

大数据挖掘在工程项目管理中的应用

曾 晖

(五邑大学 土木建筑学院, 广东 江门 529020)

摘 要:大数据时代背景下传统的工程项目管理已经不能适应科学管理的要求,而数据挖掘这一技术手段为工程项目管理提供了新的提升路径。从大数据背景出发,分析了工程项目管理的困境,提出了构建大数据挖掘的管理层次和制度结构以及大数据挖掘项目组解决方法,基于房地产工程项目设计了工期进度控制模型并进行了应用性研究。

关键词:大数据;数据挖掘;工程项目;工程管理

DOI:10.6049/kjbydc.2014GC0303

中图分类号:F062.4

文献标识码:A

文章编号:1001-7348(2014)11-0046-03

0 引言

随着信息化和知识化时代的到来,尤其是大数据的蓬勃发展^[1],我国工程项目管理受到了严峻挑战,以往的管理方法已经难以适应时代要求。大量的信息和数据是工程项目建设的根基,从科技管理角度来看,这些信息资源的科学管理成为现阶段工程管理的重要瓶颈^[2]。工程项目管理是一种以工程项目为对象的系统管理方法,通过对工程项目的全过程动态管理来实现整体目标。鉴于工程项目的系统性、动态性以及时代要求,本文认为大数据技术的出现为工程项目管理带来了新的发展方向,将大大提升工程项目管理各环节和整体的信息处理效率,为项目决策提供有效的信息参考,进而实现项目效益增值。

1 大数据挖掘与工程项目管理交互分析

1.1 大数据背景下工程项目管理发展趋势

在社会经济高速发展的今天,先进的科技管理成为各行各业获得竞争优势的途径之一,也是工程项目管理中的重要工作内容。科技管理以计算机为技术依托,以庞大的信息流和资源流为管理对象,为冗杂的信息资源的搜集、筛选、存档以及传送提供一个网络信息平台,以实现科技管理活动的流程化、信息化和自动化^[3]。基于科技管理的这些特点,我国工程项目管理也呈现出数据多元化、动态化以及信息化管理等发展趋势。一方面,在传统行业中,工程行业是数据量最大、项目规模最大的行业,参与主体多、覆盖地域范围广、

耗费时间长、影响因素多等特征决定了工程项目的信息管理具有多元性。信息数据的多元性体现在工程管理的各个环节,如前期决策中信息来源渠道多,人力、物力、财力管理都基于各种信息资源^[4]。另一方面,工程项目管理采取全周期管理模式,时间周期长,各种信息流在动态的时间流中持续分布。因此,工程项目的信息化管理是大势所趋。

1.2 大数据挖掘为工程项目管理优化提供了新路径

大数据的出现将为工程项目的科技信息管理创造新的发展契机,为工程项目的效率管理、质量管理、风险管理等创造优化路径。大数据挖掘有助于工程项目管理效率提升,由于项目的系统性和复杂性,工程项目管理的低效率普遍存在,而大数据挖掘技术凭借先进的技术手段提高数据管理效率。以工程项目管理的绩效评估为例,绩效评估常常出现指标过多、评价成本过高等问题,大数据挖掘为解决这一问题带来了新方法。在工程项目管理中引入大数据挖掘技术,可以从庞大的数据库中找到最符合项目要求的绩效指标即关键绩效指标,这将减少工程项目管理的工作量,提高绩效管理效率。

大数据挖掘为工程项目管理的全面风险管理带来了新思路。在工程行业中,庞大复杂的数据中隐藏着各种风险,对项目乃至企业长期发展带来隐患。大数据管理中,数据仓库不仅能及时收集现有和历史数据,还能对各个孤立存在的数据进行初步处理和转换,形成相互联系的统一数据集,为项目中各数据使用者提供一个透明的信息平台,减少信息流通中虚假信息和交流障碍等因素带来的风险。

收稿日期:2014-03-31

作者简介:曾晖(1969—),女,四川成都人,博士,五邑大学土木建筑学院副教授,研究方向为工程管理、房地产经济。

2 大数据时代背景下工程项目管理困境

(1)随着需求多元化的发展,生产贴合市场个性化需求的工程产品面临新的挑战。工程设计和评估过程中由于存在固有的刚性和惯性,使得很难实现与市场需求的高度贴合。在大数据背景下,市场需求不断转化为各类数据,如果不能对这些数据作及时、科学的处理,就可能造成困境:一是由于对数据的不完全解读,使得工程设计和评估与市场不完全贴合,即产出的最后产品不能最佳地满足市场需要;二是由于对数据的误判,使得工程的设计和评估完全偏离市场需要,即最后产品不能为市场所接受^[5]。由此可见,市场需求的多元化使得数据呈爆炸式增长,而工程项目管理极容易在众多数据中迷失方向,从而陷入困境。

(2)经济环境的快速变化给工程项目管理带来了诸多不确定性,使得工程项目管理时刻面临风险。技术更新频率加快,社会经济环境突变的可能性也随之增加,这对保障工程项目的进度、成本、质量、安全都带来了巨大挑战^[6]。例如,工程规模不断增大,所需资金量也随之增加,这必然产生海量的成本数据和资金数据,传统的工程预决算管理模式根本无法适应大工程项目建设,极易影响工程进度和成本控制。再如,工程规模的增大必然导致工程项目基础数据的巨量膨胀,传统的施工管理模式不仅容易造成安全隐患,而且无法保证工程整体质量。

3 大数据挖掘对工程项目管理优化路径

(1)构建大数据挖掘的管理层次和制度结构。首先,按照集中控制和分层管理的思路,确立项目公司作为数据收集者、集团公司作为数据决策者的回路模式。以数据为控制载体,项目公司按照集团公司的数据要求及时准确地采集数据,集团公司以总体数据为依据进行进度、成本、质量、安全方面的分析和决策。这里的总体数据既包括项目公司采集的内部数据,还需要集团公司采集外部数据,以保证数据完整性^[7]。其次,按照数据集中、业务集中、管理集中、控制集中的原则,建立数据处理中心及业务审批、项目施工、公司决策层数据沟通制度。项目部与施工现场人员业务往来形成的各类数据,由项目部整理和识别后录入信息系统中心,数据处理中心对总体数据进行挖掘处理后向公司决策层提供分析和辅助决策支持,各职能部门可以随时调用项目数据进行管理,项目部根据数据指标及其提示进行施工作业和相关管理。

(2)构建大数据挖掘项目组,解决项目管理中的主要问题。构建大数据挖掘项目组的目的是保证在一定资源约束的前提下,使工程项目以尽可能快的速度、尽

可能低的成本达到最好的质量效果^[8]。①建立工期进度数据挖掘项目组。整合资金数据、供应商数据、工程计划数据、施工基础数据等,通过数据挖掘建立相应的控制体系以保证工期进度有效推进;②建立工程质量数据挖掘项目组。整合施工基础数据、质量检测数据、物流仓储数据、工期进度数据等,通过数据挖掘建立相应的控制体系,避免因物料管理不规范、阶段验收和隐蔽工程验收不规范、计划安排不科学导致盲目抢工期,以及设计本身缺陷导致质量失控等问题;③建立成本控制数据挖掘项目组,整合物料数据、成本核算数据、质量控制数据、工程进度数据、资金数据等,通过数据挖掘建立相应的控制体系,避免工期拖延、质量控制不当等问题^[9]。

4 基于数据挖掘的工期进度控制模型

工期进度是工程项目管理中最难控制的问题,工期延误的原因是多方面的,包括前期准备工作不足、项目设计偏差、施工管理不当、设备未按时按要求到位等。在实际施工中,这些导致工期延误的原因并不是必然的^[10],如通常认为下雨等天气情况会影响项目进度,但并不是所有项目都会因雨季而延误工期,所以肯定还存在其它原因。在这些主要原因之外,还有很多被忽略的因素会影响工程进度,实际施工中客观存在一些我们不了解的因果关系,要想更好地了解这些因果关系,需要对大量工程项目管理数据进行挖掘,基于数据分析结果作出决策。

数据挖掘工作首先需要建立相关部门和管理机制,使数据可以从施工一线实时传输至后方管理层并存储起来以供挖掘分析,其行政管理机制可以采用集团总部—地区公司—项目公司—前线模式。项目部与施工现场人员存在业务往来,项目部将各种施工现场信息录入信息系统并传递至公司数据库,同时公司也可调用数据库数据传输至项目部。传输至公司数据库的信息可以提供给各个职能部门作管理调度参考,还可以通过建立各种数据挖掘项目组进行数据挖掘分析。

不同类型的工程管理数据特点不同,要进行数据挖掘必须对不同数据进行分类。对于结构化数据的挖掘,必须在各个专业工程领域建立一个统一的数据仓库,这是提高分析精度的根本保障。数据仓库体系构建应符合一定的关联规则,如图 1 所示。

对于非结构化数据的挖掘,可以采用检索技术进行分类管理。MOLDA 的多维联机数据分析(Multi-dimensional On-Line Data Analysis)理论是一种支持分析和决策管理项目的模型,适合种类繁多的工程项目分类。其设计思路包括多维业务空间设计、挖掘方法设计、数据连接设计 3 个部分。

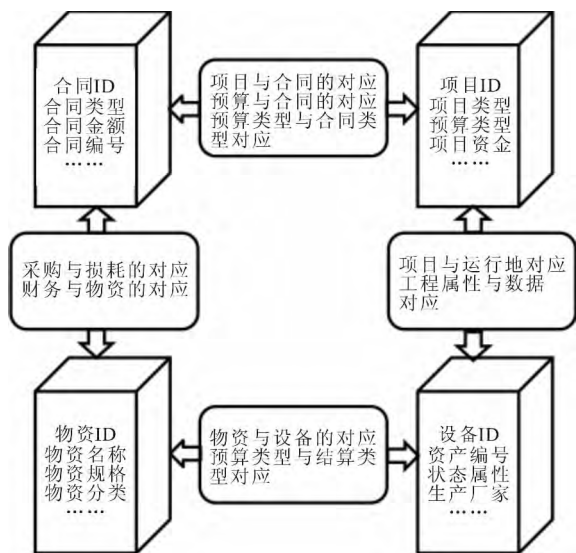


图1 要素关联

4.1 工期延误原因关联分析

分析工期延误原因可以采用关联规则方法,数据挖掘最重要的工作就是发现数据间的关联规则方法。关联即两个或多个数据之间存在一定规律性,关联分析的目的是找出其关联网。本文利用 Apriori 算法建立单层布尔关联关系对工期延误原因进行挖掘模拟。将所有存在工期延误的项目组组成一个集合 Y , 将 Y 中所有导致工期延误的原因组成集合 I , 包括事故、资金短缺、施工监管不当、雨雪、天气炎热、地质变化等, 用 $I = \{i_1, i_2 \dots i_m\}$ 表示。规范的取值比较复杂, 有些原因是隐形的, 这项工作需要在数据准备中完成。通过关联规则方法可以获得一组频繁项集, 从两方面对其进行描述: 一是支持度, 即一个项集在 Y 中出现的百分比, 如果所有工期延误的项目中由于雨雪和施工监管不当造成的工期延误占 20%, 则支持度为 20%; 二是置信度, 如果因为雨雪天气导致的施工监管不当占 50%, 则该频繁项集的置信度为 50%。

4.2 多任务进度汇总权重问题

工程项目一般有多个任务, 通过划分子项目进行管理。管理层希望获得关于工程进度的清晰简单的数值, 如项目完成的百分比等。当存在多个子项目时, 总工程项目完成百分比不是子项目进度的简单相加, 而是根据施工时间采用加成算法得到总工程项目完成的百分比。

进度汇总权重估值的确定可以通过数据挖掘从特征面向属性进行归纳, 将项目进度数据化, 获得子项目中的汇总权重。例如打地基的工作量为 20 天, 水泥浇筑的工作量为 30 天, 项目工期为 30 天, 地基打完后总项目完成 40%, 则项目还需 $50 \times 60\% = 30$ 天即可完成。通过周密的计划将工作进行精确划分, 这样根据工作量计算的工程进度才是精确的。如果一个以上的子项目完成时间大于工作量, 那么总项目进度汇总就

不准确。房地产施工中前期物资采购、设备到位等工作都可能出现完成时间大于工作量的情况, 这时需要通过数据挖掘得到各个子项目的工作权重估值, 以便进度汇总时得到更准确的项目进度结果。

以一组模拟数据进行面向属性归纳的数据挖掘步骤为: ①选定数据仓库中进度控制数据完整、工期没有延误的项目作为分析关系数据集; ②收集关系数据集中项目的所有属性; ③将属性概念进行分层; ④数据概化, 将每个概化记录插入概化关系中。项目概化关系表中重复记录数值显示了符合同一属性项目的个数, 其数值越大表明该权重可信度越高。权重的概化关系可以在制定项目进度计划时作为各个子项目的权重参考估值。随着项目数据的积累, 关系数据集越来越大, 权重的概化关系也越来越有参考价值。得到权重的估值后, 即可通过 Apriori 算法对工程进度作准确估计和管理。

5 结语

大数据挖掘已经在社会经济各个领域得到运用, 本文探讨了其在工程管理领域的应用, 创新之处在于提出了基于工程管理的关键词是建立大数据挖掘项目组, 并对建立“工期进度”数据挖掘项目组进行了模型设计的应用性研究。工程管理领域甚广, 其关键点在不同的工程项目中也不尽相同, 笔者将在以后的研究中进一步加强应用性研究。

参考文献:

- [1] 徐子沛. 大数据: 正在到来的数据革命[M]. 桂林: 广西师范大学出版社, 2012.
- [2] CLIFTON PHUA, VINCENT LEE, KATE SMITH, et al. A comprehensive survey of data mining-based fraud detection research[A]. Working Paper, 2005.
- [3] 田也壮, 周军, 石春生, 等. 科技成果价值的分析及对评估的影响[J]. 科研管理, 2000, 21(2): 62-67.
- [4] 陈士俊, 赖迪辉. 数据挖掘在科技评估中的应用[J]. 科学与科学技术管理, 2005(4): 40-43.
- [5] 隋红建, 王刚. 项目管理在 R&D 中应用的问题及对策[J]. 科学管理研究, 2005, 23(1): 47-49.
- [6] 刘朝明. 地铁基坑安全评估研究[D]. 上海: 同济大学, 2005.
- [7] 苏新宁. 数据挖掘理论与技术[M]. 北京: 科学技术文献出版社, 2003.
- [8] 刘涛. 基于数据挖掘的基坑工程安全评估与变形预测研究[D]. 上海: 同济大学, 2007.
- [9] 段云峰, 夏建国. 我国城镇土地定级估价信息系统研究评述[J]. 四川农业大学学报, 2004, 22(4).
- [10] 吴细珍. 数据挖掘在电力系统中的应用[J]. 广东电力, 2004(2).

(责任编辑: 万贤贤)