $\{D_k \subseteq D_T \mid P_{i,j} \subseteq \exists R_2. D_k\}$ 。

步骤5 从相应的知识集中得到知识推送的备选集合,知识服务集, $\{K_s \in K_T \mid K_s \subseteq D_k\}$ 。

3 知识推送

在文中提出的基于情境知识建模和本体相似度基础上,借助本体建模工具 protégé 实现对本体的编辑和维护,本体推理需要借助于 Jena 应用程序接口,知识推送实现框架见图1。

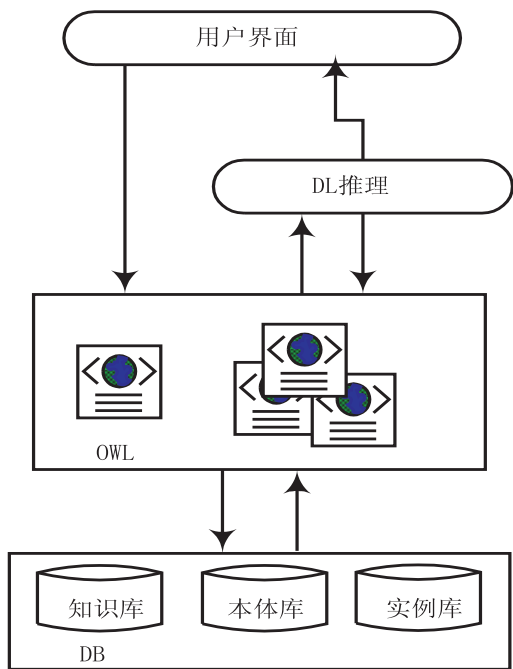


图1 知识推送框架

制造领域知识管理工具框架底层为数据存储层,主要负责制造领域知识及用户信息的本体存储;中间推理层包含了本体的语义推理以及情境相似度匹配计算,并负责推送引擎的实现;表示层负责根据用户最容易接受的方式将知识展示给用户。

在上述知识推送模型中,知识用户对系统每次推

送的知识服务可以进行评价,知识用户对知识服务的评价将作为知识推送引擎选择某条知识实例为服务推送对象的一个依据。知识推送系统将是一个持久的过程,随着系统的持续应用,推送结果将更加全面和准确,以达到企业知识共享的目的。

4 结束语

文中研究了云制造领域基于情境的知识推送技术,通过情境化技术让领域知识与实际应用紧密结合,知识的本体表示能更加准确地描述领域概念,提出的综合本体相似度计算算法使计算结果更加全面和准确。跨学科制造领域知识管理工具中知识推送的初步应用证实了该理论的可行性。文中依然存在许多需要改进的地方,比如提出的情境相似度比较方法中许多权重是根据经验设置的,情境的识别获取也值得进一步研究。

参考文献:

- [1] Ricadela A. Microsoft's Knowledge Push[N/OL]. [2000-09-25]. <http://www.information.com>.
- [2] 周明建,陶俊才.知识管理系统中的知识推送[J].计算机辅助设计与图形学报,2006,18(8):1218-1223.
- [3] 冯勇,樊治平,冯博,等.企业客户服务中心知识推送系统构建研究[J].计算机集成制造系统,2007,13(5):1015-1020.
- [4] 王海林.基于OWL DL的学科知识本体建模[J].计算机技术与发展,2010,20(8):108-111.
- [5] 朱林立,夏幼明,李军华,等.一种改进的模糊知识匹配方法-IDM法[J].计算机技术与发展,2008,18(8):140-143.
- [6] 潘旭伟,顾新建,王正成,等.集成情境的知识管理方法和关键技术研究[J].计算机集成制造系统,2007,13(5):971-977.
- [7] 潘旭伟,顾新建,程耀东,等.集成情境的知识管理模型[J].计算机集成制造系统,2006,12(2):225-230.
- [8] Bobillo F, Delgado M, Gómez-Romero J. Representation of context-dependant knowledge in ontologies: A model and an application[EB/OL]. 2007. <http://descai.ugr.es/~fbobillo/papers/ESWA08a.pdf>.
- [9] 丁政建,张路.一种改进的本体相似度计算方法[J].计算机工程,2010,36(24):39-41.
- [10] 郑诚,秦多荣.本体映射中的概念相似度计算[J].计算机技术与发展,2008,18(11):125-127.
- [11] 徐德志,谭毅.基于本体映射的多本体查询方法[J].计算机技术与发展,2008,18(11):13-17.
- [12] 刘文剑,郭宁,金天国.制造资源本体的相似度计算模型[J].计算机集成制造系统,2010,16(11):2468-2475.