

• 软件技术与数据库 •

文章编号:1000-3428(2010)23-0036-04

文献标识码:A

中图分类号:TP18

基于案例推理的 ISP 知识重用方法

刘冀琼^{1,2}, 李兴国¹, 顾东晓^{1,3}, 冯 帅^{1,2}

(1. 合肥工业大学管理学院, 合肥 230009; 2. 阜阳师范学院经济与商业系, 安徽 阜阳 236041;
3. 威斯康星大学卢芭商学院, 威斯康星 密尔沃基 53211)

摘要:针对企业信息化中信息系统规划(ISP)在较大程度上依赖先前规划经验和知识的特点,为有效支持ISP知识重用和管理创新,将案例推理技术应用于ISP知识重用的全过程,提出基于案例推理的ISP知识重用方法(HRM-KGANNA),研究应用框架、案例表示、案例检索算法与重用等关键技术。通过组织案例库进行ISP知识检索实验,结果表明HRM-KGANNA法具有良好的准确性和较高的效率。

关键词:案例推理; 信息系统规划; 知识重用; 案例表达

Case-Based Reasoning ISP Knowledge Reuse Method

LIU Ji-qiong^{1,2}, LI Xing-guo¹, GU Dong-xiao^{1,3}, FENG Shuai^{1,2}

(1. School of Management, Hefei University of Technology, Hefei 230009, China;
2. Department of Economics and Business, Fuyang Normal University, Fuyang 236041, China;
3. Lubar School of Business, University of Wisconsin-Milwaukee, Milwaukee 53211, USA)

【Abstract】Information System Planning (ISP) of enterprises mostly depends on previous knowledge and experience. In order to effectively support ISP knowledge reuse and management innovation, this paper uses the Case-Based Reasoning (CBR) technology to the whole process of ISP knowledge reuse and proposes a case-based knowledge reuse method for ISP: Hybrid Retrieval Method combining Knowledge-Guided Approach with Nearest-Neighbor Approach (HRM-KGANNA) method. The application framework, case representation, case retrieval algorithm and case reuse are thoroughly explored. After organizing the cases base, the ISP knowledge retrieval experiment is conducted, and experimental results demonstrate that HRM-KGANNA can satisfy retrieval requirements with high retrieval accuracy and efficiency.

【Key words】Case-Based Reasoning(CBR); Information System Planning(ISP); knowledge reuse; case representation

1 概述

企业信息系统规划(Information System Planning, ISP)在企业的信息化建设中起着至关重要的作用,即企业信息系统建设的基础,是企业了解自身情况、整合企业信息资源与战略、提高企业自身能力的复杂过程^[1]。但企业的信息系统规划具有知识不完备、不确定性等特点,不可能存在一套完备的规则来分析指导其规划过程。近年来,国内外不少学者提出了关于ISP的主要内容框架和其规划方法。其中,文献[2]提出了一个七步骤的包含输入-处理-输出的ISP方法框架。文献[3]提出了一个具有箱子结构的ISP方法以解决规划与实施间的隔阂问题。文献[1]分析总结了信息系统的规划纬度。文献[4]结合我国信息技术的实施情况,提出了一个IS/IT战略规划模型框架,并对规划过程进行了分析。文献[5]提出了一个以输出为动力的ISP规划方法。

以上文献虽对其提出的方法进行了案例分析,并提供个别企业ISP的详细描述,但也存在着一些问题:

(1)有些规划方法的适用面较窄。如文献[2-3]提出的方案仅面向战略信息系统(Strategic Information Systems, SIS)的规划,而SIS仅是信息系统的一部分。该方法对于业务层、管理层的信息系统规划并不适用。

(2)有些规划方法偏重于对信息系统规划效率影响因素

的研究,并不能形成一套完整的信息系统规范方案。如文献[1]提出的方法主要是对信息系统规划纬度的研究。

(3)以上方法没有根本解决信息系统规划知识的表示、快速获取和重用问题,也无法在较短的时间内对企业信息系统规划提供知识支持和决策辅助。

基于案例推理(Case-Based Reasoning, CBR)是一种基于经验知识(案例)进行推理的人工智能方法。它通过构造并组建丰富主体的案例库,在问题求解机制中采用基于案例的推理策略,其方法模仿人类决策过程中类比的认识方式,对于解决非结构化的、知识贫乏的领域问题较为有效^[6]。因此,CBR技术的特点使其非常适用于解决知识共享和知识重用问题。本文将CBR引入到信息系统规划工作中,提出了一种基于案例推理的ISP知识重用方法。

基金项目:国家自然科学基金资助重点项目(70631003);国家自然科学基金资助项目(70871033);合肥工业大学科学发展基金资助项目(2009HGXJ0039)

作者简介:刘冀琼(1984—),女,硕士研究生,主研方向:案例推理,物流与供应链管理;李兴国,教授;顾东晓,讲师、博士研究生;冯 帅,硕士研究生

收稿日期:2010-06-25 **E-mail:**liupaper@gmail.com

2 信息系统规划的知识重用框架

将 CBR 技术引入到企业信息化的信息系统规划中,提出了一个基于案例推理的信息系统规划(CBR-ISP)的知识重用框架,如图 1 所示。

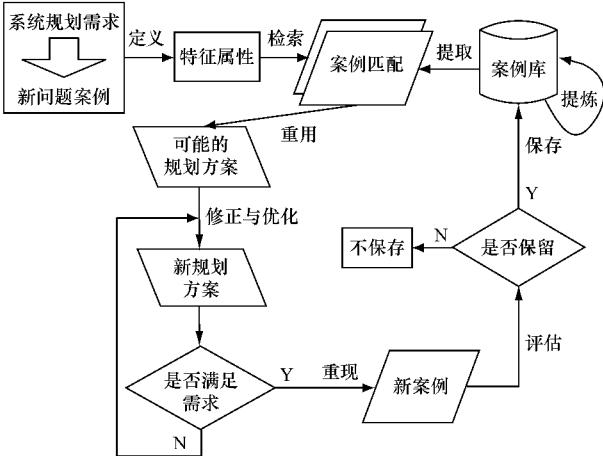


图 1 基于案例推理的信息系统规划的知识重用框架

CBR-ISP 重用技术的基本工作原理为:以案例的形式来映射 ISP 知识和企业的知识需求,通过检索并修正优化以前相似的规划方案,从中得出一个新的 ISP 方案,而不需要从零开始。

新案例需要经过预处理和过滤才能保留在案例库中,有时还会对以往案例及其相应的规划方案进行适应性修改,加入专家知识,组织新案例。值得指出的是,无论案例推理是否成功,都会导致案例库中的案例索引方案和特征权重的改进^[7]。

3 信息系统规划案例知识重用的关键技术

3.1 案例知识表示

案例知识表示是 ISP 知识重用过程中的基础性工作,其表示方法直接关系到案例推理的效率和准确度。非结构化已经成为实际应用的案例中普遍存在的性质,为了更加灵活、丰富、完整地表示和存储非结构化案例,本文将采用 XML 语言表示与组织案例。与传统的案例表示方法相比,采用基于 XML 方法描述案例具有明显的优势^[8],如:案例结构相对灵活,存储相对独立,XML 文档中的节点自定义功能有助于建立新的案例描述和功能,案例信息更加丰富与完整,案例资源可共享、跨平台应用等。

在 CBR-ISP 中,案例分为求解案例和问题案例。

定义 1 求解案例是指在满足信息系统规划要求和满足企业知识需求条件下所获得的规划结果案例、其组织与存储于案例库。

根据 ISP 特点,参照文献[5]所指出的以输出为动力的 ISP 规划方法,将 ISP 求解案例表示成一个八元特征向量组,即:

Case {case_info, enterprise_info, objective_mix, status_anlysis, IS_SWOT, business_flow_analysis, project_evaluation_analysis, planning_result}

其中,case_info 表示案例的基本信息,包括案例编号、案例名、案例输入时间和案例整理人等内容;enterprise_info 表示企业的基本信息,包括企业名称、所属行业、企业性质和企业规模

等内容;objective_mix 表示目标集合,包括企业发展目标、企业发展战略、ISP 目标、ISP 战略、ISP 范围和 ISP 责任群体等内容;status_anlysis 表示现状分析,包括分析行业现状、IS/IT 能力现状、软硬件现状、人员素质、技术水平、外部环境、资源约束等内容;IS_SWOT 表示 IS 的 SWOT 分析,包括分析 IS 的优势、劣势、机会和威胁;business_flow_analysis 表示业务流程分析与整合,包括识别企业业务功能、业务流程分析、业务流程重组、识别关键成功因素等内容,其中业务流程重组包含重组后的业务流程及其与数据库文件映射表和数据结构表等;project_evaluation_analysis 表示项目评估分析,包括 IS 架构、项目规模范围、成本分析、效率分析和开发时间计划等内容;planning_ruslut 表示 ISP 规划结果方

基于 XML 的案例表示一般采用 DTD(Document Type Definition)或 XML Scheme 定义其 XML 文档结构。本文采用 DTD 方式定义案例结构,DTD 表示如下所示。限于篇幅,仅给出部分 DTD 定义:

```

<! ELEMENT Case (case_info, enterprise_info *, objective_mix *,
*, status_anlysis *, IS_SWOT *, business_flow_analysis *,
project_evaluation_analysis *, planning_result *)>
<! ELEMENT case_info (case_ID, case_name, case_input_time,
case_input_person)>
<! ELEMENT case_ID (#PCDATA)>
...
<! ELEMENT enterprise_info (enterprise_name, industry_type,
enterprise_nature, enterprise_scale)>
...
<! ELEMENT objective_mix (enterprise_objective, enterprise_
strategy, ISP_objective, ISP_strategy, ISP_range, ISP_groups)>
...
<! ELEMENT status_anlysis (industry_status, IS_capacity,
hardware_soft_status, personnel_quality,
technical_level, external_environment, resource_constraints)>
...
<! ELEMENT IS_SWOT (strength, weakness, opportunity,
threat)>
...
<! ELEMENT business_flow_analysis (business_functions, flow_
analysis, business_process_reengineer, critical_success_factors)>
...
<! ELEMENT business_process_reengineer (reengineered_
business_process, database_business_process, data_structure)>
...
<! ELEMENT project_evaluation_analysis (IS_framework,
project_scale, cost_analysis, efficiency_analysis, time_scheme, risk_
analysis)>
...
<! ELEMENT planning_result (attachment)>
<! ELEMENT attachment (#PCDATA)>
  
```

根据 XML 的特点,基于上述 DTD 的求解案例具有可扩充性,并且其结构也可做适应性修改。在求解案例的 8 类信息中,enterprise_info、objective_mix、status_anlysis、IS_SWOT、business_flow_analysis、project_evaluation_analysis 对应于案例描述,各元素或其下属子元素均可选作为案例的特

征属性;planning_result 对应案例的解决方案。

定义 2 问题案例是反映企业现状和知识需求、需要在 CBR-ISP 中被求解的新案例。

求解案例是已经存储在案例库中的案例,其拥有特定的知识结构和内容描述,为了使得由企业现状和知识需求所形成的问题案例在案例库中更好地检索与匹配类似案例(求解案例),则要求问题案例与求解案例有相同或相似的组织结构。

3.2 案例检索

案例检索模型及其检索策略对于案例推理系统的学习和推理性能有着重要影响,其本质是问题案例与求解案例之间的相似匹配^[7]。目前常用的方法有最近相邻法、归纳索引法和知识引导策略等。根据 ISP 方案制定的特点,本文采用知识引导策略和最近相邻法相结合的混合检索策略(Hybrid Retrieval Method combining Knowledge-Guided Approach with Nearest-Neighbor Approach, HRM-KGANNA)。然而由于 ISP 案例具有针对性,并难以定义关键的特征检索属性,因此所研究的案例检索方法侧重于检索的准确性,而非自动检索的效率。按照该方法,对 ISP 案例的检索将分为初步检索和最佳匹配检索 2 个阶段。

3.2.1 初步检索

本文采用知识引导策略进行案例的初步检索。知识引导策略是根据已有的知识和经验来确定案例检索的关键字,并根据这些关键字检索与匹配案例,如:企业所属行业与企业发展目标对企业 ISP 方案的制定有着重要影响,初步检索可以将企业所属行业作为第 1 检索条件,以企业发展目标作为第 2 检索条件。在初步检索阶段,根据已有的知识、经验和企业需求,选取多关键字在案例库中检索出与问题案例相关的求解案例,直至形成满意的候选案例集合,并选取主要特征属性,为下一阶段的检索做准备。

3.2.2 最佳匹配检索

最佳匹配检索的核心是相似度计算。设论域 C^* 是案例库中所有求解案例的集合, c_0 为输入待求解的问题案例, C 为通过初步检索得到的候选案例集合, 则 $C \subseteq C^*$ 。

设 A 是经过挑选后的主要特征属性的集合:

$$A = \{\text{属性 } 1, \text{属性 } 2, \dots, \text{属性 } n\} = \{a_1, a_2, \dots, a_n\} \quad (1)$$

首先将定性属性特征值定量化处理,搭配元素 (c_i, a_j) , 即案例 $c_i (i \leq m)$ 的第 j 个特征属性 a_j 的属性值为 v_{ij} , 则 $V = (v_{ij})_{m \times n}$ 。再对 V 中的属性值 v_{ij} 进行标准化处理, 处理后得到新矩阵:

$$V' = (v'_{ij})_{m \times n} \quad (2)$$

考虑到 ISP 的复杂性,大多知识需求仅能选取有限个特征属性进行分析,CBR 系统具有信息不完全,或者“灰色”的特征^[7,9]。本文将采用灰色关联分析结合欧氏距离加权来计算相似度。

案例空间 C 中各案例 $c_i (1 \leq i \leq m)$ 在属性 $a_j (1 \leq j \leq n)$ 时与问题案例 c_0 的关联系数为:

$$\xi_{ij} = \frac{\min_{i'} |v'_{0j} - v'_{ij}| + \rho \max_{i'} |v'_{0j} - v'_{ij}|}{|v'_{0j} - v'_{ij}| + \rho \max_{i'} |v'_{0j} - v'_{ij}|} \quad (3)$$

其中, v'_{0j} 为问题案例 c_0 在第 j 点标准化处理后的属性值, $1 \leq i \leq m, 1 \leq j \leq n$; ρ 为分辨系数, $0 \leq \rho \leq 1$, 其值可以根据需要人为设定,一般选取 $\rho = 0.5$ 。可见, $0 \leq \xi_{ij} \leq 1$ 。

式(3)确定了灰色关联系数矩阵,但是并非每一特征属性的重要度是一样的,因此,结合系统的具体运行特点,运用层次分析法确定归一化特征属性权值矢量 $w = (w_1, w_2, \dots, w_j, \dots, w_n)^T$, 且 $\sum_{j=1}^n w_j = 1$ 。

则 c_0 与 c_i 的相似度可定义为 1 与欧氏距离之差:

$$sim_{0i} = 1 - d_{0i} = 1 - \sqrt{\sum_{j=1}^n [w_j (1 - \xi_{ij})]^2} \quad (4)$$

根据式(4)的物理意义, sim_{0i} 值越大, 表明其相应的案例越类似于问题案例 c_0 。因此,比较 sim_{0i} 大小 ($1 \leq i \leq m$), 求出相似度最大的案例,作为最佳匹配案例。对于实际问题,最佳匹配案例的规划方案是否可以作为问题案例的解决方案,应根据设定的阈值确定,只有相似度大于规定的阈值时,方可采用此案例对应的规划方案作为可能的方案。也可以依据相似度对类似案例进行排名,再根据企业知识需求从排名靠前的案例中进一步筛选得到最佳匹配案例。

需要说明的是,为了能有效地进行检索,需要首先确定案例属性的初始权重。在检索过程中,如果发现案例属性特征权重不能正确赋值或者案例检索质量将难以得到保证,此时也需要对权重进行必要的调整和优化。在 ISP 案例的检索中,对于初始权重的确定采用的是经典的德尔非法进行,依据系统的程序,采用 ISP 专家匿名发表意见的方式,以反复填写问卷和集结问卷填写人的共识来获得 ISP 案例各属性的初始权重。在检索过程中,采用文献[10]所提出禁忌遗传算法来优化案例属性特征权重,通过利用禁忌算法的自适应性和具有记忆功能的特点来改善遗传算法的全局搜索能力和提高其收敛速度,从而对特征权重进行自动优化,提高案例检索精度。

3.3 案例重用

在 CBR-ISP 中,若检索结果相似度 $sim_{0i} = 1$, 则基本上可以认为求解案例的规划方案符合企业的知识需求,可以直接重用。但在实际问题中,更多的情况是 $0 \leq sim_{0i} < 1$, 这时需要适当调整、修正与优化在检索操作中获得的求解案例的规划方案,使之适合企业的知识需求。由于案例修订所涉及的知识大多属于专家或者企业所拥有的隐形知识,若完全由系统自动修订将十分困难,因此在 CBR-ISP 中,将采用用户干预的形式进行案例的修订。

整合修订后重现新案例,并评估新案例是否有保留价值,若有保留价值,则作为新知识添加进案例库,形成了对问题求解能力的扩充,即完成了案例学习。

4 实验与应用分析

为验证本文所提出的 HRM-KGANNA 法的有效性,用 Java、XML 和 SQL Server2000 作为开发工具和 DBMS 实现一个融合 HRM-KGANNA 法、面向 ISP 应用的“信息系统规划知识获取与方案生成系统(CBR-ISP2009)”原型。在 CBR-ISP2009 实际应用之前,大量收集 ISP 案例,按照 XML 的 DTD 规范描述案例,并组织存储于案例库。在 CBR-ISP2009 中,对于新问题案例的求解,首先根据企业现状和 ISP 规划需求,应用 XML 解析器初步检索得到候选案例集合,并选取主要的特征属性。然后根据主要特征属性对候选案例集合与问

题案例进行相似度匹配计算,从而检索得到求解案例号。根据最类似案例号得到其对应的 ISP 结果方案,通过学习,可以很容易地从知识库中提取相关知识或拥有知识的专家形成规划方案,然后再调用知识库中的知识对 ISP 结果方案进行修正与优化。对优化后的案例再次审查与评估是否具有保留价值,若有则还要作为新案例充实到案例库,完成案例学习。

初始阈值和初始权重可以由领域专家通过 DELPHI 法得到。为了让专家的意见更具有合理性和更符合系统实际和实验要求,得到最佳的初始阈值和初始权重,在 CBR-ISP2009 的系统设计中增加了训练库,并准备了 100 个 ISP 方案进入训练库让专家操作。各个专家在 CBR-ISP2009 系统中利用训练库里的数据通过多次反复试验,给出阈值和权重,通过若干次集结和调整,最终确定初始阈值为 $\lambda_0 = 0.67$, 初始权重集 $W_0 = (0.04, 0.05, 0.12, 0.14, 0.21, 0.23, 0.11, 0.10)$ 。接下来再组织 120 个 ISP 方案作为实验案例,以 20 个为一批分 6 次输入 CBR-ISP2009 系统作为初始案例库,进行某医院 ISP 方案的检索实验。输入各特征属性值进行检索,6 次检索所得结果及其对数图趋势见图 2。

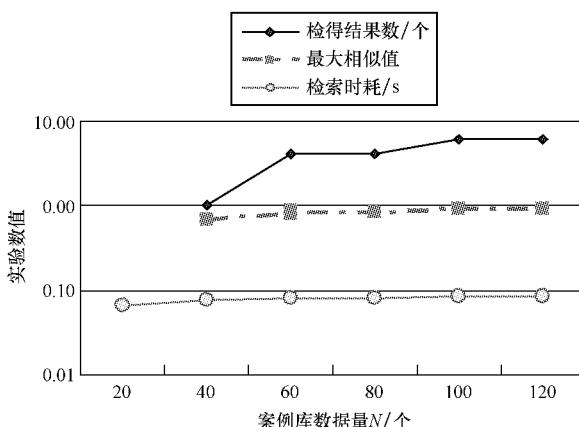


图 2 HRM-KGANNNA 法的检索实验结果

由图 2 可见,HRM-KGANNNA 法可以有效地进行 ISP 知识的获取,随着案例库数据量的增加,检索的精确度逐渐提升(检得的最大相似值呈单调不减趋势),检索的时耗总体较小,并且在数据量不断递增的情况下保持相对的稳定,体现了较好的检索效率。

下面在 CBR-ISP2009 平台上进行某高校实验室建设与维护管理 ISP 方案辅助设计的应用。某高校为进一步提高学校实验室建设、维护和管理水平,该校实验室与装备处组织规划和建设一个面向本单位实验室建设、维护与管理业务的信息系统。利用 CBR-ISP2009 平台,输入问题案例各个特征属性值,进行问题求解,获得的结果如表 2 所示。

表 2 ISP 方案辅助设计实验结果

案例编号	案例名称	相似性	案例日期
ISP0077	某高校实验室建设与管理信息系统规划方案	0.925	2005-10-17
ISP0109	某高校实验室管理信息系统规划	0.783	2009-10-26
ISP0115	某研究院实验室管理信息系统规划	0.686	2008-09-28

该处与某信息系统研究所结合本单位实际情况,对 ISP0077 号方案进行了较小幅度的修改,在较短的时间内就获得了该校实验室建设与维护管理 ISP 方案,较大幅度地降低了信息系统规划时间。

5 结束语

本文将案例推理技术运用于信息系统规划工作和知识重用的过程中,提出了案例推理在 ISP 知识重用系统中的应用框架,研究探讨了案例表示、案例检索和重用等关键技术,并进行了相关实验。实验表明,本文所提出的基于案例推理的 ISP 方法具有良好的准确性和较高的效率。某单位利用 CBR-ISP2009 平台进行该高校实验室建设、维护管理 ISP 方案获取的具体应用进一步说明了该方法的可应用性、输出结果的高质量性和方案获取的快捷性。作为研究的初步成果,本资料尚存在原型系统简单、数据量小等问题,今后的研究将围绕开发功能完善的信息系统规划 CBR 系统并进行海量数据下的实验和更深入的应用。

参考文献

- [1] Wang E T G, Tai J C F. Factors Affecting Information Systems Planning Effectiveness: Organizational Contexts and Planning Systems Dimension [J]. Information and Management, 2003, 40(4):287-303.
- [2] Lederer A L, Salmela H. Toward a Theory of Strategic Information Systems Planning [J]. Journal of Strategic Information Systems, 1996, 5(3): 237-253.
- [3] Hevner A R, Bemdt D J, Studnicki J. Strategic Information Systems Planning with Box Structures[C]//Proc. of the 33rd Hawaii International Conference System Sciences. Hawaii, USA: [s. n.], 2000.
- [4] 张玲玲,林 健.企业 IS/IT 战略规划模型框架研究[J].系统工程,2001, 19(2): 33-35.
- [5] 傅湘玲,方 红.以输出为动力的信息系统规划方法[J].情报科学,2004, 22(8): 989-1015.
- [6] 孙岩清,尹树华,王 技.基于粗糙集的 CBR 系统属性约简改进算法[J].计算机工程,2010, 36(10): 38-40.
- [7] 王 君,潘 星,李 静,等.基于案例推理的知识管理咨询系统[J].清华大学学报: 自然科学版, 2006, 46(z1): 990-995.
- [8] Changchien S W, Lin M C. Design and Implementation of a Case-based Reasoning System for Marketing Plans[J]. Expert Systems with Applications, 2005, 28(1): 43-53.
- [9] 徐晓臻,高国安.案例推理在多准则评价智能决策支持系统中的应用研究[J].计算机集成制造系统, 2001, 7(1): 16-28.
- [10] 梁昌勇,顾东晓.面向不确定多属性决策问题的范例检索算法研究[J].中国管理科学, 2009, 17(1): 131-137.

编辑:任吉慧