

doi: 10.12052/gdutxb.180135

面向问题智能处理的基元-因素空间模型研究

李兴森¹, 许立波², 刘海涛³

(1. 广东工业大学 可拓学与创新方法研究所, 广东 广州 510006; 2. 浙江大学宁波理工学院 计算机与数据工程学院 中澳智能计算与数据管理研究中心, 浙江 宁波 315100; 3. 辽宁工程技术大学 理学院, 辽宁 阜新 125105)

摘要: 互联网已成为获取信息的重要来源, 但仅依靠信息搜索引擎等现有技术仍难以从海量信息中智能化挖掘知识为解决问题服务, 面向问题处理的互联网知识搜索的模型与算法等问题一直未得到有效解决. 知识搜索引擎要实现智能化解决问题的服务, 关键是必须研究面向问题处理的互联网信息的智能重构模型. 基元-因素空间模型以可拓学基元理论、因素空间理论和智能知识管理理论为基础, 交叉研究从互联网海量信息资源中抽取对象、属性和量值, 重组互联网信息, 自动构建领域基元的方法及从领域信息中挖掘特定问题的因素空间以获取领域知识的智能算法. 在动态基元信息、专家经验知识与因素空间融合的模式与算法基础上, 开发模拟仿真软件以验证动态知识基元及因素库协同进行互联网信息重组的理论与方法, 为利用互联网生成解决问题的创新策略和研发新一代面向问题处理的知识搜索引擎提供理论基础和新的方法支撑.

关键词: 可拓学; 因素空间; 信息结构; 知识挖掘; 基元
中图分类号: TP315, F49 **文献标志码:** A **文章编号:** 1007-7162(2019)01-0001-09

A Research on Problem Oriented Intelligent Processing Model by Basic-Element and Factor Space

Li Xing-sen¹, Xu Li-bo², Liu Hai-tao³

(1. Institute of Extenics and Innovation Methods, Guangdong University of Technology, Guangzhou 510006, China; 2. Research Center of Intelligent Computing and Data Management, Zhejiang University Ningbo Institute of Technology, Ningbo, 315100, China; 3. College of Science, Liaoning Technical University, Fuxin 123000, China)

Abstract: The Internet has become an important way of obtaining information, but it is still difficult to provide aids to problem solving. Problem solving oriented knowledge searching in the Internet has not been solved thoroughly. The key factor is the reconstruction of web information. The combination of network and intelligence will bring the knowledge-based application of web information into a new stage. A cross-study is proposed on the novel hybrid method of extracting the objects, attributes, and values from the mass information of internet with simulation and automatically building the field basic elements on macro information based on Extenics, factor space theory and intelligent knowledge. In addition, a lot of future research directions are put forward, such as on intelligent algorithm which collects correlative factors of special problem from the field of information for acquiring domain knowledge. The research may provide theoretical basis and new methods for using the internet to generate the innovative strategy of problem solving and examine a new generation of problem-solving oriented knowledge search engine.

Key words: Extenics; factor space; information structure; knowledge mining; basic-element

互联网搜索引擎已成为人类获取信息的重要途径. 现有搜索引擎的核心技术是采集主题相关的 web 信息, 通过关键词匹配从海量网页中快速、被动

地提供信息^[1], 并以各类算法排序展示给用户. 但现有搜索引擎提供的是堆砌的信息, 用户难以从数以百万计的网页中(以“客户流失”为关键词, 百度中查

收稿日期: 2018-10-12

基金项目: 国家自然科学基金资助项目(71271191); 浙江省自然科学基金资助项目(LY18F020001, LY16G010010); 教育部人文社会科学研究项目(18YJAZH049)

作者简介: 李兴森(1968-), 男, 教授, 博士, 主要研究方向为可拓学与智能创新.

出319万条信息)提取有价值的信息进而获取知识,更难以整理和综合知识智能解决问题.如前微软亚洲研究院常务副院长马维英所言:“目前的搜索引擎,提供的主要是链接而不是信息,其实并不能真正理解用户的搜索需求,智能搜索是未来的大势所趋.”以智能化为用户提供问题解决方案将是下一代搜索引擎的主要特征.然而,目前互联网信息的智能化应用还面临如下挑战.

(1) 现有技术难以解决多源数据、信息、知识的智能融合问题.网页文本、图片、声音、视频等多种类型的异构数据堆砌成动态、粗糙的海量信息(截至2016年底,中国网页数量为2 360亿个),且以11.2%以上的速度逐年加速增长(中国互联网络信息中心第39次《中国互联网络发展状况统计报告》).现有的局部数据挖掘手段仅能产生碎片化知识,而信息整合面临事物的分类不具唯一性,海量信息和知识在不同领域有不同的概念连接结构及知识图谱,不同的目标需要不同的分类和图谱方式等.如何自动融合、匹配这些数据与信息并为后续大规模、高效地挖掘应用搭建新型的知识库,是一件极具挑战性的任务.

(2) 现有技术难以解决web信息的知识智能挖掘与策略生成问题.互联网的信息处于动态增长和不断更新的过程中,如何能够在保证获取知识的质量不变的前提下,增量地实时挖掘不断涌现的新数据是知识智能化挖掘的难题^[2].此外,现有基于静态数据挖掘的算法仅能从网页中获取多源碎片化知识,难以满足用户使用搜索引擎的真正需求.例如,查询房源信息背后的真正目标可能是为解决资金不足又必须买高价学区房的矛盾问题,获取互联网上的房源信息只是解决问题的前期准备.

要解决上述2个主要问题,使新一代互联网实现智能化解决问题,其关键是必须研究面向问题处理的互联网信息的智能重构模型,为智能搜索与知识挖掘准备信息条件.从动态快速变化的海量web信息中,面向特定领域的问题,对多源异构信息进行智能化的过滤、提取、融合、挖掘及重组等,建立web网页信息层上一层的结构化领域信息知识库(类似整合各管理信息系统构建数据仓库、数据集市).并将这作为问题智能处理的信息输入,为web大数据环境下矛盾问题智能处理等基础性科学问题研究提供结构化信息基础.面向问题处理的互联网信息的智能重构模型研究极为迫切.

1 因素空间理论与可拓学理论在问题智能化处理方面的作用

1.1 因素空间理论

(1) 因素.1982年产生了3个直接以认知为描述对象的数学理论:形式概念分析、粗糙集和因素空间^[3].因素是指确定事物产生的原因的要素.如降雨量是粮食丰收的一个因素,遗传基因是健康的一个因素.随机性是由于掌控因素的不充分而导致因果律的破缺,概率是不充分条件下的广义因果律,问题处理要从不确定性和随机性向必然性转化.同样,模糊性是描述因素的不充分而导致的排中律的破缺,因素的增加会降低认识的模糊性.互联网的海量信息对特定问题是模糊的,有可能通过因素的系统挖掘逐步聚类而清晰化^[4].

(2) 因素空间.投掷一枚硬币落地后会得到硬币状态的基本空间 $\Omega=\{\text{正}, \text{反}\}$,把硬币图案形状、硬币材质均匀度、手投掷的动作、落地的地面条件和其他环境等影响硬币正反面朝向的因素映射到不同的坐标轴上进行综合,将这些轴交叉起来,形成一个以因素为轴的坐标空间,叫做因素空间.每个因素对应一个状态变化的维度,所有因素张成一个高维空间 $X(F)$, F 是因素的集合, F 的充分性便确定了从 $X(F)$ 到 Ω 的一个映射,使偶然现象得到必然性的结果.任何事物都可以被相应的因素空间所描述.因素空间被认为是事物描述的普适性框架^[5].

(3) 因素库.因素空间存储形成的数据库叫做因素库.因素库具有4个特色:①明确了数据挖掘的目标是用数据样本逼近背景分布.背景分布决定概念生成、因果归纳和逻辑推理以及知识结构.②明确了数据维度化大为小的关键.在很宽泛的条件下,背景分布具有凸性,凸集可以由其顶点决定. R 的顶点所构成的集合叫做背景基 B ,它的数量比 R 要少得多.从 R 到 B ,是信息压缩的关键.相比网页信息,因素数据库保持着非大的数据.③对于非结构化数据,可在关系数据库的右端留下影像资料作备份,按目标需求询问这些资料的参考价值,随时调用.④背景样本是去掉对象而留下的性状分布,不存在隐私泄露的问题,样本求并可以并行计算.

以上特点显示了因素库处理互联网大数据的明显优势.因素库以影响结果状态的关键因素为单元,在互联网上收集处理数据,逐步培养出以问题背景关系为核心的知识基.通过对大数据的因素化处理,

始终保持较低的维度,在不侵犯隐私的前提下与同类知识包并行合作,以上下关系与异类包进行连接,形成融合监测、组织、管理、控制的人机认知体这一软硬件系统^[6].用因素空间研究事物状态的确定性与不确定性的转化,已经在概念生成、知识表示、模糊控制、模糊推理等领域形成理论及应用体系^[7].

因素神经网络^[8]、决策科学^[9-11]、分类^[12]等应用成果表明,因素空间作为海量信息低维度描述的数学框架,不但简明清晰地叙述信息与智能科学的一些基本问题,而且算法简单,涉及的N-hard问题在因素空间中都可以化为复杂度不高于 $O(m^2n)$ 的算法(m, n 分别表示对象和因素个数).另外,因素库能处理的数据除了结构化数据以外还有非结构化数据和异构数据.近年来,因素空间在数据科学领域显示出新的生命力,如为粗糙集提供了样本分析的坐标系统和母体理论等^[5].

根据以上分析,因素空间理论不但把非结构化大数据转化为小数据,而且将为信息和智能科学提供统一的数学语言和普适的描述框架.从web海量信息中抽取解决领域问题的要素子集是策略生成的必要准备.因素空间理论将为基元-因素空间模型研究提供必要的数学方法.

1.2 可拓学理论与方法回顾

1.2.1 可拓学理论概述

可拓学是一门以矛盾问题为研究对象的原创性横断学科,它用基元等形式化模型研究事物拓展的可能性以及矛盾问题处理的规律与方法,根本目标在于借助计算机等智能化解决矛盾问题^[13-16].可拓学是哲学、数学和工程学交叉而成的学科,对数学和逻辑做了较大拓展,构建了解决矛盾问题的方法论体系^[17],成为连接自然科学和社会科学的桥梁^[18].可拓学理论和方法已应用到数据挖掘、知识管理、策略生成及多个人工智能领域^[19-20].罗马尼亚科学院及美国学者将可拓学拓展到多维空间^[21-22]并用于机器人控制^[23],部分成果获得日内瓦国际发明专利金奖.

1.2.2 可拓学与知识智能化管理

可拓数据挖掘理论与方法立足于挖掘“变换”的知识^[24],近年来提出了可拓转化知识的挖掘方法^[25],基于基元理论的内容推荐算法^[26]等.李兴森等^[27]聚焦数据挖掘与知识管理的交叉融合,提出了智能知识管理理论框架以及知识培育孵化的方法,并与可拓集合结合提出可拓创新模型^[28]和大数据环境下的协同创新模型^[29].

1.2.3 基元理论与矛盾问题处理

通过基元理论建立可拓模型,利用基元库与可拓推理和可拓变换、传导变换等方法结合,研究创意生成和解决矛盾问题的理论与方法^[30-33].借助Agent的智能引导,建立了基于可拓学和HowNet的策略生成系统的基本流程和系统框架结构,并在矛盾问题信息的形式化基元模型及主要矛盾识别方法、基元问题的相关度计算、自适应软件形式化方法、解析用户自然语言需求语句并自动建立可拓模型的方法^[20]、基于知网的可拓领域信息元库的构建方法^[34]等方面取得系列进展.

另外,可拓学在互联网信息处理^[35]和可拓距与关联函数^[36]为多维信息融合及矛盾问题处理的研究打下了很好的基元建模和因素完备性、有效性科学评价的方法基础.

2 问题智能化处理的现状与趋势

国内外学者针对基于互联网信息的问题智能处理这一方向,主要有如下3类研究方案.

2.1 语义元数据分析与搜索引擎优化

主要研究集中在如下3个方面.

(1) 元数据分析与搜索算法优化.该领域的研究主要集中在检索效率、检索结果的排序、扩大信息搜集的范围等^[37-39].对语义元数据的分析是提高web服务和搜索引擎智能性的重要方面.如利用语义元数据服务弥补信息搜索的质量差距,构建理解问题的语境等^[40-41],基于Agent的智能元搜索引擎等相关算法研究对知识的搜索有一定的参考价值,但需要提升推理和信息知识智能衔接等的智能性.

(2) 研究智能性更强的新搜索引擎.如提高对用户检索提问的理解,研究语义搜索引擎^[42]、模糊语义搜索引擎^[43-44]和可拓展标记语言(XML)内容驱动的搜索引擎^[45].很多研究者开发了本体搜索引擎和原型系统,提供本体搜索与排序服务^[46],或结合情景和本体知识的搜索^[47-49]等,但搜索内容仍然是信息为主,很少涉及深层知识的搜索.

(3) 开发新一代面向特定领域应用的专业化垂直搜索引擎^[50].如商业搜索引擎^[51],基于本体的云服务检索系统^[52],搜索云平台上的各类云计算服务基元-因素空间模型.

上述大部分研究聚焦在搜索引擎技术层面,如关键词搜索算法、信息的呈现方式等^[53],对用户的特定情景和动机及如何智能化解决问题等缺乏系统全

面的研究。

2.2 知识搜索引擎

一些学者已经认识到,互联网知识的智能化应用是一大趋势^[54],未来应研究智能性更高的知识搜索引擎,挑战在于如何从粗糙的海量的web网页中自动抽取结构化的知识。

较早着手尝试基于自然语言理解的知识搜索系统是WolframAlpha。由沃尔夫勒姆研究公司开发的新一代的搜索引擎网站,2013年5月上线,能根据用户的问题进行后台运算后直接给出答案,而不是返回一大堆信息及网页链接。它从公开的和获得授权的信息资源中,重新组织信息建立一个庞大的新数据库,再利用高级的自然语言算法进行处理后,构造出一个计算知识引擎。例如,在被问到“婶婶的女儿的儿子”之类的问题时,WolframAlpha能告诉你正确的称谓、亲缘关系度,还绘制一张类似家谱的系列图表。但其智能性还不够高,目前比较好的应用是一些常识的运算,缺乏行业知识库支撑,难以提供全面的行业知识来实际问题。

2.3 互联网架构的升级换代

认识到现有互联网信息爆炸难以高效利用等问题,国内外一些学者提出了信息中心网络(ICN)的新构想,为构建以内容为中心的下一代网络提供方法^[55],让现在以IP为核心的架构能在将来演进成以内容为核心的架构。但架构的升级换代是一件复杂的事情,需要达成共识并完成一系列准备工作。而在升级换代过程中,现有互联网的信息仍在持续增加中。

认识问题和系统性解决问题需要从子集到全集,逐步实现对问题要素的智能识别和处理即策略生成。而问题求解和模式识别等都依赖信息加工的特殊形式。缺乏突破性的新的信息组织结构支撑新一代搜索引擎的研发,在以文本挖掘和智能知识发现为主要技术的策略生成等方面,系统性融合的信息基础架构仍相对缺乏。近年来,基于互联网解决问题的研究逐渐增加,但西方学者擅长以微观思维研究特定边界条件下数量化问题的优化和求解^[56-57],缺乏基于web信息融合的开放性创新策略生成的研究。

互联网信息爆炸的背景下,如何借助大数据处理的最新理论和方法,研究互联网信息在智能化生成解决问题的策略方面的应用基础科学问题,特别是非结构化信息面向问题的重新组织问题。大数据使web数据的常态从存储变为流动。新阶段信息革命的核心是网络和智能的结合。而中国学者创立的因

素空间理论和可拓学理论将为面向新一代策略搜索引擎构建的信息重构提供理论和方法基础。

利用可拓学的基元理论,将网络信息和知识进行属性抽取和量值融合,进而构建因素空间,设计面向行业问题处理的策略搜索引擎,其搜索的主要内容是互联网上解决矛盾问题相关的深层知识(在web信息抽取基础上的二次抽取),这将是新一代搜索引擎技术研发的重要方向之一。动态基元库与因素库融合的智能构建方法,是为未来的知识搜索和策略搜索提供信息基础的关键步骤之一。

3 基元-因素空间模型构建

3.1 模型框架及主要内容

面向特定领域,将互联网信息分为物、事、关系三类,分别抽取属性和量值构建基元库。基于内部数据库以数据挖掘方法挖掘关键因素,构建因素空间,因素空间和基元库融合生成面向特定问题的因素库,因素库以树结构组织,多个因素库构成因素仓库,成为支撑未来知识搜索引擎和策略搜索引擎构建的理论基础。

模型针对面向问题处理的策略智能化生成的目标,研究多类、动态web信息的因素化重构方法,为新一代知识搜索提供理论基础。模型的主要模块之间的逻辑关系如图1所示。各模块的内容说明如下。

(1) 从web信息中提取基元对象、属性和量值的算法。该模块研究领域本体与基元的互补性及属性词典的生成方法,研究从web提取的关键词中识别对象和属性关联的算法,设计领域信息与知识的基元表达结构与云存储方案;研究属性聚类与量值的区间化集合智能生成算法,提出自动构建行业领域基元的理论方法及动态更新机制,该内容提供策略生成的结构化宏观基础信息。

(2) 面向特定问题的因素挖掘及因素空间构建方法。研究量值分析获取属性权重的算法、面向特定领域挖掘关键因素的组合法及因素挖掘效果的评价方法;研究结合管理信息系统构建领域因素空间的理论与算法;研究面向问题处理的因素运算、因素组态的概率分布以及多个因素空间的交叉维护方法,实现因素词典的自动生成及更新等。

(3) 动态基元信息、专家经验知识与因素空间融合的模式与算法。研究面向特定问题的动态基元信息、专家经验知识与因素空间的融合算法、管理模式和统一表达,设计云存储方案,形成基元-因素耦合

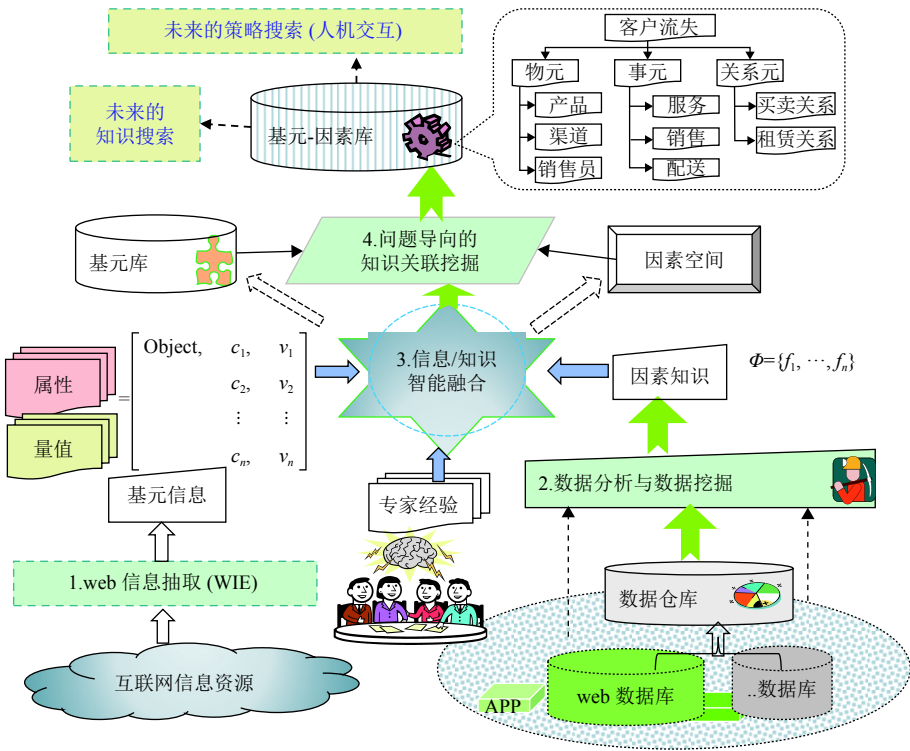


图1 模型的逻辑关系框架图
Fig.1 Logic framework of the model

结构. 研究基元-因素结构信息的聚类与关联算法, 构建因素与目标结果之间的函数方程以及可拓距为主的三元融合完备性评价函数等.

(4) 基元-因素结构在面向问题处理的动态关键知识链生成中的模拟与检验. 建立描述矛盾问题的人机交互基元模型, 优化面向矛盾问题处理的基元-因素结构信息的复杂网络. 研究问题的基元表达与策略可信度和相容度评价方法, 利用基于基元-因素结构的关键知识链构建挖掘算法, 采用因素空间建立简明统一的数学模式并编制快捷的算法. 模拟人机交互的关键知识的识别及智能补充方法, 完成信息与知识基元化融合生成解决问题策略的动态网络结构在策略搜索中试用与检验.

基于信息智能采集与云存储技术, 将集成结构化数据与非结构化数据利用可拓学基元理论构建统一结构的信息基元, 实现异质数据的离散化及增量集成. 信息基元的基本结构以三元组(对象, 属性, 量值)表示, 多个属性形成矩阵结构:

$$B = (O, C, V) = \begin{bmatrix} \text{Object}, & c_1, & v_1 \\ & c_2, & v_2 \\ & \vdots & \vdots \\ & c_n, & v_n \end{bmatrix}.$$

其中, O (Object) 表示某对象(物、动作或关系词), $c_1, c_2, \dots, c_n$ 表示对象 O 的 n 个特征, $v_1, v_2, \dots, v_n$ 表示对

象 O 关于上述特征的相应量值.

多个信息基元通过特征或量值相互联系, 形成复杂网络. 在复杂信息基元网络基础上, 研究问题处理的可拓变换知识的挖掘算法. 例如从客户注册、日常业务使用等数据中挖掘客户流失的规则, 然后通过可拓数据挖掘得到引导冻结客户(即将流失客户)向正常客户转化的规则知识. 以互联网基元库信息为基础, 针对数据的动态性、多态性和时效性, 综合运用数理统计、可拓数据挖掘和智能知识管理技术, 研究面向矛盾处理的因素知识挖掘算法.

3.2 实现方案

基于领域问题的因素空间构建因素知识图谱, 运用智能知识二阶挖掘方法, 得到高阶智能知识, 发现问题产生的根本原因和求解的方向. 智能知识是引入心理学、复杂系统、机器学习等理论, 对数据挖掘获取的知识进行存储、评价、筛选、审计, 对有价值的知识进行标记和基元化处理, 孵化出自身具有智能的知识元并进行组织重构、自我更新和智能应用等过程的管理, 使知识能识别情境, 在合适的时间自动提供给需要的人, 从而提高数据挖掘获取的知识智能性和实用性. 以规则表达的智能知识示例如下.

R =

规则知识A302,	来源,	销售数据
	类型,	决策树
	适用部门,	设计中心
	数据集开始时间	2010.8
	数据集结束时间	2016.3
	规则产生时间	2016/3/9
	审计分数	87.5
	使用次数	21
		...

将可搜索因素知识分为企业内部因素、合作商共享知识与互联网外部知识三类,统一用基元建立云知

识仓库.在此基础上,设计基于信息基元的知识爬虫算法和知识搜索算法,以多维关联函数评价矛盾相容度,根据多维分层情境模型结合语义技术研究矛盾情境的关联技术.

知识层采用可拓数据挖掘和智能知识发现方法提取规则和知识,以基元支持多源信息与知识的集成应用.通过因素知识图谱,形成多源碎片化知识的相互印证、组合和衔接,最终发现面向问题处理的针对性的知识链策略,如图2所示.

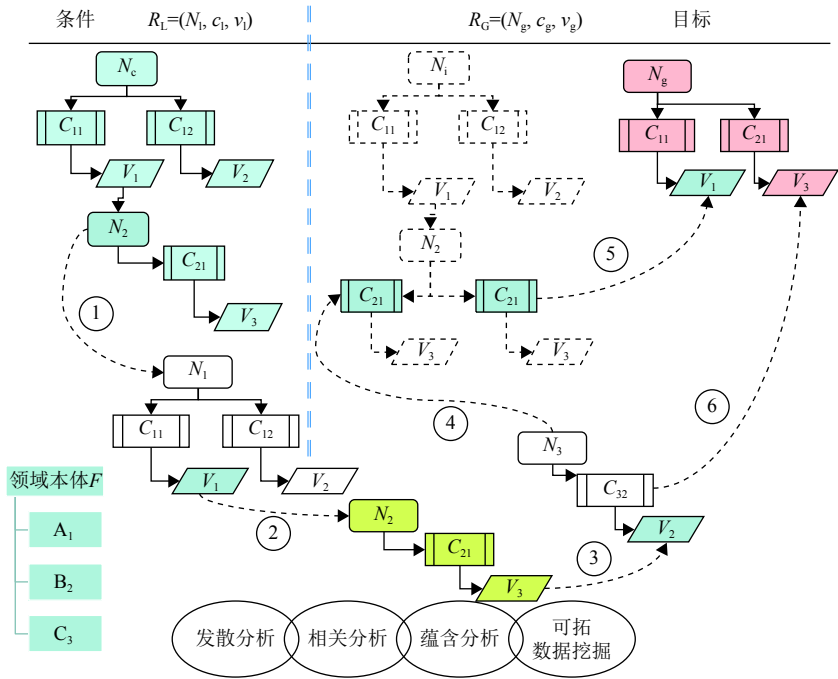


图 2 基元隐含知识链生成的路径图

Fig.2 Implicit knowledge chain generated path from basic elements

设定义在 U 上的因素集为 $F = \{f_1, \dots, f_n\}$, 即 $f_j: U \rightarrow X(f_j) (j = 1, \dots, n)$. 设 $F' = P(F) = (F'; \vee, \wedge, \neg, 0, 1)$ 是一个布尔代数, 记 $X(F) = \{X(f)\}_{(f \in F')}$, 称 $\psi = (U, X(F))$ 为 U 上的一个因素空间, 如果满足条件:

- (1) $X(0) = \{\emptyset\}$.
- (2) 对任意 $T \subseteq F$ 及 $s, t \in T$, 若 $s \wedge t = 0$ (s 与 t 叫做不可约), 则 $X(\vee\{f|f \in T\}) = \Pi_{f \in T} X(f)$ (Π 是笛卡尔乘积) 其中, $F = \{f_1, \dots, f_n\} = \{f_1 \vee \dots \vee f_n\} = 1$ 叫做全因素, 0 叫做空因素. 记号 $\vee$ 和 $\wedge$ 分别叫做因素的合成和分解运算.

对任意 $f = \{f_{(1)}, \dots, f_{(k)}\} \subseteq F$, 用同一符号来记 $f = f_{(1)} \vee \dots \vee f_{(k)}$. 因素集相当于诸因素的合成. 多个信息基元通过特征或量值相互联系, 形成复杂网络.

合成与分解运算使得因素相空间的维度随着因素繁简交变, 因素空间在数学上被定义成一个以因

素为参数的集合族, 而参数域可以按这两种运算形成偏序集或格, 甚至强化为一个布尔代数.

问题的因果归纳可化为这样的数学问题: 将所有因素分为条件与结果两类, 要通过背景分布 R 来提取从条件到结果的因果推理句, 并且要在因素约简的前提下来获取这些推理句. 一个认知单元就是对一个给定概念团粒的细化, 由一个因素空间来承担. 新的子概念又可根椐新的要求再生成一个因素空间, 形成因素空间藤 (简称因素藤). 一个实际系统可以在不同节点上生成多个因素空间, 也可被视为另一种形式的因素藤. 因素藤实现信息与知识的融合, 形成更大的认知网络.

针对利用互联网大数据处理矛盾问题的应用, 基于期望的目标状态及影响因素选定合适的树结构, 用目标驱动知识分类和知识图谱的构建, 建立以

目标为主导的信息和知识的动态联接模式. 对因素库的每个认知包都设立标志符号, 标明论域的概念名称和任务目标的相关性. 然后从目标出发逆向搜寻实现目标的充分条件因素, 递推到现有条件因素成立时得到一条解决问题的路径.

问题解决程度可用矛盾相容度评价的多维关联函数实现:

$$K_{nD}(x_1, x_2, \dots, x_n) = \frac{\rho((x_1, x_2, \dots, x_n), S_2)}{\rho((x_1, x_2, \dots, x_n), S_2) - \rho((x_1, x_2, \dots, x_n), S_1)},$$

具体含义参考文献[21]. 在此不再详述.

4 结论

从互联网信息深度利用的角度提出一种基于东方系统思维的系统组织模型与策略知识链的模型构建方法, 研究互联网海量信息背景下因素空间的智能构建, 可以丰富因素空间理论, 促进因素空间在信息科学领域的交叉应用. 基元库与因素空间的交叉研究使两大理论实现互补, 构建一种新的非结构化数据处理的基础理论.

本模型的深入研究将进一步提升互联网的问题处理资源库作用. 因素库积累到一定程度, 将有助于解读深度学习的黑箱要素, 有助于实现创新过程科学管理理论的突破. 对提升企业自主创新能力, 实现中小型企业的产业转型升级都具有重要的实践意义. 在利用大数据构建国家战略因素库和行业、企业管理因素库等方面也有广阔的应用前景. 从动态快速变化的海量web信息中, 面向特定领域的问题, 对多源异构信息进行智能化的过滤、提取、融合、挖掘及重组等, 建立web网页信息上一层的结构化领域信息知识库(类似整合各管理信息系统构建数据仓库、数据集市), 将为web大数据环境下矛盾问题智能处理等基础性科学问题提供新的视角.

深入研究的方向包括以下几个方面, 如探讨互联网信息和领域知识的基元-因素化融合的动态结构, 提出基元库与因素库互补的耦合方法; 建立领域因素空间的基元库模型; 开发原型系统, 进一步探索互联网海量信息在问题处理方面的深度利用及基础理论与方法等.

(本文部分内容曾在ITQM2017国际会议上交流)

参考文献:

[1] RAHMAN M. Search engines going beyond keyword search: a survey [J]. *International Journal of Computer Applications*, 2013, 75(17): 1-8.

[2] 张清华, 幸禹可, 周玉兰. 基于粒计算的增量式知识获取方法[J]. *电子与信息学报*, 2011, 33(2): 435-441.
ZHANG Q H, XING Y K, ZHOU Y L. The incremental knowledge acquisition algorithm based on granular computing [J]. *Journal of Electronics & Information Technology*, 2011, 33(2): 435-441.

[3] 汪培庄, SUGENO M. 因素场与模糊集的背景结构[J]. *模糊数学*, 1982(2): 45-54.
WANG P Z, SUGENO M. The background structure of factor field and fuzzy sets [J]. *Fuzzy Mathematics*, 1982(2): 45-54.

[4] 汪培庄. 因素空间与因素库[J]. *辽宁工程技术大学学报(自然科学版)*, 2013, 32(10): 1-8.
WANG P Z. Factor spaces and factor data-bases [J]. *Journal of Liaoning Technical University (Natural Science)*, 2013, 32(10): 1-8.

[5] 汪培庄. 因素空间与数据科学[J]. *辽宁工程技术大学学报(自然科学版)*, 2015, 34(2): 273-280.
WANG P Z. Factor space and data science [J]. *Journal of Liaoning Technical University (Natural Science)*, 2015, 34(2): 273-280.

[6] WANG P Z, LIU Z L, SHI Y, *et al.* Factor space, the theoretical base of data science [J]. *Annals of Data Science*, 2014, 1(2): 233-251.

[7] 汪培庄, 李洪兴. 知识表示的数学理论[M]. 天津: 天津科技出版社, 1994.

[8] 刘增良, 刘有才. 因素神经网络理论及实现策略研究[M]. 北京: 北京师范大学出版社, 1992.

[9] ZENG W Y, LI D Q, WANG P Z. Variable weight decision making and balance function analysis based on factor space [J]. *International Journal of Information Technology & Decision Making*, 2016, 15(5): 999-1014.

[10] ZHOU L, SUN K, LI H. Multifactorial decision making based on type-2 fuzzy sets and factor space approach [J]. *Journal of Intelligent & Fuzzy Systems*, 2016, 30(4): 2257-2266.

[11] 汪华东, 郭嗣琮. 基于因素空间反馈外延外包络的DFE决策[J]. *计算机工程与应用*, 2015, 51(15): 148-152.
WANG H D, GUO S Z. DFE decision-making based on outer-enveloping of feedback extension in factor spaces [J]. *Computer Engineering and Applications*, 2015, 51(15): 148-152.

[12] 崔铁军, 马云东. 因素空间的属性圆定义及其在对象分类中的应用[J]. *计算机工程与科学*, 2015, 31(11): 2169-2174.
CUI T J, MA Y D. Definition of attribute circle in factor space and its application in object classification [J]. *Computer Engineering & Science*, 2015, 31(11): 2169-2174.

[13] 蔡文. 可拓集合和不相容问题[J]. *科学探索学报*, 1983(1): 83-97.
CAI W. Extension set and non-compatible problem [J]. *Sci-*

- ence Exploration, 1983(1): 83-97.
- [14] 蔡文, 石勇. 可拓学的科学意义与未来发展[J]. [哈尔滨工业大学学报](#), 2006, 38(7): 1079-1086.
CAI W, SHI Y. Extenics: its significance in science and prospects in application [J]. [Journal of Harbin Institute of Technology](#), 2006, 38(7): 1079-1086.
 - [15] 涂序彦. 开拓新领域研究新方法[J]. 科技导报, 2014, 32(36): 117.
TU X Y. Develop new research methods in new fields [J]. Science & Technology Review, 2014, 32(36): 117.
 - [16] 钟义信. 评“可拓学”[J]. 科学通报, 2013(13): 1188-1189.
ZHONG Y X. Comments on “Extenics” [J]. Chinese Science Bulletin, 2013(13): 1188-1189.
 - [17] 杨春燕, 蔡文. 可拓学[M]. 北京: 科学出版社, 2013.
 - [18] 李幼平. 可拓学: 沟通自然科学与社会科学的桥梁[J]. 科技导报, 2014, 32(36): 1, 3.
 - [19] 陈文伟, 杨春燕, 黄金才. 可拓知识与可拓知识推理[J]. [哈尔滨工业大学学报](#), 2006, 38(7): 1094-1096.
CHEN W W, YANG C Y, HUANG J C. Extension knowledge and extension knowledge reasoning [J]. [Journal of Harbin Institute of Technology](#), 2006, 38(7): 1094-1096.
 - [20] 杨春燕, 汤龙. 中国人工智能学会系列研究报告—中国原创学科可拓学发展报告2016[R]. 北京: 北京邮电大学出版社, 2017.
 - [21] SMARANDACHE F. Generalizations of the distance and dependent function in Extenics to 2D, 3D, and n-D [J]. Global Journal of Science Frontier Research, 2012, 12(8): 47-60.
 - [22] SANDRU O I, LUIGE V, PAUL S. Multidimensional Extenics theory [J]. University Politehnica of Bucharest Scientific Bulletin-Series A: Applied Mathematics and Physics, 2013, 75(1): 3-12.
 - [23] VLADAREANU V, SMARANDACHE F, VLADAREANU L. Extension hybrid force-position robot control in higher dimensions [J]. Applied Mechanics and Materials, 2013(332): 260-269.
 - [24] 蔡文, 杨春燕, 陈文伟, 等. 可拓集与可拓数据挖掘[M]. 北京: 科学出版社, 2008.
 - [25] LI X S, ZHANG H, ZHU Z, *et al.* An intelligent transformation knowledge mining method based on Extenics [J]. Journal of Internet Technology, 2013, 14(2): 315-325.
 - [26] 崔春生, 赖锴, 陈婕. 基于基元理论的内容推荐算法研究[J]. 系统科学与数学, 2015, 35(10): 1209-1218.
CUI C S, LAI K, CHEN J. Research on content-based recommendation of ecbased on basic element theory [J]. Journal of Systems Science and Mathematical Sciences, 2015, 35(10): 1209-1218.
 - [27] 李兴森, 石勇, 张玲玲. 从信息爆炸到智能知识管理[M]. 北京: 科学出版社, 2010.
 - [28] LI X S, LI L, CHEN Z. Toward Extenics-based innovation model on intelligent knowledge management [J]. [Annals of Data Science](#), 2014, 1(1): 127-148.
 - [29] LI X S, TIAN Y, SMARANDACHE F, ALEX R. An extension collaborative innovation model in the context of big data [J]. International Journal of Information Technology & Decision Making, 2015, 14(1): 69-91.
 - [30] 杨春燕, 蔡文. 基于可拓学的创意生成与生产研究[J]. [广东工业大学学报](#), 2016, 33(1): 12-16.
YANG C Y, CAI W. Generating creative ideas for production based on Extenics [J]. [Journal of Guangdong University of Technology](#), 2016, 33(1): 12-16.
 - [31] 余志伟, 李兴森. 基元库构建模型及其应用研究[J]. [广东工业大学学报](#), 2015, 32(3): 5-9.
YU Z W, LI X S. Modeling of basic-element and its application [J]. [Journal of Guangdong University of Technology](#), 2015, 32(3): 5-9.
 - [32] 徐慧, 陈宏伟, 宗欣露. 基于可拓论的SDN管理信息语义不匹配问题研究[J]. [广东工业大学学报](#), 2017, 34(4): 17-21.
XU H, CHEN H W, ZONG X L. Semantic-mismatching of management information for software-defined networking based on extension theory [J]. [Journal of Guangdong University of Technology](#), 2017, 34(4): 17-21.
 - [33] 李文军, 杨春燕. 基于事元理论的用户需要与产品功能的形式化研究[J]. [广东工业大学学报](#), 2017, 34(6): 1-8.
LI W J, YANG C Y. A Formal research on user needs and functions of products based on affair-element theory [J]. [Journal of Guangdong University of Technology](#), 2017, 34(6): 1-8.
 - [34] 陶星, 李卫华, 汪中飞. 基于知网的可拓领域信息元库的构建方法[J]. 智能系统学报, 2015, 10(5): 790-796.
TAO X, LI W H, WANG Z F. Construction of howNet-based extendable domain information element base [J]. CAAI Transactions on Intelligent Systems, 2015, 10(5): 790-796.
 - [35] 王景丽, 李兴森. 统一内容标签的基元表达及其构建策略初探[J]. [广东工业大学学报](#), 2014, 31(2): 14-20.
WANG J L, LI X S. On basic element expression and construction of the unified content label [J]. [Journal of Guangdong University of Technology](#), 2014, 31(2): 14-20.
 - [36] LI Q X, LI X S. The method to construct elementary dependent function on single interval [J]. Key Engineering Materials, 2011, 476: 651-654.
 - [37] GUDIVADA V N, RAO D, PARIS J. Understanding search engine optimization [J]. Computer, 2015, 48(10): 145-154.
 - [38] HONG J, HE Z, BELL D A. An evidential approach to query interface matching on the deep web [J]. [Information Systems](#), 2010, 35(2): 140.
 - [39] KIM K, TSE E. Inferior search engine's optimal choice: knowledge-sharing service versus search quality [J]. [Electronic Commerce Research and Applications](#), 2014, 13(6): 387-401.

- [40] BALSIOY O, JIN Y, AYDIN G. Automating metadata web service deployment for problem solving environments [J]. *Future Generation Computer Systems*, 2005, 21(6): 910-919.
- [41] CHEN L, SHADBOLT N R, GOBLE C. Managing semantic metadata for web/grid services [J]. *International Journal of Web Services Research*, 2006, 3(4): 73-94.
- [42] 郭卫宁, 司莉. 国外语义搜索引擎调查与分析[J]. *图书情报工作*, 2013, 57(23): 121-129.
GUO W N, SI L. Investigation and analysis on semantic search engines abroad [J]. *Library and Information Service*, 2013, 57(23): 121-129.
- [43] LAI L F, WU C C, LIN P Y, *et al.* Developing a fuzzy search engine based on fuzzy ontology and semantic search [J]. *IEEE International Conference on Fuzzy Systems*, 2011: 2684-2689.
- [44] SHAH D, SOMAIYA J, NAIR S, *et al.* Fuzzy semantic search engine [J]. *International Journal of Computer Applications*, 2014, 107(15): 25-27.
- [45] TAHA K, ELMASRI R. BusSEngine: a business search engine [J]. *Knowledge and Information Systems*, 2010, 23(2): 153-197.
- [46] 李江华, 时鹏, 胡长军. 本体搜索与排序方法研究综述[J]. *小型微型计算机系统*, 2013, 34(10): 2396-2406.
LI J H, SHI P, HU C J. Overview of ontology search and ranking [J]. *Journal of Chinese Computer Systems*, 2013, 34(10): 2396-2406.
- [47] MYLONAS P, VALLET D, CASTELLS P, *et al.* Personalized information retrieval based on context and ontological knowledge [J]. *The Knowledge Engineering Review*, 2008, 23(1): 73-100.
- [48] MIZZARO S, VASSENA L. A social approach to context-aware retrieval [J]. *World Wide Web*, 2011, 14(4): 377-405.
- [49] 刘剑, 许洪波, 贾岩涛, 等. 面向智能搜索的动态知识网络建模[J]. *电信科学*, 2014, 30(10): 64-70.
LIU J, XU H B, JIA Y T, *et al.* Dynamic knowledge network modeling orient to intelligent search [J]. *Telecommunications Science*, 2014, 30(10): 64-70.
- [50] PETERS S, RUECKEMANN C P, SANDER-BEUERMANN W. A new approach towards vertical search engines-intelligent focused crawling and multilingual semantic techniques[J]. *Webist 2010: Proceedings of the 6th International Conference on Web Information Systems and Technologies*, 2012: 181-186.
- [51] TAHA K, ELMASRI R. XCDSearch: an XML context-driven search engine [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2010, 22(12): 1781-1796.
- [52] RODRIGUEZ-GARCIA M A, RAFAEL V, FRANCISCO G. Ontology-based annotation and retrieval of services in the cloud [J]. *Knowledge-Based Systems*, 2014, 56: 15-25.
- [53] KOO D M, CHOI Y Y. Knowledge search and people with high epistemic curiosity [J]. *Computers in Human Behavior*, 2010, 26(1): 12-22.
- [54] SHROFF, GAUTAM. The intelligent web: search, smart algorithms, and big data[J]. Oxford University Press, 2013.
- [55] 陈震, 曹军威. 信息中心网络[M]. 北京: 清华大学出版社, 2013.
- [56] HONG, JUNE S. Development of constraint-based problem solving framework for the semantic web [J]. *Entrue Journal of Information Technology*, 2011, 10(2): 239-250.
- [57] ANNELIES R, SCHELLENS T, DE W B. Promoting meta-cognitive regulation through collaborative problem solving on the web: when scripting does not work [J]. *Computers in Human Behavior*, 2016, 58: 325-342.