

数据驱动的专家知识资源建模方法与原型系统开发研究

张琪 战洪飞 余军合 魏保伟

(宁波大学机械工程与力学学院 浙江 宁波 315211)

摘要 专家知识资源是企业进行业务执行时,最重要的关键资源。快速高效组织专家资源的能力将是决定企业竞争力的重要因素。然而目前企业在知识管理方面的欠缺及专家资源信息获取能力的不足造成很多企业在组织专家资源方面的困扰。缺乏科学合理的专家资源建模方法与技术是其中的关键。大数据的快速发展及网络开放数据的增加,形成数据驱动方法,构建专家知识资源模型。基于该构思,提出一种数据驱动的,基于映射规则的专家知识资源建模方法。利用多源数据融合与抽取技术,构造比较全面的专家知识资源描述模型,为企业配置、业务绩效提升提供方法和技术支持。该研究对提升企业的知识利用能力、提升企业创新水平有重要价值。

关键词 数据驱动 映射规则 专家知识资源模型 多源数据融合 数据挖掘

中图分类号 TP3

文献标识码 A

DOI:10.3969/j.issn.1000-386x.2019.02.012

DATA-DRIVEN EXPERT KNOWLEDGE RESOURCE MODELING METHOD AND PROTOTYPE SYSTEM DEVELOPMENT

Zhang Qi Zhan Hongfei Yu Junhe Wei Baowei

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

Abstract Expert knowledge resources are the most important key resources for business execution. The ability to organize expert resources quickly and efficiently is an important factor in determining a company's competitiveness. However, the current lack of knowledge management and access to expert resource information has caused many companies to be confused about the organization of expert resources. The key is the lack of scientific and reasonable methods and techniques for expert resource modeling. The rapid development of big data and the increase of open data in the Internet, lead to the construction of an expert knowledge resource model with a data-driven manner. Based on this conception, this paper proposed a data-driven modeling method for expert knowledge resources based on mapping rules. It used a multi-source data fusion and extraction technology to construct a more comprehensive model of describing expert knowledge resource, which provided supports of techniques and methods for enterprise configuration and business performance improvement. The research is of great value to enhance the ability of enterprises to use knowledge and improve their innovation level.

Keywords Data-driven Mapping rules Expert knowledge resource model Multi-source data fusion Data mining

0 引言

知识资源是决定企业竞争优势的重要资源,但由于企业知识管理实践的缺乏及知识管理应用理论的不

足,使企业在业务执行中的知识资源配置能力较差,不能够很好地将正确的资源用在正确的业务上。同时,知识资源配置上的不足也造成企业业务执行绩效低下,阻碍企业创新能力的提升。因而,如何有效经营企业的知识资源,在盘活企业内部知识资源的基础上如

何高效利用企业外部的专家资源,这都是对企业竞争力有重大影响的核心问题。然而,目前专家知识资源信息建模技术与方法上的研究成果不能满足企业的需求,企业难以获取这些专家资源信息,优化知识配置更无从谈起。针对这一问题,本文提出基于大数据技术及广域的开源数据,构建比较合理的专家资源模型,为企业开展面向业务执行的专家知识资源提供基础理论参考。为了提高知识资源的求解效率,并及时提供需求的专家知识资源,国内外学者已开展了较丰富的研究。

文献[1]认为知识资源的获取在专家知识系统的构建是需要通过解决的实际问题,并基于“洞察力”的假设理念为知识获取和表示工具提供适合的哲学环境。文献[2]认为知识获取在未来应该从狭隘的技术、结构和模型概念转移到对业务问题求解的专家系统的需求上。文献[3]引入知识获取技术提高组织知识的位置、所有权及影响的理解,而不是机构执行的数据流和任务,并为组织问题解决知识获取提供了指导原则。文献[4]基于问题领域特征来规定某些知识获取技术,以生产和运营管理(P/OM)领域作为研究的试点领域,将P/OM任务映射到问题域的一般分类方式。根据分类的结果构建描述具有显著知识的获取技术。文献[5]在访谈的层面上,结合工业界的高级管理人员担任实验课题,比较非结构化知识和结构化知识获取的方法,发现采用结构化访谈方法获取知识可以提高绩效。文献[6]认为在业务领域中问题以及求解的策略在选择方面存在很大的差异,采用问题分解和简化技术以便为管理决策提供适当的信息。郭健美等^[7]在基于业务问题求解的实际需求方面,提出了一种对于知识供应过程的业务问题求解模型。王德川等^[8]针对企业业务问题求解时方案制定不合理的现象,将企业中的知识库用领域建模的方式进行重新整理建模,设计了一种面向特定领域建模方法。姚平等^[9]为了提取决策知识,采用将模糊集和粗糙集理论相结合的方法获取知识资源。罗琳等^[10]从数据的内涵出发,提出了知识流动的特点,采用“数据-人-知识”模型,构建了其理论框架,应用在产学研协同创新管网研究。刘骊等^[11]以服装数据作为驱动力,构建三维服装模型,提高服装建模的效率。徐建国等^[12]采用数据驱动的方式,采用技术评论数据,应用在技术网络模型构建。杨青等^[13]以工程管理的角度,采用数据驱动的方式,将工程管理按照时间维度、领域维度和方法维度三个维度,挖掘数据信息,及各个生命阶段的数据类型。程进等^[14]以个性化制造企业流程工艺参数数据作为驱动力,研究其参数匹配方法,有效地为制造

工艺知识提供服务。

综合国内外的相关研究可以看出,以数据作为驱动,结合映射规则获得更加全面的专家知识资源应用在业务问题的求解方面的研究和应用较少,尤其是在在专家知识资源获取方法方面仍需要进一步研究^[15-21]。

以开放数据资源进行知识模型构建仍旧存在较多的有价值性的工作。本文从企业业务问题求解进行分析,以数据作为驱动力,利用已有的网络数据资源,基于映射规则的方法进行推理,填充企业业务问题求解的专家知识资源模型,令专家知识资源模型更加全面细致,并研究数据驱动的专家知识资源模型方法,进行专家知识资源原型系统开发。

1 模型构建

1.1 模型维度的确定

面向业务问题求解的专家知识资源建模是从业务问题求解需求的角度出发构建的,在业务的执行过程中,会遇到比较复杂的业务难题或困难,本文将这类困难与难题的解决称为业务问题求解。而对于业务问题求解的发生源于业务问题的出现,业务问题的出现是企业业务人员在执行不同业务活动发生,因此,根据找到业务问题求解的源头,对其进行解析,使出现业务问题时能够及时得到求解的方案,从而继续推进整个业务活动。从中可以发现,业务问题求解的需求知识与专家知识资源的构建之间有密不可分的关系,如图1所示。

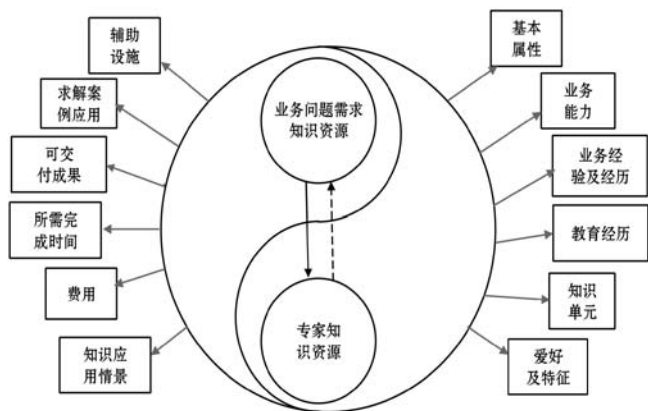


图1 业务需求与专家知识资源的关系模型图

在实际求解业务问题中的过程中,业务问题是具有多层次的结构,所以,业务求解需求就是按照层次粒度集成的集合体。根据此特性,业务求解需求是为业务问题提供服务,所以,业务需求也具有多层次的结构。围绕业务需求的涉及的知识资源角度,结合其多层次结构的特点,以企业中一系列的业务案例为例,分

别从企业业务项目层次需求维度、知识领域需求维度、产品生命周期需求维度将企业业务需求的知识资源进行分类,为专家知识资源建模做准备。

结合刘则晴等^[22]构建的业务问题模型,在此基础上,构建了业务需求的模型:

$$P_k = \{X_k, Y_k, Z_k\} \quad k=1, 2, \dots, n$$

式中: k 为企业业务求解需求编号, n 为当前企业内的业务求解需求总数。

$X_k = \{x_{ki}, i=1, 2, 3\}$, x_{ki} 表示第 k 个业务求解需求处于 i 层次的需求层次维度。

$Y_k = \{y_{kj}, j=1, 2, \dots, m\}$, y_{kj} 表示第 k 个业务求解需求的知识领域维度总数。

$Z_k = \{z_{kt}, t=1, 2, \dots, n\}$, z_{kt} 表示第 k 个业务求解需求的产品生命周期维度个数。

由于业务需求所处的项目层次维度不同,其求解的过程中所需要的知识资源也是存在差异性的,因而需求的专家知识资源也存在很大的差异性。

尽管项目层次不同,但专家知识资源每个维度的知识领域却存在着相似性。因此,该专家虽然处于不同项目、任务、活动层面中,但是都需要找到能够与专家知识资源属性一致的知识结构,而该结构的构成则可以定义为专家的知识单元维度。

在项目层次维度所处的任务层次,以及解决何种业务问题是根据专家的业务经验及经历。因而,确立了该专家的业务经验及经历维度和教育经历维度。

在业务问题求解的配置过程中,会出现专家之间的相互配合情况,若要业务问题能够顺利得到解决,就需要对专家的脾气秉性特征,及爱好等有所了解,因此,我们建立了专家知识资源模型中的又一个维度即爱好及特征,旨在业务问题求解的知识应用情景中为业务配置能够顺利进展推进,并为专家在解决问题时创造舒适环境,从而使业务问题能够高效合理地解决。

根据产品生命周期需求维度分析,主要是考虑该产品需求分析、设计、生产、销售、售后、到回收阶段,需要专家面对该流程中业务运营产生的问题,求解业务问题的能力,从中可以得到该专家知识资源模型中的业务能力维度。

通过对业务需求进行解析,本文确立了专家知识资源的六个维度信息,分别为专家的基本属性、业务经验及经历、业务能力、教育经历、知识单元、爱好及特征。对每个维度又可以进一步细分,以此作为专家知识资源与求解业务问题需求契合度的依据,并为专家知识资源模型提供基础。

数据驱动中数据即为专家知识资源模型的维度信息,驱动专家知识模型的构建。

1.2 模型构建

结合上述的专家知识资源维度的分析,本文提出依托网络数据资源,基于规则抽取、数据检索及数据挖掘技术的专家知识资源模型,如图2所示。首先专家模型的建立需要满足业务问题求解时对专家需求,利于实现专家知识资源的配置;另外如何获取专家资源信息及建立专家资源信息库,这也是业务求解过程中的关键。

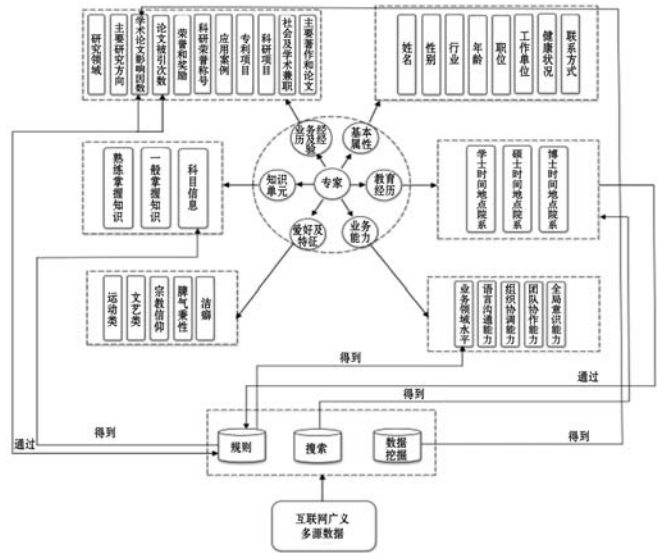


图2 数据驱动的专家知识资源模型框架

图2中对专家知识资源模型中的6个维度及其维度进行细分,并表示了每个维度信息之间的相互关系,及如何获取专家知识模型资源中相对应的数据获取方式。以教育经历为例,其细分的维度为该专家的受教育时间(when)、受教育的学校即地点(where)、受教育的学习的专业信息(major),即为“2W1M”。该维度信息的获得可以采用搜索方式,但有时该专家的某些信息又不能直接通过网络搜索获得,例如张三的毕业学校无法获取,但从网络数据中发现张三与李四是大学同学,而李四的毕业学校可获取。因而,就可以推断出张三的毕业学校。基于这样的构思,本文提出基于数据挖掘或基于规则推理得方式从网络数据中抽出不能直接获取的相关信息。

(1) 专家的基本属性 基本属性是作为业务求解过程中便于业务人员联系专家,了解专家的基本状况之用。该维度信息知识特征具有平行结构,无先后的顺序,采用表格的形式记录该基本属性的细分维度。该数据是由搜索方式得到的数据,若不能直接获得数据则由推理规则获取,如表1所示。

表 1 专家的基本属性

专家编号	基本属性							
	姓名	性别	行业	年龄	职位	工作单位	健康状况	联系方式
专家 1	× ×	× ×	× ×	× ×	× ×	× ×	× ×	× ×
专家 2	× ×	× ×	× ×	× ×	× ×	× ×	× ×	× ×
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

(2) 专家的教育经历 专家所掌握的基本知识与能力信息,是关于个体比较全局的知识与能力描述。该知识结构为“2W1M”,相互之间是关联关系。具体包括专家学习旅程中不同阶段,分别为大学、研究生、博士生等,以及在什么时间进入什么样的院校学习什么专业。如表 2 所示。

表 2 专家的教育经历

2W1M 阶段	大学	研究生	硕士	博士
时间	× ×	× ×	× ×	× ×
院校	× ×	× ×	× ×	× ×
专业	× ×	× ×	× ×	× ×

(3) 专家的业务能力 与业务求解密切相关的信息需求,了解专家能力特点是否可以业务岗位需求相匹配。其知识结构是按其掌握的熟练程度等级划分,所以以金字塔的方式描述。该维度是由业务领域水平、语言沟通能力、组织协调能力、团队协作能力、全局意思能力组成。同时,根据基于规则的推理进行该维度的等级分类。

该维度主要是以研究领域作为研究专家业务能力的的关键信息。如图 3 所示,把研究领域细化到 10 个相应的维度,分别是主要研究方向、学术论文影响因素、论文被引次数、荣誉和奖励、科研荣誉称号、应用案例、专利项目、科研项目、社会及学术兼职、主要著作和论文。根据上述 10 个细分维度,将专家业务能力更加具体展示给企业业务需求人员,作为企业业务应用专家知识资源提供参考。

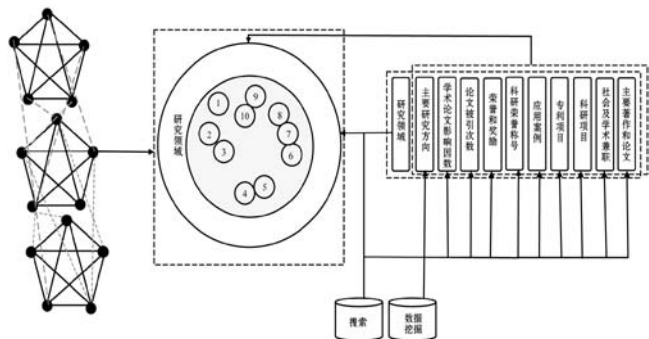


图 3 专家的业务能力图

(4) 专家的业务经历及经验 专家的业务经历和经验的描述,进一步展示专家能力范围,而且此维度也最能反映专家过往与当前的能力特长,是进行专家资源配置的重要理论依据。图 4 为该专家的业务经历及经验知识模型图。

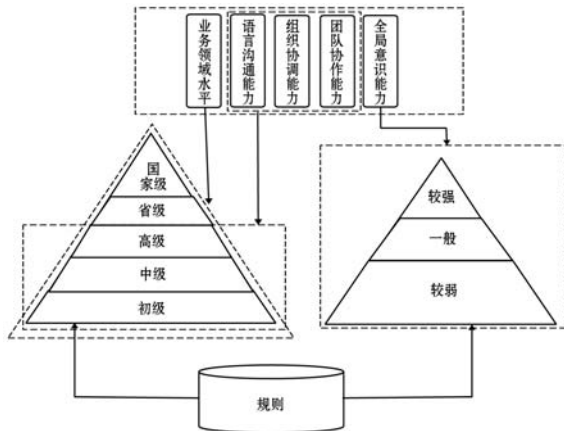


图 4 专家的业务经历及经验模型图

每个专家在人生的不同阶段都会遇到不同的专家,专家们相互交流,共同工作,知识会相互组合,并且创造形成新的知识。从专家自身的阅历和经验,形成知识体系,专家知识是具有包含的关系等,从知识的宽度领域逐渐向深度迈进,形成专家自身的知识体系结构,为业务问题求解效率带来提升和改进。

(5) 专家的知识单元 知识单元是对专家所在知识领域中具体的知识结构的最直接阐述。结合面向业务求解的知识资源配置的需要,对其知识结构进行细致的分类。如图 5 所示,其相互之间的关系是包含的关系,对于知识体系进行逐级分层,从而在业务问题求解时能够及时找到需要的专家,并将专家的知识单元按照对知识掌握的熟练程度进行划分。

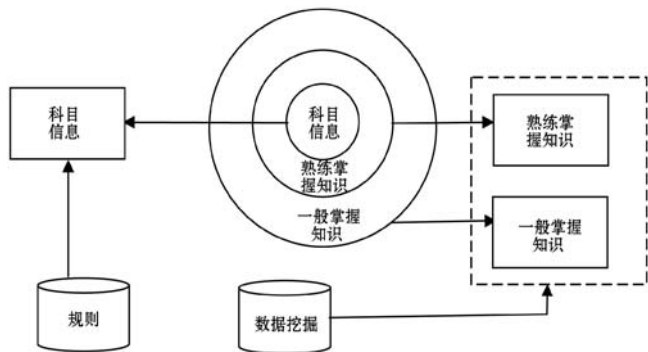


图 5 专家的知识单元

知识单元主要是按照对于知识的掌握的熟悉度,从一般到熟练,挖掘到专家构建和掌握其知识体系的方式是源自于其教育生涯中科目信息,每个专家在知识获取上都是以科目信息的传授和深入挖掘,来构建其完整的知识体系。因此,以科目信息建立规则,寻找相似知识单元结构的专家,构建专家知识资源模型。

(6) 专家的兴趣爱好及特征 该维度的模型信息是为业务求解中团队成员间可能的默契和配合程度而考虑的,在知识资源配置的协调决策中发挥辅助作用。如图 6 所示,该维度信息主要是按照树状结构进行分类,找到专家的兴趣点。其中主要包括专家喜好是运动类、文艺类,及专家的特征是有无宗教信仰、有无洁癖等特殊习惯。

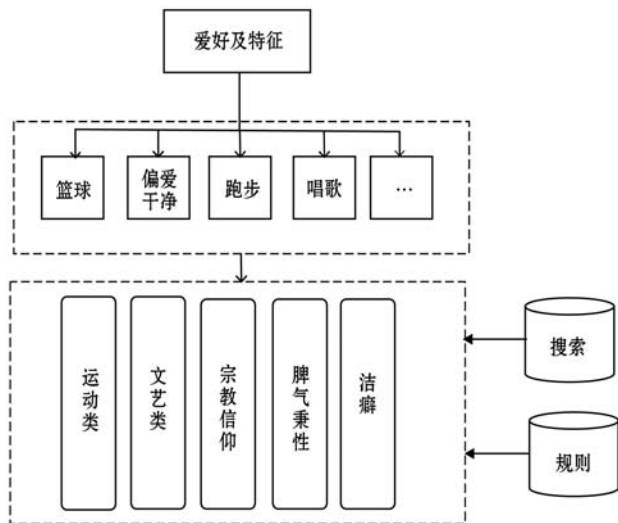


图 6 专家的爱好及特征图

2 建模方法研究

如图 2 所示,数据获取的方式有一种或者两种方式结合甚至更多,本文结合专家知识资源模型构建的数据需要,总结出 3 种数据获取的方式,分别为基于规则的推理、基于数据检索、基于数据挖掘的获取方式。

2.1 基于规则的推理获取方法

由于数据的广泛性,图 2 中各维度模型中的信息很难从数据源中直接获取,因而本文定义数据分析挖掘的映射规则,建立模型映射的规则库。所谓的规则就是从数据源信息推导出模型信息的推理原则,如图 7 所示。

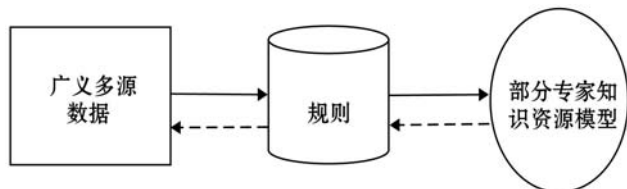


图 7 基于规则的获取流程图

广义多源数据是指互联网中开放的数据源;获取规则,主要是基于推理规则的方式获取专家模型,根据业务需求的知识资源,利用推理规则填充专家知识模型某些维度信息的数据。例如:A 专家 1994 年本科毕业于清华大学的机械制造及其自动化专业,那么从这条

专家的个人信息中,可以定义规则“该专家应该基本具有机械制造及其自动化领域的基本专家知识与技能,并基本具有 1994 年清华大学所开设专业知识中的知识单元结构”。基于这条规则,所需获取的进一步信息就是获取 1994 年清华大学机械制造及其自动化专业的本科课程体系,或获取与 A 专家相近就学情形的其他专家的个人信息,以此推导出 A 专家可能的知识单元维度信息构成及教育经验维度的求学信息。

通过不断积累的大量规则的建立,利用软件技术即可实现专家资源模型信息的自动获取。使专家模型的数据信息更加完整。

专家知识的推理规则可表示为:

规则 1: If “时间 = $\times \times \times$ ” and “地点 = 某高校名称” and “专业 = 教育部开设某专业名称”, then “该专家的科目信息 = 具体根据教育部制定与修订《普通高等学校教育学校本科专业目录》中的信息确定该专家科目信息的知识结构”。

规则 2: If “A 专家与 B 为同时同学校同专业”, then “该专家 A 与专家 B 具有相同的科目信息和知识结构”。

规则 3: If “申请项目团队人数 > 5 ” and “撰写论文的团队人数 > 5 ” and “专家参与的科研活动” and “该专家的存在社会及学术兼职”, then “团队能力较强”。

If “申请项目团队人数 > 5 ” and “撰写论文的团队人数 > 5 ” and “该专家的存在社会及学术兼职”, then “团队能力较强”。

If “申请项目团队人数 > 5 ” and “撰写论文的团队人数 > 5 ” and “专家参与的科研活动”, then “团队能力较强”。

If “申请项目团队人数 > 5 ” and “专家参与的科研活动” and “该专家的存在社会及学术兼职”, then “团队能力较强”。

If “撰写论文的团队人数 > 5 ” and “专家参与的科研活动” and “该专家的存在社会及学术兼职”, then “团队能力较强”。

If “申请项目团队人数 > 5 ” and “撰写论文的团队人数 > 5 ”, then “团队能力一般”。

If “申请项目团队人数 > 5 ” and “专家参与的科研活动”, then “团队能力一般”。

If “申请项目团队人数 > 5 ” and “该专家的存在社会及学术兼职”, then “团队能力一般”。

If “撰写论文的团队人数 > 5 ” and “专家参与的科研活动”, then “团队能力一般”。

If “撰写论文的团队人数 > 5 ” and “该专家的存在社会及学术兼职”, then “团队能力一般”。

If“专家参与的科研活动”and“该专家的存在社会及学术兼职”,then“团队能力一般”。

If“申请项目团队人数 > 5”,then“团队能力较弱”。

If“撰写论文的团队人数 > 5”,then“团队能力较弱”。

If“专家参与的科研活动”,then“团队能力较弱”。

If“该专家的存在社会及学术兼职”,then“团队能力较弱”。

满足以上三条及其以上的内容,该专家的业务能力较强。满足两条,该专家的团队能力一般。满足上面一条内容则该专家的团队能力较弱。

专家业务能力维度的描述,以该专家的团队能力维度进行其推理规则的详细阐述,其他细分维度信息获取规则方式与此相同,由于篇幅问题,暂不赘述。

规则4:If“描述该专家的信息出现干净,整洁”,and“出现程度副词,较,极其”,then“该专家有洁癖习惯”。

If“该专家信仰某种宗教”or“该专家信仰佛教”,then“该专家有宗教信仰”。

If“描述该专家做事严谨”and“被评为某种具有严谨工作性质的奖项,带有程度副词”or“该专家被评为某种带有感情色彩,我最喜爱,我最敬佩”,then“该专家做事严谨,平易近人”。

专家爱好及特征维度信息的描述,主要是根据描述信息出现的关键词及修饰该关键词的程度副词进行数据推理和填充。

2.2 基于数据检索的获取技术

根据网上已存在的数据,通过对其数据进行关键词检索得到的数据信息可直接填充进专家知识模型细分的维度信息内容,如图2所示。以专家知识资源中的教育经历为例,进行维度信息的获取。提出基于数据挖掘中广泛使用的网络爬虫技术,从互联网环境中广泛获取与专家相关的各类信息,作为数据挖掘的数据源。

某些数据经过检索得到,但是,并不能直接填充到专家知识模型中相对应的字段维度信息,需要对检索到的数据进行相关的处理,本文采用数据统计的方式,具体应用于专家的文献被引次数及专家成果影响因素和专家论文被引次数。

2.2.1 论文被引总次数及成果影响因素

论文被引总次数的确定,方便对专家进行配置时提供相应的理论参考。该方法的计算方式,主要是按照参考文献中作者的排序作为其贡献影响的依据。若

为第一作者,则其贡献量最大,则按引用次数与因子1相乘,若作者是第二位及其之后出现,将引用次数按照总人数平分,具体计算如下所示:

$$F = (a_1 + a_2 + \dots + a_n) \times 1 + (b_1 + b_2 + \dots + b_n) \times \frac{1}{2} + (c_1 + c_2 + \dots + c_n) \times \frac{1}{3} + (d_1 + d_2 + \dots + d_n) \times \frac{1}{4} + (e_1 + e_2 + \dots + e_n) \times \frac{1}{5} \quad (1)$$

式中: a_1, a_2, \dots, a_n 指专家为第一位作者的引用次数; b_1, b_2, \dots, b_n 指专家作为第二位及以后作者的引用次数,且参与学术论文创作的作者人数是2个人; c_1, c_2, \dots, c_n 指专家作为第二位及以后作者的引用次数,且参与学术论文创作的作者人数是3个人; d_1, d_2, \dots, d_n 指专家作为第二位及以后作者的引用次数,且参与学术论文创作的作者人数是4个人; e_1, e_2, \dots, e_n 指专家作为第二位及以后作者的引用次数,且参与学术论文创作的作者人数是5个人,由于一般学术论文规定的最多人数为5个人,所以,本文将专家论文被引用次数按最多人数为5个人进行其引用次数总和计算。

2.2.2 成果影响因素计算

成果影响因素,文中的成果影响因子主要是根据文章被引次数,完成其影响成果的确定,之所以要计算成果因素,是为专家在配置知识资源时提供相应的数据参考,决定哪位专家在同一领域中更适合胜任该工作。本文采用文献[23]介绍的I3(Integrated Impact Indicator)^[24],I3使用的是无参统计的方式评价偏态分布的期刊被引频次,并且利用标准差来检验评价结果,对于期刊评价法和科研人员评价比期刊因子更加合理。本文采用科研指标I3,计算公式如下:

$$Ia = \sum ix_i \cdot f(x_i) \quad (2)$$

式中: x_i 表示第*i*等级的权值, $f(x_i)$ 表示该权值为 x_i 的论文出现的数量。基于中科院对于期刊的划分^[23]并结合PR6^[24],对该权值与等级的划分范围进行调整,并根据专家学术论文的等级进行了重新界定。

2.3 基于数据挖掘的获取技术

如图1所示,专家知识资源模型框架充分考虑了企业业务问题求解时的知识需求,而对于模型各维度实例信息的获取,本文采用互联网中的开源数据挖掘技术来实现。基本流程如图8所示。

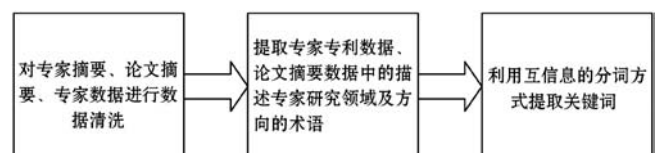


图8 基于数据挖掘的流程图

该流程主要以对专家摘要、论文摘要、专家数据进行清洗、提取专家专利数据、论文摘要数据中的描述专家研究领域及方向的术语、利用互信息的分词方式提取关键词。

对专家摘要、论文摘要、专家数据进行数据清洗,利用了数据清洗的理念,去除数据的杂质。

提取专家专利数据、论文摘要数据中的描述专家研究领域及方向的术语,采用对数据进行分词处理。

利用互信息的分词方式提取关键词:本文采用一种无监督学习的专业领域分词歧义方法^[25],采用互信息的方式,因为其分词效果较好,且具有稳定性。该方法采用定量估计两个汉字间的结合力方式,从而确定两个字之间的联系强度,以此作为依据判断两个字是否需要分词。但由于某些领域的专家的研究方向并不仅限于两个汉字之间,例如某专家擅长“知识管理”,如果只是两个词之间的紧密度的话,分词就为“知识/管理”与研究领域的专家知识描述相违背。因而,本文采用是多字词的互信息。即一个词由多个字组成时,先计算专业研究领域的两两临近字的互信息,再取平均值。算法如下:

$$mi(x_{1...n}) = \frac{1}{n-1} \sum_{i=1}^{n-1} mi(x_i, x_{i+1}) \quad (3)$$

$mi(x_i, x_{i+1})$ 指两两相近字的互信息,其计算方式如下:

$$mi(x_i, x_{i+1}) = \log_2 \frac{p(x_i, x_{i+1})}{p(x_i)p(x_{i+1})} \quad (4)$$

式中: x_i, x_{i+1} 指两个相近字。用 mi 表示两个词的相近度,以此作为研究分词的理论依据,确立得到的分词信息能够描述专家的研究领域的知识资源信息,更好地解释专家的研究领域内容。同时进行关键词提取及词频统计,为基于推理规则的方式获取模型数据做准备。

通过编写网络爬虫程序爬取专家知识资源信息,构建专家知识模型。根据上述三种方式,节省数据清洗和数据处理步骤,利用爬虫直接获取该专家模型需求的信息源。

3 原型系统开发

为了能够方便开发者进行开发,本节设计并开发该数据驱动的基于规则映射形成的专家知识资源原型系统,为专家知识资源配置服务做准备。采用 PHP + MySQL 的模式进行 Web 开发,基于 B/S 框架。数据库采用标准的 SQL 语言,由于 MySQL 是具有开放源代码的特点,因此,可以根据用户个人需要需求的要求进行修改。PHP 是一种嵌入到 HTML 中的脚本语言,绝大

多部分源于 C 语言,也有少数承自 Java 和 Perl,并由服务器进行解释的脚本语言。本系统使用的软件开发环境为:Windows10,MySQL Sever11.0.10。(数据库),PHP-STUDY20180211(调试环境),PhpStorm(编程软件)。图9为该系统的框架图。

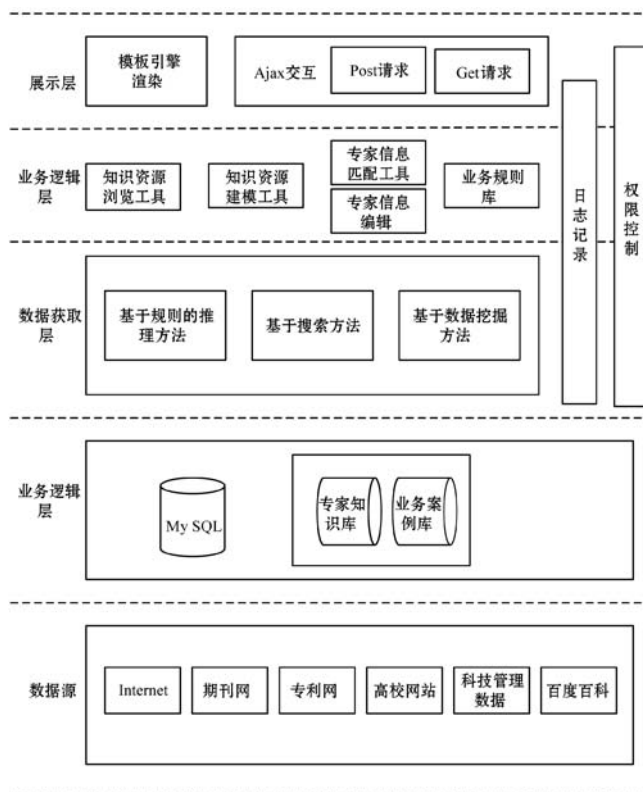


图9 系统架构图设计

根据图2和图9的理论框架设计构图的模型,构建专家知识模型系统如图10所示。



图10 数据驱动的专家知识资源模型系统

4 应用实例

本模型以实验室为研究对象,进行了该系统的测试和试验,验证了该方案的可行性。具体操作是通过输入某位专家的名字得到其相应的数据信息,显示了专家信息中部分维度中的数据信息,分别是业务经验

及经历维度,进行实例展示。同时,在科目信息旁边有一个红色惊叹号的标志,可以看到该数据的获取方法,如图11所示。



图11 数据驱动的专家知识资源部分模型系统

5 结 语

数据驱动的基于规则的专家知识资源建模及获取技术,能通过数据本身相关的规则进行推理,实现智能化获取数据的方式,减少人工编辑工作量。并且,数据的多源性使得获取知识多样性,满足知识的个性化要求,并快速地为企业提供合理的配置专家作准备工作,在业务问题求解方面提高效率,将数据可视化,为业务人员在求解过程中提供参考。

参 考 文 献

- [1] Compton P, Jansen R. A philosophical basis for knowledge acquisition[J]. Knowledge Acquisition, 1990, 2(3):241-258.
- [2] Poulymenakou A, Cornford T, Whitely E A. Pragmatic considerations for effective knowledge acquisition: the case of business expert systems[C]//Proceedings of the 1990 ACM SIGBDP conference on Trends and directions in expert systems. ACM, 1990:237-246.
- [3] Poulymenakou A, Cornford T, Whitely E A. Knowledge acquisition to facilitate organisational problem solving[J]. Expert systems with applications, 1990, 129(8):181-188.
- [4] Agarwal R, Tanniru M R. Knowledge acquisition using structured interviewing: an empirical investigation[J]. Journal of Management Information Systems, 1990, 7(1):123-140.
- [5] Wagner W P, Najdawi M K, Chung Q B. Selection of knowledge acquisition techniques based upon the problem domain characteristics of production and operations management expert systems[J]. Expert Systems, 2001, 18(2):76-87.
- [6] Valiris G, Laios L. Design and validation of knowledge acquisition tools in a business domain[J]. Behaviour & Information Technology, 1995, 14(2):121-131.
- [7] 郭健美,王英林,张申生. 面向业务问题求解的知识供应过程模型[J]. 计算机应用与软件, 2009, 26(1):3-5, 12.
- [8] 王德川,战洪飞,余军合. 面向业务问题求解的知识资源特定领域建模方法[J]. 计算机应用, 2016, 36(3):37-842.
- [9] 姚平,梁静国,刘丙泉. 基于模糊聚类与粗糙集的决策知识获取方法研究[J]. 情报杂志, 2007, 26(11):46-48.
- [10] 罗琳,顾新. 智慧数据驱动的产学研协同创新知识管网研究[J]. 软科学, 2017, 6(31):15-18.
- [11] 刘骊,王若梅,罗笑楠,等. 数据驱动的三维服装快速建模[J]. 软件学报, 2016, 27(1):2574-2586.
- [12] 徐建国,李孟军,姜江,等. 数据驱动的技术创新网络模体分析[J]. 系统工程与电子技术, 2017, 39(5):1072-1077.
- [13] 杨青,戴高宁,王丽珍. 大数据:数据驱动下的工程项目管理新视角[J]. 系统工程理论与实践, 2017, 37(3):710-719.
- [14] 程进,王坚. 数据驱动的流程制造工艺参数匹配方法[J]. 计算机集成制造系统, 2017, 23(11):2361-2370.
- [15] 徐晓滨,郑进,徐冬玲,等. 基于证据推理规则的信息融合故障诊断方法[J]. 控制理论与应用, 2015, 9(32):1170-1181.
- [16] 苏立,南海鹏,余向阳,等. 关联规则分类的数据流挖掘方法在水电机组故障诊断的应用[J]. 西安理工大学学报, 2012, 28(1):106-110.
- [17] 魏海洋,周纪超,穆俊斌,等. CRH3型动车组故障诊断代码编码规则研究[J]. 大连交通大报, 2014, 8(35):102-105.
- [18] 谢龙君,李黎,程勇,等. 融合集对分析和关联规则的变压器故障诊断方法[J]. 中国电机工程学报, 2015, 35(2):277-286.
- [19] 罗俊海,王章静. 多源数据融合和传感器管理[M]. 北京:清华大学出版社, 2015.
- [20] 宫欣伟,余国林,孙勇,等. 基于数据驱动的全流程模锻生产过程专家系统[J]. 锻压技术, 2017(4):27-32.
- [21] 朱杰,游雄,夏青. 基于任务驱动的战场环境分析数据映射模型设计与实现[J]. 地理信息世界, 2017, 24(4):80-85.
- [22] 刘则晴,战洪飞,余军和. 企业业务问题解决方案的建模方法研究[J]. 科技与管理, 2014, 7(16):72-76.
- [23] 陈福佑,杨立英. 新科研影响力评价指标分析[J]. 情报杂志, 2014, 7(14):81-85.
- [24] Wagner C S, Leydesdorff L. An Integrated Impact Indicator (I3): A New Definition of "Impact" with Policy Relevance[J]. Research Evaluation, 2012, 21(3):183-188.
- [25] 修池,宋柔. 适用于不同领域的中文分词方法研究与实现[D]. 北京:北京工业大学, 2013:114-115.