

知识驱动的企业间协作参与者动态推荐方法



王铁鑫^{1,2} 李文心¹ 曹静雯¹ 杨志斌^{1,2} 黄志球^{1,2} 王 飞¹

1 南京航空航天大学计算机科学与技术学院 南京 210016

2 高安全系统的软件开发与验证技术工业和信息化部重点实验室(南京航空航天大学) 南京 210016

摘 要 信息技术的飞速发展有力地推动了市场全球化的进程。全球化的经济趋势给中小型企业带来了前所未有的机遇与挑战,孤岛式的企业发展模式已失去了生存空间。为了快速响应多变的市场需求,中小型企业在聚焦核心业务的同时,须与其他企业建立动态的协同合作关系。面向如何高效构建动态协作企业联盟的问题,通过构建相关领域本体并结合使用语义检测技术,提出了一种知识驱动的企业间协作过程中动态推荐最佳参与者的方法。该方法致力于弥补传统企业协作中“合作参与者固定”“合作模式单一”等缺陷,针对特定协作目标与协作偏好,通过多维度的匹配计算,来快速高效地推荐能力、属性相匹配的参与者企业。在理论层面,针对企业建模及企业协同合作管理,归纳总结模型驱动的企业管理相关方向的研究现状;在此基础上,结合企业间协同合作目标的动态变化特点,定义了描述企业间协作上下文的元模型并构建了企业间协作领域本体及结合该领域本体使用的语义检测方法,以提高动态推荐参与者企业的效率。通过可拆卸连接机械制案例论证了所提推荐方法的有效性,并对其性能进行了评估。

关键词: 企业间协作;领域本体;知识驱动;语义检测;元模型

中图法分类号 TP311

Knowledge-driven Method Towards Dynamic Partners Recommendation in Inter-enterprise Collaboration

WANG Tie-xin^{1,2}, LI Wen-xin¹, CAO Jing-wen¹, YANG Zhi-bin^{1,2}, HUANG Zhi-qiu^{1,2} and WANG Fei¹

1 College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 210016, China

2 Key Laboratory of Safety-Critical Software (Nanjing University of Aeronautics and Astronautics), Ministry of Industry and Information Technology, Nanjing 210016, China

Abstract The rapid development of information technology strongly promotes the process of market globalization. The trend of economic globalization has brought unprecedented opportunities and challenges to small and medium-sized enterprises (SMEs). Enterprises can no longer survive in an isolated-island way. In order to quickly respond to the changing market demand, SMEs need to establish dynamic collaborative relationship with other enterprises while focusing on their core business. To solve the problem of how to construct dynamic collaborative enterprise alliance efficiently, a method of dynamically recommending the best partners in the process of inter-enterprise collaboration is proposed by building domain ontologies and using semantic detection technology. This method aims to break through the defects of traditional enterprise collaboration, such as “fixed cooperative participants” and “single cooperative mode”, and can quickly and efficiently recommend competent participants considering the matching between the specific cooperative goals (& preferences) and capabilities and attributes. By studying enterprise modeling and enterprise collaboration management and summarizing the research status of model-driven enterprise collaboration construction methods, a meta-model to describe the context of inter-enterprise collaboration is defined. Furthermore, the corresponding domain ontologies and semantics detection methods are proposed to improve the efficiency of dynamic recommendation of partners. Finally, the effectiveness of this recommendation method is demonstrated by a case study of disassembly and connection machine manufacturing and its performance is evaluated.

Keywords Inter-enterprise collaboration, Domain ontology, Knowledge-driven, Semantic checking, Meta-model

1 引言

信息技术的快速发展有力地推动了市场经济全球化的进

程。全球化的市场经济趋势减弱了地域差异对企业的影响,并加剧了企业间的竞争;对中小型企业而言,孤岛式的发展模式已失去了生存空间。在此背景下,中小型企业为了实现可

到稿日期:2019-07-27 返修日期:2019-11-12 本文已加入开放科学计划(OSID),请扫描上方二维码获取补充信息。

基金项目:国家自然科学基金项目(61872182);工信部重点实验室(NJ2018014)

This work was supported by the National Natural Science Foundation of China (61872182) and Fundamental Research Funds for the Central Universities (NJ2018014).

通信作者:王铁鑫(tiexin.wang@nuaa.edu.cn)

持续发展,需要不断提高自身的信息化程度以增强市场竞争力。有效的做法包括:将有限的资源与精力集中在核心业务上,并通过信息化技术实现企业间的资源共享、优势互补,最终实现共赢。

在市场经济全球化的大趋势下,企业间的交互与协作变得日益频繁,传统的单一企业内部的合作逐渐升级为企业间的协作。组织间协作网络 CNO^[1] (Collaborative Networks of Organizations) 的理念与应用应运而生,其概念如图 1 所示。为了实现特定的协作目标,若干企业可组成一个管理结构清晰、分工明确、工作流程规范的(多层次)临时协同合作组织。

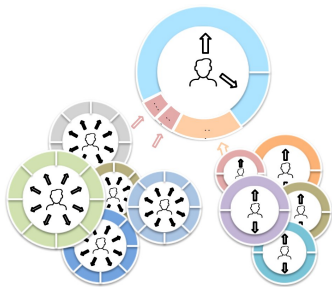


图 1 CNO 的概念示意图

Fig. 1 General illustration of CNO

传统的企业合作往往由企业内的个别员工负责,这种合作方式主要依赖于个人的能力与经验,存在合作企业的选择范围有限、合作模式单一、协作度低等缺点^[2]。受“智能制造”“集成制造”等理念的影响,企业间的“合作”逐步升级为“协作”,呈现出了新的特征,如协作关系的动态性(参与企业按需随时进入、退出合作关系)、协作周期短以及参与企业多元化(异地、异构)等^[3-4]。如何快速、高效地建立高强度的跨企业协作是一项尚未解决的难题。

为了解决这一难题,需要从以下 3 个方面展开研究:1) 如何提升企业自身的互操作性(Enterprise Interoperability)^[5-6]; 2) 如何根据特定的协同合作目标高效地选择最佳的企业参与者; 3) 如何管理(自动推导)多企业间的协作流程。

本文以“如何根据特定的协同合作目标来高效地选择最佳的企业参与者”为研究内容,提出知识驱动的协作参与者推荐方法,主要解决的问题及贡献如下。

(1) 针对如何统一、规范地描述企业间多维度的“协作目标”这一问题,本文定义了一个描述企业间协同合作上下文的元模型 ECMM(Enterprise Collaboration Meta-Model)。该元模型定义协作目标模型、参与企业信息(模型)以及二者间的关联关系的建模方式。

(2) 针对如何根据“协作目标”判别“最佳”参与者企业这一问题,构建企业间协作领域本体,实现“协作目标”到“企业能力”的量化映射规则,设计(模板)用户可扩展的多维度“协作偏好”与“企业属性”间动态映射的机制。

(3) 针对如何根据协作目标自动推荐最佳参与者这一问题,本文提出了基于领域本体的知识推导方法,并结合语义检测实现用户输入与领域本体的自动匹配,从而实现知识驱动的企业协作最佳参与者推荐方法 KdPRM(Knowledge-driven

Partners Recommendation Method for Inter-enterprise collaboration)。

本文第 2 节列举并总结了本文的相关工作;第 3 节在介绍 ECMM 元模型及相关领域本体的基础上,论述了 KdPRM 的总体架构;第 4 节通过可拆卸连接机械制造案例,说明了 KdPRM 的有效性并对其性能进行了评估;最后总结全文并展望未来。

2 相关工作

2.1 协同合作中的参与者选取

企业间协作是“多组织协作”的特例。多组织协作情景普遍存在,如何建立、管理多组织协作是多领域共同的研究课题,如管理学科中的“供应链管理”、计算机学科中的“网络服务组合(web service composition)”等。该问题的研究涉及“多目标优化”“服务发现与匹配”“流程管理”等多个方向。

以企业间的协作为研究背景,文献[2]提出了基于蚁群算法(Ant Colony Algorithm)的协作流程推导方法。该方法应用的前提条件是:假设协作参与者已知,且参与者的角色固化。

以构建虚拟企业为研究背景,文献[7]提出了基于禁忌搜索算法(Tabu search algorithm)的最佳参与者选取方法。该方法将企业间的协作视为一个有中心点(中心决策者)的网状模型,潜在参与者都是网络中已知的节点。首先设定多维全局目标及约束条件,然后执行算法搜索得到参与者集合。文献[8]提出了基于多代理的本体驱动的参与者(多层次)选取方法。该方法在每个代理筛选过程中使用独立的本体,使用“模糊-层次分析法”来逐层缩小参与者的选取范围。该方法在选取过程中使用语义相似度检测方法,但没有考虑各代理本体之间的协同一致性问题。

以供应链管理为研究背景,文献[9]提出了基于遗传算法(Generic Algorithm)的最佳协作参与者选取理论。该理论假设协作的具体流程已知,根据已知协作流程所需的服务选取协作参与者。参与者选取是一个多目标决策过程,通过考虑多目标的综合影响,选取全局最优解。

以网络服务组合为研究背景,文献[10]提出基于语义检测的业务服务(BPMN 模型)到网络服务(BPEL 代码)的映射方法。映射转换建立在基于特定知识库的语义推理上。文献[11]提出使用语义网络及知识驱动相结合的方法来实现网络服务组合。

综上,在管理学科中多组织协同合作参与者推荐(选取)多使用启发式算法;在计算机学科中,该问题的解决多倾向于使用知识驱动及语义相似度检测相结合的理论与方法。

2.2 模型驱动与企业间协同合作管理

在企业管理领域结合使用模型驱动(架构)方法得到了学术界和工业界的共同关注,并取得了一系列的研究成果。Chen 等提出的企业互操作性框架 EIF^[12] (Enterprise Interoperability Framework)、Kosanke 等提出的 CIMOSA^[13] (CIM Open System Architecture),以及 Benaben 等提出的中间件信

息系统 MISE^[14] (Mediation Information System Engineering)等。

EIF 研究框架包含 3 个维度:关注 (concerns)、障碍 (barriers)、解决方案 (solutions)。其中,关注分为数据、服务、过程、业务 4 个层次,障碍包括概念、技术、管理 3 个层面,3 种解决方案分别为联合、统一与集成。该框架有效地对应了模型驱动理论中元模型的理论与应用。

CIMOSA 旨在定义并维护企业运营与管理过程中的知识库,并利用该知识库对企业决策进行有效的支持。企业建模在构建企业知识库的过程中扮演着重要的角色,在整合企业工作流与做出运营决策时发挥着至关重要的作用。

MISE 是基于模型驱动理论的层次性理论架构。其以抽象的模型层为核心,可支持多种应用,如企业协同管理、危机管理等。工作流程分为 4 个阶段:1)构建合作状态模型(静态);2)构建合作行为模型(动态);3)实现可运行的工作流模型;4)部署并运行工作流模型。

以上 3 项研究均围绕模型驱动的企业管理问题开展。

2.3 工业工程领域本体

本体 ontology 是领域概念(词汇)及其关系的规范化描述,具有计算机可读(可处理)性。知识本体呈现领域中一致公认的知识,反映相关领域中公认的概念集。本体可以用于知识表达、查询,是语义网的基础,可以有效地支持自动化的基于语义的服务发现、匹配与组合。目前,广泛使用的知识本体描述语言主要包括资源描述架构模式 RDFS(Resource Description Framework Schema)与网络本体语言 OWL(Web Ontology Language)等。

工业工程领域中,面向不同的应用场景,多种知识本体、标准性文献被提出并应用于实践。表 1 列举了 7 种工业工程领域本体,它们服务于工业界的实践应用。同一领域内的(大部分)知识本体是可扩展且支持复用的。

表 1 工业工程领域本体

Table 1 Ontologies of enterprise engineering

序号	名称	面向应用
1	AIAI Enterprise Ontology	企业建模
2	Toronto Virtual Enterprise Ontology	虚拟企业
3	The business Process Management Ontology	商业流程管理
4	Process Specification Language Ontology	生产应用
5	MASON	工业生产相关
6	ADACOR	工业生产相关
7	Collaboration Network Organization Ontology	协同组织网络

当前,关于企业间协同合作的研究工作主要集中于企业互操作性(如 EIF,CIMOSA 等)及合作过程管理(自动推导)(如 CIMOSA 和 MISE),但围绕推荐、选取合格参与企业的研究工作相对缺乏。EIF,CIMOSA 和 MISE 的关注点如表 2 所列。

表 2 企业间协同合作研究工作的对比

Table 2 Comparison of inter-enterprise collaboration related works

	合作企业选择	互操作性	合作过程管理
EIF	—	是	—
CIMOSA	—	是	是
MISE	—	—	是

2.4 语义检测技术

语义指数据信息的含义,表示数据在某个领域的解释和逻辑。如文献[15]所述,语义检测主要用于处理某个元素或图标背后的意义,可用于判断两个确定对象或者概念之间的语义关系(相似度、相关度)。文献[16]对语义相关性度量的相关技术与方法做了系统的比较。随着语义检测技术的逐步成熟,其应用领域越来越广泛,语义相似度计算在服务发现、问答系统、推荐系统等都有广泛的应用。表 3 列举了 5 项应用语义检测技术的研究工作。

表 3 语义检测的应用研究

Table 3 Research works of semantic checks

相关研究	应用领域	应用目的
文献[17]	模型转换	整合建模语言
文献[18]	数据管理	数据设计匹配
文献[19]	知识工程	本体整合
文献[20]	特征词抽取	文本分析
文献[21]	情感计算	舆情分析

应用领域的差异性导致语义检测技术的使用目的与使用方式有差别。现有的语义相似度计算方法大致可分为基于知识库和基于语料库两种,前者需要领域专家构建结构化的知识表示,而后者通过将词汇或文档表示为向量形式,利用统计方法计算向量特征值,进而计算出语义相关度。

通常,语义和语法检测技术可以结合使用。语法检测专注于句子与单词的组成结构规则;在使用更少资源的情况下,语法检测可以揭示某些潜在的语义关系。因此,通过结合语法检测技术,可以有效地提升语义检测方法的性能。

结合领域本体使用语义相似度检测可以实现知识的推导与衍生,是目前知识驱动方法中普遍使用的技术。

3 KdPRM 概述

KdPRM 是知识驱动的面向企业协作的最佳参与者推荐方法。如图 2 所示,使用 KdPRM 需要首先定义两种规范模型:合作需求模型和企业参与者信息模型(前期工作成果)。

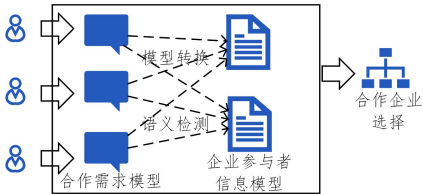


图 2 KdPRM 方法概览

Fig. 2 General illustration of KdPRM

合作需求模型描述了特定的合作目标和合作偏好;企业参与者信息模型描述了企业的能力及其相关属性。模型转换与语义检测被用来在上述两类模型间自动建立关联,实现由目标到最佳参与者的映射。为了规范以上两种模型的构建并实现建模元素间的关联,定义了描述企业间协同合作上下文的元模型 ECMM。ECMM 能够统一、规范地描述企业间多维度的协作目标。

3.1 ECMM 元模型

元模型是对模型的进一步抽象,其定义了模型构建规

则^[22-24]。常用元模型的建模语言包括 Ecore, KM3^[25] 及 MOF^[26-27] 等。

图 3 展示了 ECMM 的基本元素及其关系。协同网络代表由一组用户(企业)构成的协同合作组织,具有自包含性。用户作为协同网络的参与者既可以是组织者也可以是参与者。用户作为组织者时,需要为协同网络定义合作目标与合作偏好;作为参与者时,需要描述其具体的合作能力与属性。

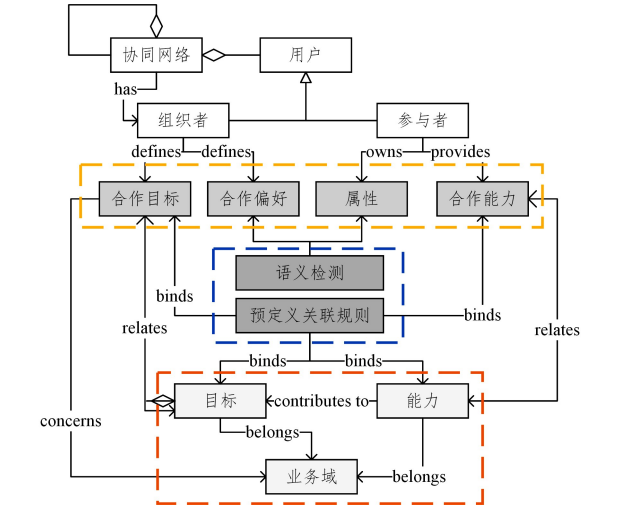


图 3 企业间协作元模型 ECMM

Fig. 3 Inter-enterprise collaboration ECMM meta-model

业务域的定义请参考领域本体的定义,用于限定用户业务(生产、服务)的选取范围,从而有效缩短潜在合作者的搜索时间,提高选取效率。目标与能力是领域本体内部维护的二元组,定义了为实现某特定目标所应具备的一组能力。此二元关系抽取自领域协同本体。

语义检测用于自动创建用户输入(合作目标、合作能力)与领域本体(目标、能力)间的关联关系,从而实现领域本体到用户输入的映射。预定义关联规则定义了一组预定义的用户输入合作偏好与企业属性之间的对应关系。

以上用户信息都是以模型实例的形式作为 KdPRM 的输入。本文依据建模元素的作用以及建模信息的来源对 EC-MM 中的建模元素进行了分类,如表 4 所列。

表 4 ECMM 建模元素

Table 4 ECMM modeling elements

	元素名称		备注
	协同网络		自包含
辅助建模元素	用户		具有唯一标识
	组织者		与协同网络一对一
	参与者		亦可成为组织者
用户输入	合作目标		应与目标对应
	合作偏好		与属性相关
	合作能力		应与能力对应
	属性		与合作偏好相关
领域本体	目标		关联一组能力
	能力		用于实现目标
	业务域		关联用户
KdPRM 内	语义检测		关联用户输入及领域本体
	预定义关联规则		

(1)预定义关联规则(可扩展属性集)

在 KdPRM 中合作偏好是筛选参与者的次要参考依据;当合作组织者输入的合作偏好与参与者输入的属性相匹配时,该企业作为最佳参与者被选中的概率会相应地提高。在 KdPRM 中,合作偏好与属性是两组可扩展属性集,其属性的类型被限定为枚举及布尔类型。

如图 4 所示,当前合作偏好有 4 个属性,即预算、位置、质量、安全,这些属性的数据类型都是枚举类。例如,在构建一个协同网络时,位置偏好可以限定在 城市、区县等。参与者的属性有 6 种,即位置、所属地区(发展程度)、员工数量、建立时间、企业类型与科技属性,以建立时间为例,它描述了企业的创建时间,可选的枚举项有:1)大于 10 年;2)介于 5~10 年之间;3)介于 1~5 年之间;4)小于 1 年。

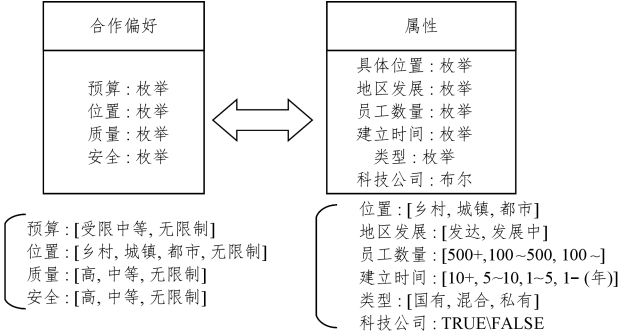


图 4 可扩展预定义属性集及关联规则

Fig. 4 Extensible predefined property set and the association rules

两组属性集间的关联关系是多对多的。一种合作偏好可能关联到多个合作企业属性,如合作偏好中的属性质量可以关联到属性中的科技公司与建立时间(建立时间与企业服务质量正相关)等。协作组织者定义协作目标时,要依据真实的协作背景,为不同合作偏好定义适当权重,并为与协作偏好相关的企业属性定义影响合作偏好的具体权重值。

(2)用户输入模型

如图 5 所示,基于 ECMM 元模型定义的组织者模型由业务域以及合作目标与合作偏好构成。其中合作偏好的输入由系统模板限定为具体的枚举类型,合作业务域从用户自业务域本体中选定,如果用户自主输入其业务域,系统则通过模糊匹配的方式自动将其映射到业务域本体中的相近业务。

业务	业务域 限定性描述	额外
ID	UNSIGNEDINT	AUTO INCRE
名称	STRING	NOT NULL
类型	ENUM	NOT NULL
...	...	...

组织者业务域

键	值	额外
目标	ENUM/STRING	NOT NULL
时间	DATE	PROPERTY
位置	STRING	PROPERTY
质量	ENUM	PROPERTY

组织者需求

图 5 组织者输入模型

Fig. 5 Input model from organizers

在 ECMM 元模型中,参与者模型需要描述合作能力及其相关属性。其中,合作能力由用户参考协作本体自行定义,相关属性由参与者依据系统输入模板按规则填入。参与者输入模型如图 6 所示。

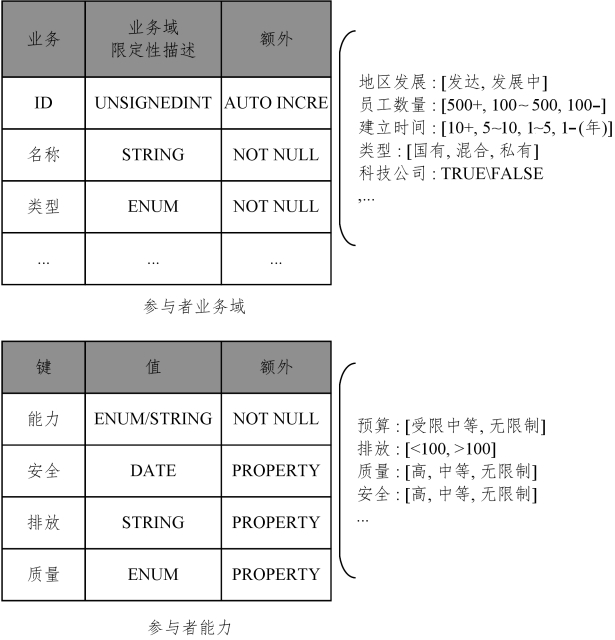


图 6 参与者输入模型
Fig. 6 Input model from partners

协作偏好与参与者属性模型及其关系的设定是本文对前期工作的扩展,亦是本文的创新点。

3.2 KdPRM 中的服务发现计算方法

如上文所述,在 KdPRM 中引用了两个工业工程领域本体^[28](前期工作成果):协作本体 CO(Collaboration Ontology)和业务域本体 BFO(Business Field Ontology)。其中,CO 是基于 MIT Process Handbook^[29]构建的,与 ECMM 中的目标、能力及两者之间的关系相对应;BFO 是基于工业分类国际标准 ISIC(International Standard Industrial Classification)^[30]构建的,与 ECMM 中的业务域相对应。

如图 7 所示,CO 中的目标可以进一步分解为一组子目标,每个子目标又可以进一步关联到一组能力。为了实现特定目标,所有与该目标关联的能力都应由参与企业提供。BFO 包含业务域与子业务域,二者之间是继承与泛化的关系,业务域的限制有助于提高选取合作企业的效率与准确性。

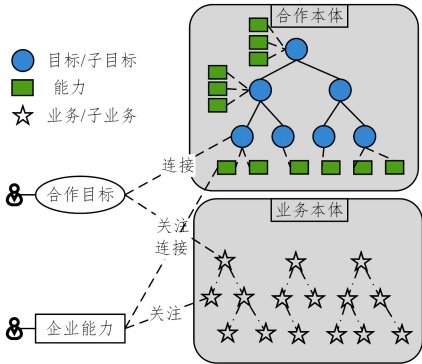


图 7 领域本体 BFO 与 CO 的结构
Fig. 7 Structures of CO and BFO

CO 与 BFO 内定义的具体条目通过语义检测技术与系统用户的输入(如协作目标、合作能力等)自动关联,是实现知识驱动的协作参与者自动选取(服务发现)的基础。

基于 BFO 的业务域匹配计算过程如下:首先,提取协作组织者输入模型中的业务域(结构化)信息;然后,通过复合语义检测,在参与者之间进行不区分上下位关系的一对多业务域匹配;最后,利用 BFO 中定义的业务域之间的(继承及泛化)关系,结合预先设定的判断阈值,初步筛选潜在的协作参与者企业范围。

基于 CO 的协作目标与合作能力匹配过程如下:基于 BFO 业务域匹配过程,筛选出潜在的参与者企业(缩小筛选范围);进一步的筛选则需要验证参与者提供的能力是否可以满足特定的协作目标;最后,基于组织者需求与 CO 协作本体,获取一个与组织者需求相关的合作能力向量,并将该向量与潜在参与者的能力描述模型的能力向量进行复合语义检测,通过判断向量距离得出合作域的匹配值。

在不使用预定义关联规则做协作匹配计算的情况下,经过以上步骤计算得出匹配值最高的企业会被选定为潜在参与者。如果采用预定义关联规则进一步细化组织者的协作需求,则可以通过计算合作偏好模型实例与属性模型之间的语义距离得到权值 k ,并将其作为影响最终协同合作匹配度的因子。

KdPRM 结合使用语法检测和语义检测,语法检测作为语义检测的补充。

语义检测阶段采用部分波特词干提取算法^[31],用于检测是否存在特殊的语义关系(如单复数变型情况等);如未发现符合检测规则的语法关系,则执行语义检测。语义检测包括对直接语义关系的检测和迭代语义关系的检测。如果无法确定一对比较概念(字符串)间的语义关系,则通过编辑距离^[32]算法进一步进行语法检测。

语义相关性 SCV(Semantic-based Comparing Value)的计算方法如式(1)所示:

$$SCV = \begin{cases} 1, & \text{判定为相同概念} \\ SrV, & \text{判定出语义关系} \\ SsV, & \text{编辑距离算法输出} \end{cases} \quad (1)$$

SCV 的值域为 0 到 1。在算法执行过程中,如果检测到对比词对间有语义关系存在,则 SCV 值由语义关系值 SrV (Semantic relation Value)决定,如果未检测到语义关系,则 SCV 值由 SsV (编辑距离算法测定的 syntactic similarity Value)决定。

KdPRM 中语义检测的过程如图 8 所示。在企业合作匹配的过程中,主要考虑如下几种关系:A 业务是否等同 B 业务、A 业务与 B 业务的效果是否相同、企业 X 是否具有 A 技术、实现 B 技术的基础是具有技术能力 C 和 D、A 业务的效果与 B 业务相似但不相等等,这些场景分别对应着同等、同义、反义、下位、类似等语义关系。表 5 列出了词缀检测的 8 种情形。

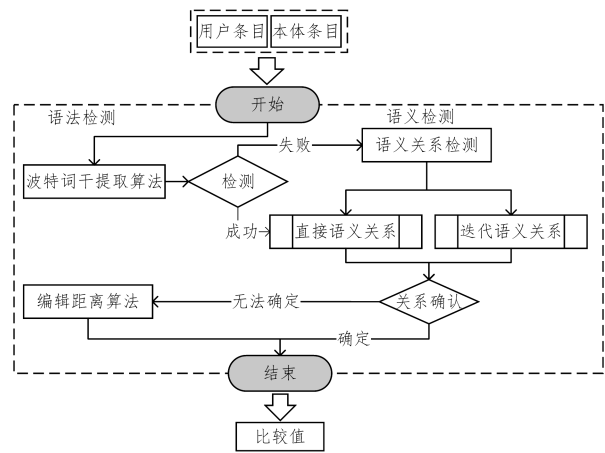


图 8 语义检测算法的流程

Fig. 8 Flow chat semantic-based checking algorithm

表 5 8 种被检测的词根词缀情形

Table 5 Eight stemming situations being detected			
前缀	后缀	例子	语义关系
1	s,es	product &. products	同等
2	ing	transport &. transporting	同义
3	ness	great &. greatness	同义
4	a,ab,abs	normal &. abnormal	反义
5	sub	section &. subsection	上位
6	super	super &. superstar	上位
7	demi-, semi-	Automatic &. semi-automatic	类似
8	nent, ment	manage,management	同义

KdPRM 中的编辑距离算法是基于 Levenshtein distance algorithm^[33]进行扩展实现的,其计算结果为 SsV 值。字符串间的编辑距离是单字符最少编辑次数,即将一个字符串(a)转换为另一个字符串(b)所需要的最少插入、删除或替换的操作步骤。该算法的使用在本文的前期工作中有详细的介绍。

语义检测阶段 KdPRM 中的语义检测采用基于知识库的概念语义相似度计算方法,使用的语义库为通用语义数据库 WordNet^[34]。为了更好地结合应用背景,在 WordNet 基础上定义新的领域词义库 IeC_ST(Inter-enterprise Collaboration Semantic Thesaurus),并利用 Neo4J 实现了 IeC_ST 的可视化查询与展示。

IeC_ST 包含 147 306 个词汇、206 941 个词义(义元)和 114 038 个近义词集。其结构与应用示例如图 9 所示。

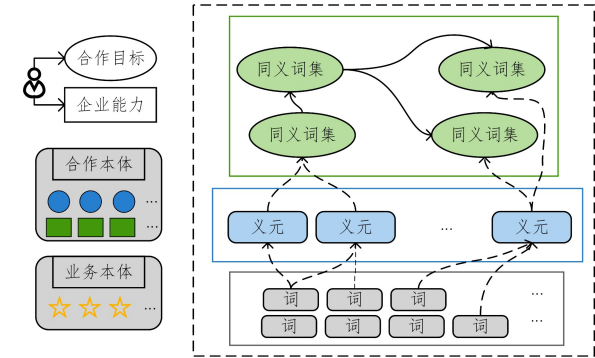


图 9 IeC_ST 的结构与应用

Fig. 9 Structure and usage of IeC_ST

对任意一组 CO(或 BFO)与用户输入的字符串,其语义关系检测包括 4 个步骤:1)在 IeC_ST 中定位两个比较字符串;2)搜索比较字符串的所有词义,形成两组词义集合;3)追踪词义所属的近似词集,构成两个近义词集合;4)检测每一对近义词集组合之间的语义关系。

如表 6 所列,KdPRM 检测 7 种直接语义关系及其嵌套语义关系。

表 6 在 KdPRM 中检测的语义关系

Table 6 Semantic relations being detected in KdPRM			
语义关系	SrV	示例	
直接语义关系	同根	0.90	good—better
	同义	0.85	good—goodish
	上位词	0.70	creator—person
	下位词	0.90	person—creator
	部分组成词	0.55	screw—bolt
	部分构件词	0.75	bolt—screw
	同等	0.95	telephone &. phone
直迭代语义关系	迭代上位词	0.7 ⁿ	maker—creator—person
	上位词 & 同等	0.7 * 0.95	phoneme—telephone
...	...	...	...

两个比较词之间可能存在多重语义关系(因其拥有不同的词义所致),亦可能通过计算得到多个 SrV 值。考虑到 KdPRM 的应用目的——推荐最佳参与者企业,在存在多组 SrV 值的情况下只取其最大值,SrV 的计算方法如式(2)所示:

$$SrV=\max\{\prod_{j=1}^{n_i}sr_{v_i_j}\},i\in[1,m]$$
 (2)

其中,srv_i_j 为当前比较词之间第 i 条语义关系通路中第 j 个关系的 SrV 值,其中假设共有 m 条语义关系通路,在第 i 条语义关系通路上共有 n_i 层迭代语义关系。

4 案例分析

4.1 案例描述

聚焦 KdPRM 的核心思想与执行过程,案例着重展示了基于领域本体与语义检测的协作目标与合作能力的自动匹配过程,并对相关模型的构建过程、合作偏好与企业属性的匹配环节等做了适当简化。本文对案例做出了以下约定。

在相关模型的构建过程中,用户输入的业务信息、需求目标和能力属性等都采用了结构化的描述方式(基于 ECMM 的标准模型,如图 4—图 6 所示)。

用户在输入信息的过程中,对 BFO 和 CO 中的知识或行业背景知识有一定的掌握。

在模型的匹配过程中,将参与者输入模型作为原模型,将组织者输入模型作为目标模型。

实现 CO 协作目标与能力向量匹配的语义计算工作时,做出如下 5 组假设:1)针对一个特定协作目标——生产可拆卸钉子,提取出了需求能力向量[螺丝钉,螺母];2)此协作目标与业务本体 BFO 中的可拆卸连接机械制造业务相关;3)基于业务域 BFO 相关性阈值的限定,初步筛选出 3 个企业成为潜在参与方;4)语义检测规则将在这 3 个企业的合作能力与协作本体 CO 中对应能力生产螺丝钉之间进行应用;5)此案例暂不考虑非功能性因素(合作偏好与企业属性)的影响。基于 CO 的合作匹配过程如图 10 所示。

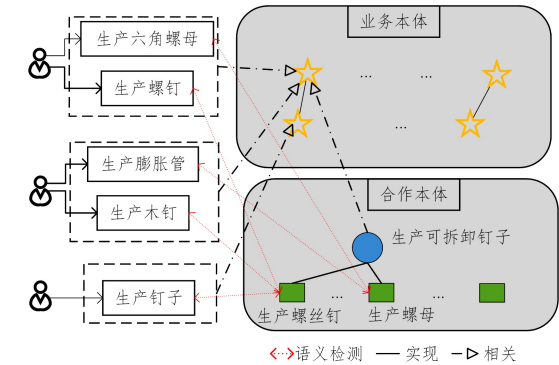


图 10 案例描述

Fig. 10 General overview of use case

在匹配计算过程中,语义检测的规则限定在特定的语义关系上,如同根关系、上位词关系、下位词关系、部分组成关系、部分构件关系等。

4.2 基于语义检测的能力匹配计算

潜在参与者协作能力与组织者的协作目标关联能力向量(CO 决定)构成了 3 组二维对:[螺钉 bolt-螺丝钉 screw,六角螺母 hex_nut-螺母 nut],[钢钉 nail-螺丝钉 screw,无 null-螺母 nut] 以及[木钉 nog-螺丝钉 screw,膨胀管 stuffing_box-螺母 nut]。

在语法检测阶段未发现词缀、词根的情况下,对以上 3 组对比词向量执行语义检测。以螺钉 bolt-螺丝钉 screw 检测为例,结果如图 11 所示。

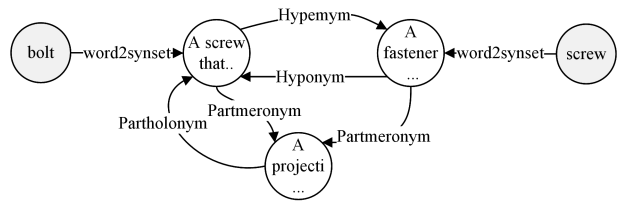


图 11 螺钉 bolt-螺丝钉 screw 语义检测结果

Fig. 11 Semantic detection between bolt and screw

在向量单词对[螺钉 bolt-螺丝钉 screw,六角螺母 hex_nut-螺母 nut]中 bolt 是 screw 的直接下位词,同时 hex_nut 是 nut 的直接下位词,则此词对的 SCV 为(最大)SrV,即 $(0.9 + 0.9)/2 = 0.9$ 。

在向量单词对[钢钉 nail-螺丝钉 screw,无 null-螺母 nut]中,nail 与 screw 之间的语义关系是一层迭代的同等下位词关系以及同等的一层迭代的同等部件词,而在该参与者能力向量中,与生产螺母相关的生产能力为 null,因此此单词对的 SCV 为 SrV,即 $(0.9 * 0.7 + 0)/2 = 0.315$ 。

在向量单词对[木钉 nog-螺丝钉 screw,膨胀管 stuffing_box-螺母 nut]中,nog 与 screw 之间的语义关系是多层迭代语义关系,其中关系图连接的关系为下位词关系、上位词关系、部件词关系。同时,stuffing_box 与 nut 之间也是多层迭代语义关系,顺序为部件词关系、下位词关系。因此此单词对之间的 SCV 为 SrV,即 $(0.9 * \text{MAX}(0.7 * 0.9, 0.75 * 0.55)) + 0.9 * 0.55)/2 = 0.531$ 。

综上所述,参与者能力向量为[螺钉 bolt,六角螺母 hex_nut]的参与者与组织者需求目标匹配度最高。在仅考虑合作

企业能力与实现协作目标所需能力的匹配度时,提供最高匹配度的生产能力的参与者应成为合作的对象。

4.3 方法评估

(1)数据评估

本文提出的基于知识驱动的企业间协作参与者推荐方法基于富含语义关系的结构化数据,匹配过程是上述数据基于知识库的推理,具有可视化、可配置、易修改等特性。

已有的相关研究中,企业协作参与者推荐往往是基于海量数据分析且由大量的数据指标驱动的。文献[35]实现了基于多源数据的企业跨产业研发合作伙伴识别方法,数据主要来源于领域资讯相关网站、文献数据库等。文献[36]实现了企业合作型原始创新互惠共生伙伴选择方法,该方法应用了三角模糊数、模糊网络分析、模糊 TOPIS 法实现对潜在伙伴的评价。

以上两个研究都是在分析、研究特定指标后,基于数据分析实现的;从数据量和数据维度的角度,以上研究相对本研究对数据的需求是巨大的,且推算过程有不可解释性、不可视。

表 7 企业合作推荐数据评测

研究	数据来源	数据格式	分析方法	数据量级
文献[35]	网络爬取	非结构化	TOPIS	MB 数据
文献[36]	网络爬取与企业输入	半结构化	模糊 TOPIS	MB 数据
本文	企业输入	结构化	语义相关度计算	MB 先验+KB 数据

从数据角度出发,基于海量数据分析的推荐方案通常可以获得一个近似最优解(企业协作参与者匹配度评估);获得近似最优解的根本原因是底层的描述模型经过了精密的考量与验证。对于推荐企业间协作参与者,可借助合理的企业合作模型,基于有效知识(领域本体)推理方案,实现具有一定指向的服务计算。而本文正是从语义的角度出发考虑企业的表述,而不是从网络上获取无法证明有效性的统计数据。

(2)性能评估

性能评估关注语义关系检测的时间消耗与准确性。运行此案例的平台配置如下:主频 3.3 GHz,CPU I7-7700,RAM 16GB DDR4 2333 Hz,Win10 系统,JDK 为 JAVA 8。

4.2 节的测试结果显示,3 组对比词之间的语义关系分别为:螺钉-螺丝钉(直接语义关系)、钉子-螺丝钉(一层迭代语义关系)和木钉-螺丝钉(多层迭代语义关系)。检测的顺序是先执行直接语义关系检测,再进行迭代语义关系检测(多层语义关系)。图 12 给出了时间消耗性能的评测结果,横轴为检测语义关系的层数,前三层为直接语义关系,单词对之间迭代语义关系从第四层开始。

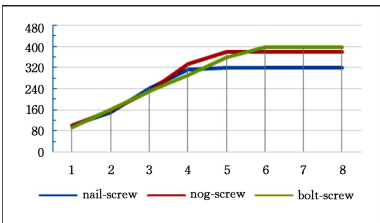


图 12 时间消耗性性能的评测结果

Fig. 12 Evaluation result of time-consuming issue

直接语义关系(前三层)的检测时间按照层数的增加以固定比率递增;一层语义关系的检测时间消耗约 45 ms(由于 Neo4J 基于图存储结构的优化,其时间消耗与是否检测到语义关系无关)。迭代组合语义关系的检测时间消耗与检测语义关系层数的比率与前者相似。因此,计算时间主要取决于首次检测到语义关系的层数。

语义检测可以发现比较词之间潜在的语义关系。检测的语义关系(直接、迭代)的数量越多,发现潜在关联的可能性就越大,消耗的资源也会随之增多。然而,弱化的语义关系(如四层及以上迭代所发现的)不能用于制定转换规则,进而实现最终的映射。因此,对检测直接语义关系的数量、迭代语义关系的层数的限定,都要视具体的应用而定。

4.4 KdPRM 方法讨论

面向企业间协作的最佳参与者选择问题,KdPRM 拟结合使用知识驱动(基于领域本体)及语义检测(基于知识库 WordNet)的方式,来实现针对特定协作目标的最佳参与者自动推荐。

传统的企业间协作参与者的选取依靠企业个别人员的能力与经验,因此存在参与者选取范围小、选取过程不透明及不公平等诸多弊端。KdPRM 能够促进多元化企业成为协作参与方,扩大协作参与方的选择范围,一定程度地屏蔽人为因素的干扰,提高选取过程的公平性与准确性。

基于启发式算法的参与者选取方法^[2,7-9]普遍对算法的输入有较严格的限制,且只有在满足相对严格的假设条件下,算法才能取得相对较好的输出效果。与之相比,KdPRM 为用户提供信息输入模板,在不改变用户输入习惯的大前提下结构化、标准化用户的输入信息;利用领域本体作为知识库,通过语义检测方法建立用户输入与知识库的有效关联。KdPRM 打破了使用启发式算法的诸多限制,具有更好的适用性。

与其他结合领域本体与语义检测技术实现服务发现的研究工作^[2,10]相比,KdPRM 具有以下特点及优势:1)为 KdPRM 构建了两个特殊的领域本体 CO 与 BFO;2)语义检测方法是面向结构化信息(ECMM 标准)与领域本体结构而特别设计实现的;3)结合 Neo4J 图形化数据库,使匹配的过程可视化,便于追踪与验证。

KdPRM 仍在被不断地优化,其中领域本体内容的可持续化更新方法、语义检测方法的效率及相关支持软件的性能都在持续地完善。

结束语 面向动态构建高强度企业间协作联盟问题,针对协作参与者推荐方法,本文结合使用模型驱动、知识驱动与语义检测技术,提出了理论架构并提供了软件工具支持。

本文通过定义元模型来规范用户对企业信息的统一模型描述,统一、规范地描述企业间多维度的协作目标。为了提高模型间转换、整合的效率,定义并构建了领域本体与语义检测方法。该方法根据协作目标判别最佳参与者企业,能够有效解决传统企业间协作人为选取参与者的弊端,并突破基于启发式算法选取参与者的多重限制与约束,在一定程度上提高

了最佳参与者选取的准确性、公平性与选取效率。

参 考 文 献

- [1] CAMARINHA-MATOS L. Classes of Collaborative Networks [M]// Encyclopedia of Networked and Virtual Organizations. Information Science Reference,2008:193-198.
- [2] MONTARNAL A,MU W X,BENABEN F,et al. Automated deduction of cross-organizational collaborative business processes[J]. Information Sciences,2018,453:30-49.
- [3] TOUZI J,LORRE J P,BENABEN F,et al. Interoperability through Model-based Generation:The Case of the Collaborative Information System (CIS) [M]// Enterprise Interoperability. London:Springer,2007.
- [4] LI L. Effects of enterprise technology on supply chain collaboration:analysis of China-linked supply chain [J]. Enterprise Information Systems,2012,6(1):55-77.
- [5] KONSTANTAS D,BOURRIERES J P,LEONARD M,et al. Interoperability of enterprise software and applications [M]. Springer Science & Business Media,2006,1.
- [6] IDE N,PUSTEJOVSKY J. What does interoperability mean, anyway? Toward an operational definition of interoperability for language technology [C]// Proceedings of the Second International Conference on Global Interoperability for Language Resources. Hong Kong,China,2010.
- [7] CRISPIM JA,SOUSA J P. Partner selection in virtual enterprises[J]. International Journal of Production Research,2010,48(3):683-707.
- [8] SADIGH B L,NIKGHADAM S,OZBAYOGLU A M,et al. An ontology-based multi-agent virtual enterprise system (omave): part 2:partner selection[J]. International Journal of Computer Integrated Manufacturing,2017,30(10):1072-1092.
- [9] SHA D,CHE Z. Supply chain network design:partner selection and production/distribution planning using a systematic model [J]. Journal of the Operational Research Society,2006,57(1):52-62.
- [10] BENABEN F,BOISSEL-DALLIER N,PINGAUD H,et al. Semantic issues in model-driven management of information system interoperability[J]. International Journal of Computer Integrated Manufacturing,2013,26(11):12.
- [11] GARCIA-CRESPO A,RUIZ-MEZCUA B,LOPEZ-CUADRA DO J,et al. Semantic model for knowledge representation in e-business[J]. Knowledge-Based Systems,2011,24(2):282-296.
- [12] CHEN D,DOUMEINGTSB G,VERNADATC F. Architectures for enterprise integration and interoperability:Past,present and future [J]. Computers in Industry,2009,59(7):647-659.
- [13] KOSANKE K,VERNADAT F,ZELM M,CIMOSA. enterprise engineering and integration [J]. Computers in Industry,1999,40(2/3):83-97.
- [14] BENABEN F,MU W X,BOISSEL-DALLIER N,et al. Supporting interoperability of collaborative networks through engineering of a service-based Mediation Information System (MISE

2. 0) [J]. Enterprise Information Systems, 2014, 9(5/6): 1-27.

[15] ALI N H, IBRAHIM N S. Porter stemming algorithm for semantic checking [C]// Proceedings of 16th International Conference on Computer and Information Technology. 2012: 253-258.

[16] FENG Y, BAGHERI E, ENSAN F, et al. The state of the art in semantic relatedness: a framework for comparison [J]. The Knowledge Engineering Review, 2017, 32: 1-30.

[17] KAPPEL G, KARGL H, KRAMLER G, et al. Matching Meta-models with Semantic Systems-An Experience Report [M]// BTW Workshops on Model Management. 2007: 38-52.

[18] SHVAIKO P, EUZENAT J. A survey of schema-based matching approaches [M]// Journal on Data Semantics IV. Springer Berlin Heidelberg, 2005: 146-171.

[19] LIN F, SANDKUHL K. A Survey of Exploiting WordNet in Ontology Matching [C]// Artificial Intelligence in Theory & Practice II. Italy: DBLP, 2008.

[20] LI P F, ZHOU G D, ZHU Q M. Semantics-Based Joint Model of Chinese Event Trigger Extraction [J]. Journal of Software, 2016, 27(2): 280-294.

[21] RAO Y, WU L W, WANG Y M, et al. Research Progress on Emotional Computation Technology Based on Semantic Analysis [J]. Journal of Software, 2018(8): 2397-2426.

[22] THOMAS K. Matters of (Meta-) Modeling [J]. Software and Systems Modeling, 2006, 5(4): 369-385.

[23] FAVRE J M, NGUYEN T. Towards a megamodel to model software evolution through transformations [J]. Electronic Notes in Theoretical Computer Science, 2005, 127(3): 59-74.

[24] ALBERTO S R D. Model-driven engineering: A survey supported by the unified conceptual model [J]. Computer Languages, Systems & Structures, 2015, 43: S1477842415000408.

[25] JOUAULT F, BÉZIVIN J. KM3: A DSL for Metamodel Specification [C]// Ifip Wg 61 International Conference on Formal Methods for Open Object-based Distributed Systems. 2006.

[26] OMG. Object Management Group-MDA (Model Driven Architecture) Guide Version 1. 0. 1 [OL]. <http://www.omg.org/mda/>.

[27] OMG. Object Management Group-Meta Object Facility (MOF) Core Specification, v2. 4. 2 [OL]. <http://www.omg.org/mof/>.

[28] WANG T X, MONTARNAL A, TRUPTIL S, et al. A Semantic-checking based Model-driven Approach to Serve Multi-organization Collaboration [OL]. <https://www.sciencedirect.com/science/article/pii/S1877050918311918>.

[29] MALONE T W, CROWSTON K, HERMAN G. A. Organizing business knowledge: the mit process handbook [J]. Journal of Product Innovation Management, 2010, 22(2): 218-219.

[30] United Nations and Statistical Division. International Standard industrial classification of all economic activities (ISIC) [M]. New York; United Nations, 2008.

[31] PORTER M F. An algorithm for suffix stripping [J]. Program, 1980, 14(3): 130-137.

[32] COHEN W, RAVIKUMAR P, FIENBERG S. A comparison of string metrics for matching names and records [C]// Kdd Workshop on Datacleaning and Object Consolidation. 2003: 73-78.

[33] HIRSCHBERG D. Serial computations of Levenshtein distances [OL]. <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.55.5863>.

[34] FELLBAUM C. WordNet. Blackwell Publishing Ltd [OL]. https://xueshu.baidu.com/usercenter/paper/show?paperid=38f9291923c02bf94ea9b712c610f714&site=xueshu_se.

[35] WU F F, LI R M, HUANG L C, et al. Cross-industry R&D Partners Identification and Selection of Enterprises [OL]. <http://kns.cnki.net/kcms/detail/61.1167.G3.20190710.1021.016.html>.

[36] LI B Z, GAO S. Research on the mutualism partner selection of enterprise collaborative original innovation [J]. J. Harbin Inst. Technol., 2019, 40(7): 1367-1374.



WANG Tie-xin, born in 1987, Ph.D, assistant professor, M. S supervisor. His main research interests include model-driven engineering and collaboration management, etc.