

企业数据应用影响力评估模型方法研究



乐雯娇^{1,2} 李朋¹ 文俊浩² 邢镔^{1,3}

1 重庆工业大数据创新中心 重庆 400000

2 重庆大学大数据与软件学院 重庆 400000

3 工业大数据应用技术国家工程实验室 北京 100000

(wenjiao_yue@163.com)

摘要 针对企业数据利用率低、数据质量评估难等问题,考虑中国企业数据治理和应用需求,联合美国 RMDS 实验室从企业数据应用的角度,创造性地加入数据科学评估维度,提出了兼容现有主流评估模型且更满足中国企业需求的企业数据影响力评估模型(Data Impact Assessment Model, DIAM)框架。考虑到现有 DIAM 模型尚未提出具体可行的评估方式,在模型框架研究的基础上,对 DIAM 模型的评估方法和评级策略进行研究。首先采用改进的层次分析法对 DIAM 模型的 4 个层次,涵盖数据顶层设计、数据科学、数据管理三大维度的 240 个评估指标进行权重计算;然后在权重计算的基础上,研究自上而下的模型评估方法;进而提出五级的数据影响力评估等级,定义了评级结果的综合性评级调整策略。通过分析,改进的 DIAM 模型能够适用企业数据应用影响力评估,为企业进行数据治理和应用能力提供科学的评估依据。

关键词: 数据治理;数据成熟度模型;层次分析法;DIAM 模型;数据应用影响力

中图法分类号 TP319

Study on Impact Assessment Model of Enterprise Data Application

YUE Wen-jiao^{1,2}, LI Peng¹, WEN Jun-hao² and XING Bin^{1,3}

1 Chongqing Industrial Big Data Innovation Center, Chongqing 400000, China

2 School of Big Data and Software, Chongqing University, Chongqing 400000, China

3 National Engineering Laboratory of Industrial Big Data Application Technology, Beijing 100000, China

Abstract Aiming at the problems of low data utilization rate and difficult data quality assessment, considering Chinese enterprise data governance and application requirements, in conjunction with the US RMDS laboratory, from the perspective of enterprise data application, the data science assessment dimension is creatively added, and compatible with existing mainstream assessments is proposed Enterprise data impact assessment model (DIAM) framework that is based on the model and better meets the needs of Chinese enterprises. Considering that the existing DIAM model has not yet proposed a specific and feasible evaluation method, based on the model framework research, this paper studies the evaluation method and rating strategy for the DIAM model. To carry out the research, first, this paper uses an improved analytic hierarchy process to calculate the weights of 240 evaluation indicators covering the four dimensions of the DIAM model, covering the three dimensions of data top-level design, data science, and data management. Then, based on the weight calculation, it researches Top-down model evaluation method. Further, it proposes a five-level data impact evaluation level, and defines a comprehensive rating adjustment strategy for the rating results. Through analysis, the improved DIAM model can be applied to enterprise data application impact evaluation for enterprises, and provides a scientific basis for enterprise data governance and application capability.

Keywords Data governance, Data maturity model, Analytic hierarchy process, DIAM model, Data application impact

1 引言

在全球化和信息化的今天,现代企业在生产经营的过程中不可避免地要接触、产生、利用大量的数据。如何最大化地管理和使用这些数据直接关系到企事业单位的盈利和生存^[1]。中国市场在数据评估方面有着一定的局限性:

1) 数据管理成熟度评估模型多由美国实体开发,中国行

业的针对性较差,这限制了模型的评估效能;

2) 作为数据管理成熟度评估的顶级方法,很少有模型评估企业利用大数据创造价值的能力。

为了弥补这种差距,中美联合提出了 DIAM 模型。但该模型尚未提出具体可行的评估方式,不具有了一定的应用性价值。因此,在该模型和现有数据成熟度评估模型的基础上,为了提高 DIAM 模型的评定结果,重新对 DIAM 模

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:国家重点研发计划课题(2019YFB1706104)

This work was supported by the National Key Research and Development Project(2019YFB1706104).

通信作者:李朋(cheney535@163.com)

型的评定方式加以调整和修正。

本文所提出的评定方法采用基于层次分析法^[2]改进的方式对 DIAM 模型指标进行 4 个层次的权重计算。同时,模型评估指标创造性地加入了与数据应用相关的数据科学这一类别,从数据顶层设计、数据科学、数据管理三大维度入手,结合数据生命周期、影响力评价两大元素构成五个方面,全方位地对企业数据成熟度进行等级评估。最后在现有指标权重和评分标准的基础上进行测评,采用降级处理等评级方式调整确定最终评级结果。评定方法利用精细的层次化架构帮助企业数据应用过程中进行规划和管理,为企业进行数据治理和应用提供科学的决策依据。

本文第 2 节介绍了企业成熟度模型的研究现状;第 3 节给出了基于 DIAM 模型的企业数据应用影响力评估研究的方法;第 4 节对案例进行分析;最后总结全文。

2 相关工作

目前国际上关于数据能力成熟度评估方面的研究有以下几种典型模型:数据管理能力评价模型(Data Management Capability Assessment Model,DCAM)、数据管理成熟度模型(Data Management Maturity Model,DMM)、CDQ(Corporate Data Quality)、DAMA(Data Management Association)、数据管理能力成熟度评估模型(GB/T36073-2018)等。本文以 DCAM 模型和 DMM 模型为例进行分析,并最终介绍 DIAM 模型。

DCAM 模型由企业数据管理协会主导,主要分为 8 个职能域,从战略、组织、技术和操作的最佳实践等方面描述了如何成功地进行数据管理,对金融业有很大的影响力^[3]。

DMM 模型是由卡耐基梅隆大学旗下机构研究所开发,通过六大职能域实现业务部门利益与 IT 的相互匹配^[4],确保更好地运用关键数据资产来实现商业目标。

这些模型都通过一定的方法为不同行业提供了软件能力成熟度评估标准。但总体而言,这些模型大多是针对特定的领域,不具备较好的普适性,无法与企业数据完美契合。此外,目前各类经典模式对于数据应用能力的评估都缺乏系统、全面的考虑。

针对以上问题,本文提出的 DIAM 模型主要由三大维度、五个方面组成^[5]。三大维度分别为:

1)数据顶层设计:负责评估企业战略层面对数据的理解、重视程度、把控能力、规划的成熟度。

2)数据科学:负责评估企业在利用数据创造价值方面的成熟度。

3)数据管理:负责评估企业在日常管理运营过程中对数据的管理行为、管理机制的建立等细节层面的成熟度。

模型的五个方面,除了三大维度各占一个主要方面之外,还包括将三大维度联系在一起的两个元素:

1)数据生命周期:负责评估企业在数据全生命周期中每个环节的管理能力。

2)影响力评价:负责评估企业已完成的项目对于业界在数据管理、数据科学领域的声誉和贡献。

在 DIAM 模型的基础上,本文提出了适合国内企业标准的各项指标和评估方法。该评估方法从 DIAM 理论研究出发,分层次、分等级地将 DIAM 模型的所有指标进行权重划分,并最终得出评定结果,为中国市场上的企业规划提供了一

个以与数据相关为主题的全方位评估视角。

3 DIAM 模型评定

3.1 模型层级划分

由本文模型框架可知,模型的指标分为 4 个层级:第 1 层级为三大维度,即模型涉及到的主体范围;第 2 层级是三大维度的细分领域,共计 19 个板块的详细介绍;第 3 层级是在每个细分领域的既定目标,共计 113 个三级目标,分别介绍了各个领域细分部分的具体目标;第 4 层级是针对每个既定目标的评价指标,共计 240 个四级指标。该模型用以评估其距离理想状态的距离,也即成熟度。整个模型框架就通过该层级模式对企业进行数据管理成熟度的评估^[6],如图 1 所示。

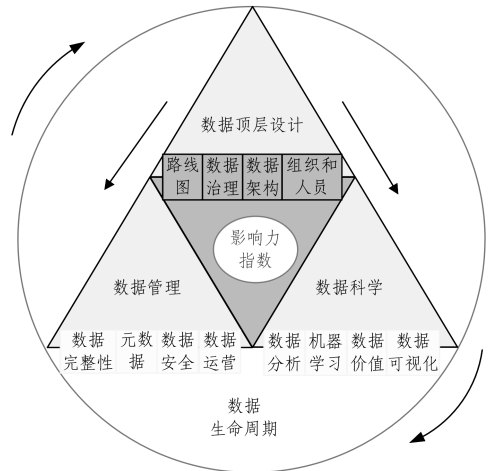


图 1 DIAM 模型框架

Fig. 1 Data impact assessment model framework

3.2 模型指标划分

整个企业数据应用影响力评估模型分为三大维度:数据顶层设计、数据科学、数据管理。其中一级指标、二级指标的具体内容如表 1 所列。

表 1 模型指标

Table 1 Model indicators

一级指标	二级指标
数据顶层设计	数据管理项目
	数据治理
	数据架构
	策略制定 路线图 人员角色和责任 组织架构
数据科学	数据分析
	机器学习
	数据价值
	数据可视化
数据管理	数据运营
	数据整合
	元数据
	数据管理策略
	数据结构
	技术支持
	数据质量 数据安全与访问

3.3 模型评定

3.3.1 指标权重

为了得到更加客观的评估方式,各项指标需要确立适宜的权重进行整体测评。本模型在结合现有成熟度模型的特点

下,将层次分析法(Analytic Hierarchy Process, AHP)进行改进,最终确定各项指标的权重^[7]。

层次分析法是由美国运筹学家萨蒂提出的一种层次权重决策分析方法^[8]。DIAM确定权重的具体步骤为:

1)确立 DIAM 模型的各项指标内容,最终得到 4 个层级,共 240 个四级指标。

2)对于本文研究的 DIAM 评估指标体系,基于常用的 9 分位比率表采取专家打分的方式得到各项指标的打分情况。

3)根据 10 位专家对各项指标的打分情况绘制出对应散点图,对每项指标的散点图进行归一化处理,并最终选定 80% 指标数在內的圆心作为当前指标的平均值。最终结果以二级指标为例,如图 2 所示。

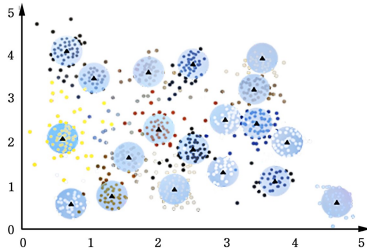


图 2 二级指标平均值

Fig. 2 Average of secondary indicators

4)专家分别对各类别指标进行两两对比,从而获得判断矩阵,得出其对应的特征向量。对特征向量进行归一化处理后,用这些归一化的特征向量近似替代其所映射的评估指标的权重。最终一级指标、二级指标的结果如图 3 所示^[9]。

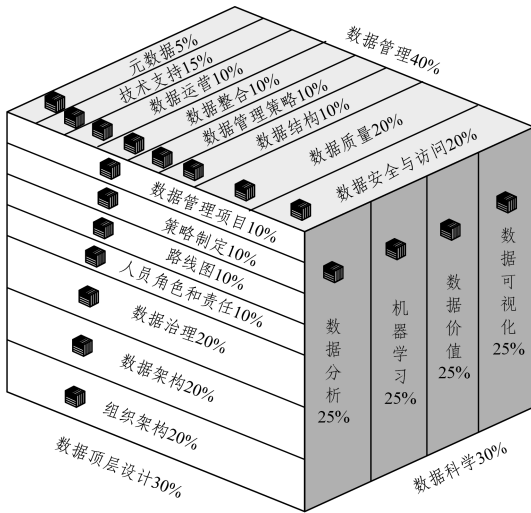


图 3 指标权重

Fig. 3 Indicator weights

3.3.2 评定方法

由于评估过程是自顶向下的,因此,DIAM 模型的评定结果只有在满足评估模型大范围的前提下,才会满足其中的小范围内容。首先计算出企业数据应用影响力评估模型的总权重,然后对其子级别的权重进行评估,具体计算过程如下。

1)评估模型总评分数

模型各级别本身均对应 100% 内的不同权重,而且权重越高越好,因此不需要对指标进行无量纲化处理,可直接应用如下公式中的加权平均模型来计算具体的评估得分情况。

$$L_1 = S * \sum_{i=1}^m \sum_{j=1}^n \sum_{k=1}^p \sum_{l=1}^q A_i B_j C_k D_l$$

其中, L_1 为企业数据应用影响力评估模型的综合评价权重, A_i 为第 i 个一级指标的权重系数, B_j 为第 i 个一级指标下的第 j 个二级指标的权重系数, C_k 为第 j 个二级指标下的第 k 个三级指标的权重系数, D_l 为该企业在第 k 个三级指标下的第 l 个四级指标的得分。 $i=1, \dots, m, j=1, \dots, n, k=1, \dots, p, l=1, \dots, q$,其中 m, n, p, q 分别为当前等级下对应的子级别个数; S 在此处设为百分制的系数,为 100。同时, S 也可根据自身需求取其他分制的系数并获得对应值。

二级、三级、四级各自的综合评分结果分别如下:

$$L_2 = S * \sum_{j=1}^n \sum_{k=1}^p \sum_{l=1}^q B_j C_k D_l$$

$$L_3 = S * \sum_{k=1}^p \sum_{l=1}^q C_k D_l$$

$$L_4 = S * \sum_{l=1}^q D_l$$

2)模型评估标准

依据上述对各级别的评估结果,本文在借鉴了国内外数据成熟度评估模型的评定标准以及《企业信息化能力成熟度研究》之后,选择 30, 60, 80, 90, 100 作为等级值,提出了企业数据应用影响力的评定结果,具体对应情况如表 2 所列。

表 2 综合分数评定结果

Table 2 Results of comprehensive score assessment

综合分数 L_1 (百分制)	评定结果
$0 < L_1 \leq 30$	一级
$30 < L_1 \leq 60$	二级
$60 < L_1 \leq 80$	三级
$80 < L_1 \leq 90$	四级
$90 < L_1 \leq 100$	五级

五级是指企业数据应用影响力各阶段都已十分完善,现阶段的各方面成熟度超过 90% 的其他公司,在企业数据应用影响力上处于领先地位,为业界标杆,整个行业权重不超过 10%。

四级是指该企业的企业数据应用影响力趋于完善,整体水平较高,但略有不足之处,在企业数据应用影响力上有比较好的借鉴意义,整个行业权重不超过 20%。

三级是指该企业的企业数据应用影响力基本完成,整体成熟度不错,但仍有较多需要改进的地方。在企业数据应用影响力上处于中间水平,整个行业大部分企业处于该区间内。

二级是指该企业的整体企业数据应用影响力不高,有很多需要完善的地方,但部分数据管理项目已经完成,处于该行业落后位置,需要加强企业数据应用影响力的完善。

一级是指该企业基本没有数据管理等方面的内容,企业数据管理十分落后,刚刚开展甚至并没有开展这方面的工作。

具体等级情况如图 4 所示^[10]。

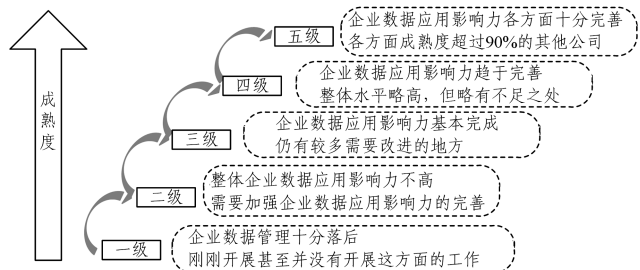


图 4 评定等级

Fig. 4 Rating

3)最终级别评定

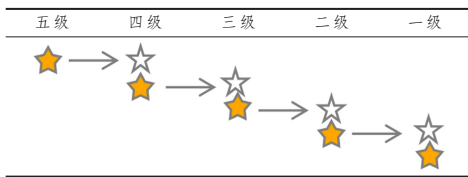
最终级别评定的主要参考对象是综合权重,但同时会在综合权重的基础上参考子级综合权重。子级的具体计分方式如下,若分数未达到60%,则进行降级处理。

Sub=L1 * w1 + L2 * w2 + L3 * w3 + L4 * w4

其中,Sub为子级综合评分的结果,L1,L2,L3,L4分别为综合评分中计算出的一级、二级、三级、四级的分数。各级权重系数由3.3.1节的权重计算流程得出,最后计算出相对应的子级分数结果。

如果在评定过程中每一项都达到了60%以上,则无须进行子级评定操作。如果其中有未达标的项目,则需要通过子级评定方式计算子级综合分数是否达标。如果达标,则仍按照综合评定结果评定级别,否则进行降级处理,如表3所列。

表3 最终评定等级 Table 3 Final rating



3.3.3 评定流程

整个企业数据应用影响力评估流程按照自顶向下的方式,首先评估企业是否已经包含了三大维度的具体内容,如果没有则直接不评定该维度的子模块,如果有再继续细分进行评定,具体评定方式参照上节。

每次评定均要对当前评定内容打分,直到评定结束,最终得出具体综合评定结果和子级评定结果。判断每个子级权重是否均超过了60%,若超过则直接对企业按照综合评估结果进行等级评定。如果有未超过60%的子级,则按照子级的计算规则算出综合子级分数是否达标,如果超过60%则企业等级不变,否则进行降级处理。具体流程如图5所示。

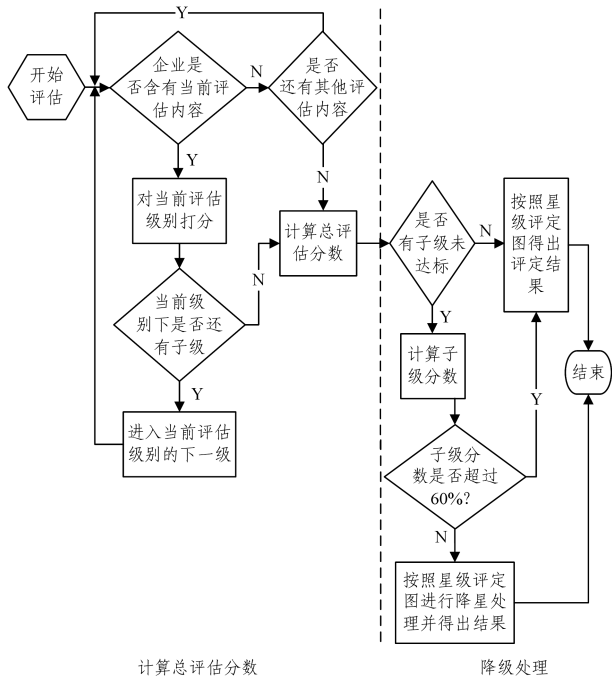


图5 具体流程

Fig.5 Specific process

4 案例分析

4.1 模型对比

我国在企业数据评估方面的模型相对较少,目前最具权威性的是国家出台的标准——数据管理能力成熟度评估模型(DCMM模型)。将DIAM模型与DCMM模型现有的层级划分以及涵盖的评估内容、评定指标和评估方式进行对比分析,具体分析结果如表4所列。

表4 对比分析

Table 4 Comparative analysis

Table with 3 columns: Assessment Item, DIAM Model, DCMM Model. Rows include Data Top-level Design, Data Management, Data Science, Hierarchical Analysis, Grade Evaluation, Downgrade Processing, and Weight Setting.

由表4的各项数据可知,在数据成熟度评估方面,本文所提出的DIAM评估模型不仅涵盖了DCMM的所有指标,而且创造性地提出了与数据应用相关的数据科学类别。与DCMM相比,本文所提出的模型更全面、针对性更强,并且增加了降级处理、权重设置等细节方面的把控,以期对企业在数据应用方面得到更全面的认识。

4.2 应用实例

为了更直观地展示企业数据应用影响力评估模型的评定结果,下面以两个企业为例进行模拟分析。

企业A为服务类公司,服务类公司更注重客户的体验感。由于A企业各个部门的员工人数众多,为了方便管理,它的运营模式和业务流程趋近成熟,公司也具备完善的人员管理和数据管理流程。同时,A企业经营时间过长,其当前使用技术较为落后,技术更新迭代的速度较慢。因此针对企业数据应用影响力模型中的三大维度而言,企业A在数据科学方面的落实十分欠缺。

企业B为制造类公司,制造类公司的技术更前沿。公司年轻人人居多,企业B对前沿技术十分感兴趣,且已经在企业中展开了较好的应用。但企业管理层面稍显欠缺,还需不断优化改进。

根据DIAM模型评估方式最终得到的结果可以发现:企业A和企业B各有所长,最终的综合评分相同,都达到了四级水平。但由于企业A的第二维度十分欠缺,企业B的第一维度、第三维度都有部分欠缺,因此两个企业均进行降级处理,最终等级结果为三级,其需要改进的地方通过生成的雷达图直观显示,用以给企业更直观的改进意见,如图6、图7所示。

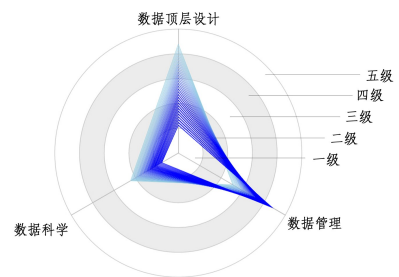


图6 企业A的三维评估图

Fig.6 Three-dimensional assessment chart of enterprise A