

基于多准则决策方法的数据资产质量评价模型

尤建新, 徐涛

(同济大学 经济与管理学院, 上海 200092)

摘要: 结合最优最劣法(best worst method, BWM)、三角模糊数和逼近理想解排序(TOPSIS)方法, 提出了一种用于评估数据资产质量的多准则决策框架。首先通过专家访谈构建评价指标体系。其次, 采用BWM方法确定评价指标的权重信息。随后, 邀请专家根据指标对数据资产质量进行语言评价, 并引入三角模糊数克服单个实数表达不完整问题。最后, 根据评价指标权重及专家评价信息, 采用TOPSIS方法对数据资产质量进行排序。通过将提出的方法应用于商业银行的数据资产质量评估, 验证方法的可行性。

关键词: 数据资产; 质量评价; 多准则决策; 最优最劣法; 逼近理想解排序法

中图分类号: C93

文献标志码: A

A Data Asset Quality Evaluation Model Based on a Multi-criteria Decision-making Method

YOU Jianxin, XU Tao

(School of Economics and Management, Tongji University, Shanghai 200092, China)

Abstract: This paper proposes a multi criteria decision making (MCDM) framework for evaluating the quality of data assets by combining the best worst method (BWM), technique for order preference by similarity to an ideal solution (TOPSIS) and triangular fuzzy number. In this framework, firstly, the evaluation indexes are constructed through expert interviews. Then, the BWM method is used to determine the weight of the evaluation indexes. After that, experts are invited to linguistically evaluate the quality of the data set based on the constructed indexes, and triangular fuzzy numbers are introduced to overcome the problem of incomplete expression of individual real numbers. Finally, the TOPSIS method is utilized to rank the quality of the data assets. The feasibility of the methodology is verified by

applying the proposed approach to data asset quality assessment of commercial banks.

Key words: data assets; quality evaluation; best worst method; multi-criteria decision-making; technique for order preference by similarity to an ideal solution

随着大数据、云计算、人工智能、区块链等技术的快速发展,以数字化、网络化、集成化、数据化、智能化为特征的数字化转型浪潮席卷全球^[1]。数据已经成为新生产要素和战略资产,为技术创新和产业升级提效增能^[2]。如何发挥数据资产价值成为政府、企业和诸多学者关注的问题。其中,数据资产质量是数据价值实现的关键,如果数据资产质量存在瑕疵还被分析、挖掘,不仅不会产生价值,甚至会出现偏差,影响企业正确决策,误导企业投资和产品研发方向^[3]。因此,评价和提升数据资产质量对发挥数据资产价值具有重要意义。

目前针对数据资产质量问题的研究仍然较少,现有研究主要关注数据资产的概念和数据资产的运营。中国信通院发布的《数据资产管理实践白皮书》将数据资产定义为由企业拥有或者控制,能够为企业带来未来经济效益,以物理或电子方式记录的数据资源,如文件、电子数据^[4]。在大数据时代,王汉生认为数据要形成产业必须有规模化的数据支撑,为了支撑规模化商业应用,数据必须是电子化的^[5]。从产业界来看,企业对于数据资产的运营场景主要包括内部使用和外部使用^[6]。关于数据资产质量研究中,孙俐丽等采用扎根理论对影响企业的数据资产质量的因素进行研究,指出数据资产质量管理的动机、标准和规范、过程管理、支撑与保障等因素对数据资产质量具有重要影响^[7]。相关研究中,数据质量问题也被诸多学者关注。从信息技术角度看,数据质量可以从精确性、完整性、一致

收稿日期: 2020-12-01

基金项目: 国家自然科学基金(71671125)

第一作者: 尤建新(1961—),男,教授,博士生导师,管理学博士,主要研究方向为管理理论与工业工程。

E-mail: yjx2256@vip.sina.com

通信作者: 徐涛(1993—),男,博士生,主要研究方向为管理理论与工业工程。E-mail: xutao0709@yeah.net



论文
拓展
介绍

性、及时性等维度来衡量^[8-9];从用户角度来看,数据质量的高低取决于用户是否可以高效地使用数据^[10-12]。

关于质量评价方法研究已经比较成熟,其中多准则决策(MCDM)方法在质量评价领域已被广泛采用。MCDM是现代决策理论的重要内容之一,该方法在比较不同方案时,根据决策问题设定相应的准则和模型进行评价^[13]。常用的MCDM方法有逼近理想解排序法(TOPSIS)^[14]、层次分析法(AHP)^[15]、决策试验与评价实验室技术(DEMATEL)^[16]、最优最劣法(BWM)^[17]等。本文将数据资产质量评价看作多准则决策问题,结合BWM、TOPSIS和三角模糊数等方法,提出一种用于评估数据资产的质量的MCDM框架。该框架中,首先通过专家访谈构建评价指标体系,并采用BWM方法通过使用1-9量表在不同指标之间进行成对比较获得权重信息。随后,邀请专家根据指标对数据资产进行语言评价,并引入三角模糊数克服单个实数表达不完整的问题;最后,根据评价指标的权重以及专家的评价信息,采用TOPSIS方法对数据资产质量进行排序。通过将提出的方法应用于某银行对其支行的数据资产质量评估,验证方法的可行性。

1 模型构建

结合BWM、TOPSIS和三角模糊数等方法,提出一种用于评估数据资产质量的MCDM框架。该框架主要包括指标体系构建、指标权重确定、专家评价以及质量排序等步骤。在确定指标权重时,采用BWM方法,该方法是对AHP方法的改进,所需的评价数据较少,可以减少专家的评价偏颇,评价数据更为准确^[17]。专家对数据资产进行质量评价时,为了克服单个实数表达不完整的问题,引入三角模糊数方法,同时解决定性指标量化问题。在质量排序时,引入TOPSIS方法评估每个方案到正理想解(PIS)和负理想解(NIS)的相对位置距离来对方案进行排序^[14]。具体如下:

步骤1:根据不同评价对象和场景构建评价指标体系。

构建评价指标是对数据资产质量进行评价的关键步骤。但不同于传统实体资产,数据资产具有规模大、流转快、类型多等特点,不同场景、不同主体对的数据资产的质量要求各不相同,数据资产的质量评价指标具有相对性和动态性。因此,数据资产质

量评价指标的确定应在特定场景下,结合文献研究、专家访谈方法进行分析构建。假设存在 n 个评价指标 $C = \{C_1, C_2, \dots, C_n\}$, l 个对指标评价的专家,用 $P_k(k=1, 2, \dots, l)$ 表示。

步骤2:评价指标比较。

专家 P_k 针对某一具体问题判断在 n 个指标中的最重要指标 C_M^k 和最不重要指标 C_L^k ;随后,使用1-9量表确定 C_M^k 对于所有指标的重要程度,重要度越高,则数值越大。同理,确定其他指标对 C_L^k 的重要程度。通过比较可获得:

$$V_M^k = (v_{M1}^k, v_{M2}^k, \dots, v_{Mn}^k) \quad (1)$$

$$V_L^k = (v_{L1}^k, v_{L2}^k, \dots, v_{Ln}^k) \quad (2)$$

式中: v_{Mj}^k 表示 P_k 评价最重要指标 C_M^k 对其他指标 $C_j^k(j=1, 2, \dots, n)$ 的重要程度,且有 $v_{MM}^k=1$; v_{jL}^k 表示 P_k 评价其他指标 $C_j^k(j=1, 2, \dots, n)$ 对最不重要指标 C_L^k 的重要程度,且有 $v_{LL}^k=1$ 。

步骤3:计算指标的最优权重。

假设基于BWM方法, P_k 评价得到的权重为 $\alpha_j^k(j=1, 2, \dots, n, k=1, 2, \dots, l)$ 。为了得到指标最优权重, $\left| \frac{\alpha_M^k}{\alpha_j^k} - v_{Mj}^k \right|$ 和 $\left| \frac{\alpha_j^k}{\alpha_L^k} - v_{jL}^k \right| j=1, 2, \dots, n$ 值应该最小化。因而,可以通过构建模型(3)求得每个指标的最优权重,即

$$\begin{aligned} & \min \xi \\ & \text{s.t.} \quad \left| \frac{\alpha_M^k}{\alpha_j^k} - v_{Mj}^k \right| \leq \xi, j=1, 2, \dots, n \quad (3) \\ & \quad \left| \frac{\alpha_j^k}{\alpha_L^k} - v_{jL}^k \right| \leq \xi, j=1, 2, \dots, n \\ & \quad \sum_{j=1}^n \alpha_j^k = 1, j=1, 2, \dots, n \\ & \quad \alpha_j^k \geq 0, j=1, 2, \dots, n \end{aligned}$$

当存在多个专家时,依据专家权重信息,聚合每个专家得到的最优指标权重如下:

$$\alpha_j = \sum_{k=1}^l \lambda_k \alpha_j^k \quad (4)$$

式中: α_j^k 是 P_k 给出的第 j 个指标的权重; λ_k 是专家 P_k 的权重。

步骤4:构造专家评价矩阵。

构造三角模糊数 $\tilde{X} = (x^L, x^M, x^U)$,其中 $x^L \leq x^M \leq x^U$, x^L, x^M, x^U 分别是 \tilde{X} 的下界元素、主元素和上界元素。当 $x^L = x^M = x^U$ 时,三角模糊数 \tilde{X} 则退化为普通的实数^[18]。假设需要评价的资产集有 m 个,记为 $A = \{A_1, A_2, \dots, A_m\}$ 。BWM方法确定的指标权重为

$\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$ 。 $\tilde{x}_{ij} = (x_{ij}^L, x_{ij}^M, x_{ij}^U)$, $i = 1, 2, \dots, m; j = 1, 2, \dots, n$ 表示第 i 个数据资产集的第 j 个指标的评价值。 P_k 对 m 个数据资产的第 n 个指标评价的矩阵如下:

$$\tilde{X}^k = (\tilde{x}_{ij}^k)_{mm} = (x_{ij}^{kL}, x_{ij}^{kM}, x_{ij}^{kU})_{mm} = \begin{bmatrix} (x_{11}^{kL}, x_{11}^{kM}, x_{11}^{kU}) & \cdots & (x_{1n}^{kL}, x_{1n}^{kM}, x_{1n}^{kU}) \\ \vdots & \ddots & \vdots \\ (x_{m1}^{kL}, x_{m1}^{kM}, x_{m1}^{kU}) & \cdots & (x_{mn}^{kL}, x_{mn}^{kM}, x_{mn}^{kU}) \end{bmatrix} \quad (5)$$

当存在多个专家时,根据专家 P_k 权重 λ_k 将 l 个专家的评价 $\tilde{X}^k = (\tilde{x}_{ij}^k)_{mm}$, $k = 1, 2, \dots, l$ 进行聚合,得到综合评价矩阵如下:

$$\tilde{X} = (\tilde{x}_{ij})_{mm} = \sum_{k=1}^l \lambda_k \tilde{x}_{ij}^k = (x_{ij}^L, x_{ij}^M, x_{ij}^U)_{mm} = \begin{bmatrix} (x_{11}^L, x_{11}^M, x_{11}^U) & \cdots & (x_{1n}^L, x_{1n}^M, x_{1n}^U) \\ \vdots & \ddots & \vdots \\ (x_{m1}^L, x_{m1}^M, x_{m1}^U) & \cdots & (x_{mn}^L, x_{mn}^M, x_{mn}^U) \end{bmatrix} \quad (6)$$

步骤5:构建加权评价矩阵。

根据 BWM 方法确定的指标权重为 $\alpha = \{\alpha_1, \alpha_2, \dots, \alpha_n\}$, 指标加权后的评价矩阵如下:

$$\tilde{R}^k = (\tilde{r}_{ij}^k)_{mm} = (r_{ij}^{kL}, r_{ij}^{kM}, r_{ij}^{kU})_{mm} = \begin{bmatrix} (\alpha_1 x_{11}^L, \alpha_1 x_{11}^M, \alpha_1 x_{11}^U) & \cdots & (\alpha_n x_{1n}^L, \alpha_n x_{1n}^M, \alpha_n x_{1n}^U) \\ \vdots & \ddots & \vdots \\ (\alpha_1 x_{m1}^L, \alpha_1 x_{m1}^M, \alpha_1 x_{m1}^U) & \cdots & (\alpha_n x_{mn}^L, \alpha_n x_{mn}^M, \alpha_n x_{mn}^U) \end{bmatrix} \quad (7)$$

步骤6:根据式(8)式(9),确定 PIS 和 NIS。

$$\tilde{Z}^+ = (\tilde{Z}_1, \tilde{Z}_2, \dots, \tilde{Z}_n) \quad (8)$$

$$\tilde{Z}^- = (\tilde{z}_1, \tilde{z}_2, \dots, \tilde{z}_n) \quad (9)$$

其中, $Z_j = \max \{\tilde{r}_{1j}, \tilde{r}_{2j}, \dots, \tilde{r}_{mj}\}$, $j = 1, 2, \dots, n$ 是指标 j 的加权指标值对应的模糊极大集。 $z_j = \min \{\tilde{r}_{1j}, \tilde{r}_{2j}, \dots, \tilde{r}_{mj}\}$, $j = 1, 2, \dots, n$ 是指标 j 的加权指标值对应的模糊极小集。

步骤7:根据式(10)、(11)计算 A_i 与 PIS、NIS 的欧式距离 D_i^+ 、 D_i^- ,再根据式(12)计算贴近度 D_i , 获得最终的质量排序, D_i 值越大,其对应的数据资产质量越高,即

$$D_i^+ = d(\tilde{r}_{ij}, \tilde{Z}^+) = \sqrt{\sum_{j=1}^n (d_{ij}^+)^2} \quad (10)$$

其中, $d_{ij}^+ = \frac{(r_{ij}^U - Z_j^U)^2 + (r_{ij}^M - Z_j^M)^2 + (r_{ij}^L - Z_j^L)^2}{3}$

$$D_i^- = d(\tilde{r}_{ij}, \tilde{Z}^-) = \sqrt{\sum_{j=1}^n (d_{ij}^-)^2} \quad (11)$$

其中, $d_{ij}^- = \frac{(r_{ij}^U - z_j^U)^2 + (r_{ij}^M - z_j^M)^2 + (r_{ij}^L - z_j^L)^2}{3}$

$$D_i = \frac{D_i^-}{D_i^+ + D_i^-}, 0 \leq D_i \leq 1, i = 1, 2, \dots, m \quad (12)$$

2 案例分析

将提出的数据资产质量评估框架应用于某银行对其支行的数据资产质量评估,并对评价结果进行讨论。

2.1 案例描述

近年来,商业银行在业务快速发展过程中,积累了客户数据、交易数据、外部数据等海量数据。数据已经成为银行的重要资产和核心竞争力。为了解数据资产质量现状,某地区商业银行分行邀请5位专家(2位研究人员和3位银行中高层管理人员)采用本文提出的评价模型,对该分行5家支行数据资产质量进行评估。

2.2 模型应用

(1)构建商业银行数据资产质量评价指标。结合文献研究、专家访谈方法,构建商业银行数据资产质量评价指标框架。指标框架构建过程中,首先对现有文献进行分析。从目前研究来看,研究者主要关注数据的相关性、一致性、准确性、完整性等指标^[8-9]。随后,通过对5位专家进行访谈并对观点归纳总结,形成数据资产质量的评价指标体系,具体指标和解释如表1所示。

表1 数据资产质量评价指标

Tab. 1 Indexes for evaluating the quality of data assets

指标	指标解释
C_1 准确性	数据与真实数据的接近程度
C_2 时效性	及时、有效地获取数据的水平
C_3 一致性	不同层次结构/数据库中同一实体的数据的一致水平
C_4 完整性	数据集中数据属性被赋值、记录的完整水平
C_5 收益性	数据获得经济效益的水平
C_6 安全性	保障数据不会流失、泄密的水平
C_7 风险性	运营数据触发道德、法律风险的可能性水平

(2)指标重要程度评价。专家基于专业知识和经验,确定在商业银行数据资产质量评价中的 C_M 和 C_L ; 随后,用1-9实数来评价 C_M 对其他指标的重要程度和其他指标对 C_L 的重要程度,结果见表2。

(3)根据 BWM 模型,利用 Lingo 软件求解出指标最优权重。邀请的5位专家权重相同,最终得到的每位专家的权重以及综合权重,结果见表3。

(4)构建指标评价语言变量与三角模糊数的对照表^[19],如表4所示。专家根据对企业调研与了解情况对5家单位的数据资产质量状况进行评价。5

表2 C_M 和 C_L 与其他指标的重要程度比较

Tab. 2 Comparison of the importance of C_M and C_L with other index

P_k	C_M	其他指标						
		C_1	C_2	C_3	C_4	C_5	C_6	C_7
P_1	C_5	5	7	5	3	1	3	6
P_2	C_4	6	4	7	1	2	5	7
P_3	C_5	6	7	8	3	1	4	6
P_4	C_4	5	7	8	1	2	6	5
P_5	C_4	6	7	7	1	2	5	6

P_k	C_L	其他指标						
		C_1	C_2	C_3	C_4	C_5	C_6	C_7
P_1	C_3	4	3	1	6	5	4	3
P_2	C_7	5	2	3	7	7	3	1
P_3	C_7	6	4	2	5	6	4	1
P_4	C_2	4	1	2	7	7	3	2
P_5	C_3	5	4	1	7	8	4	3

位专家根据构建的指标情况对5家支行的数据资产质量状况进行评价,其中风险性对应表4中的成本型指标,其他指标为效益型指标。通过将5位专家的评价聚合,可得到综合的评价表,如表5所示。

表5 综合评价

Tab. 5 Comprehensive evaluation

银行	C_1	C_2	C_3	C_4	C_5	C_6	C_7
A_1	(7.0,8.6,9.8)	(4.6,6.6,8.2)	(5.8,7.4,9.0)	(6.2,8.2,9.6)	(0.4,1.8,3.8)	(5.0,7.0,8.6)	(6.2,8.2,9.6)
A_2	(7.4,8.6,9.8)	(5.0,7.0,9.0)	(5.0,7.0,9.0)	(7.8,9.0,10.0)	(1.2,3.0,5.0)	(5.0,7.0,8.6)	(5.8,7.8,9.4)
A_3	(7.0,8.6,9.8)	(4.2,6.2,8.2)	(7.8,2.0,9.4)	(6.2,8.2,9.6)	(1.4,3.4,5.4)	(5.4,7.4,8.8)	(7.0,9.0,10.0)
A_4	(6.6,8.2,9.6)	(4.6,6.6,8.4)	(5.8,7.8,9.4)	(6.2,7.8,9.4)	(1.2,3.0,5.0)	(5.4,7.4,8.8)	(5.4,7.4,9.2)
A_5	(7.4,8.6,9.8)	(5.8,7.8,9.4)	(6.6,8.2,9.4)	(6.6,8.2,9.6)	(1.4,3.0,5.0)	(3.8,5.8,7.6)	(5.4,7.4,9.2)

(0.62, 0.76, 0.87)	(0.34, 0.50, 0.62)	(0.33, 0.42, 0.51)	(1.84, 2.44, 2.86)	(0.11, 0.53, 1.12)	(0.57, 0.80, 0.98)	(0.41, 0.54, 0.64)
(0.65, 0.76, 0.87)	(0.38, 0.53, 0.68)	(0.28, 0.39, 0.51)	(2.32, 2.68, 2.98)	(0.35, 0.89, 1.48)	(0.57, 0.80, 0.98)	(0.38, 0.52, 0.62)
(0.62, 0.76, 0.87)	(0.31, 0.47, 0.62)	(0.39, 0.46, 0.53)	(1.84, 2.44, 2.86)	(0.41, 1.00, 1.60)	(0.62, 0.85, 1.01)	(0.46, 0.60, 0.67)
(0.58, 0.72, 0.85)	(0.34, 0.50, 0.63)	(0.33, 0.44, 0.53)	(1.84, 2.32, 2.80)	(0.35, 0.89, 1.48)	(0.62, 0.85, 1.01)	(0.36, 0.49, 0.61)
(0.65, 0.76, 0.87)	(0.44, 0.59, 0.71)	(0.37, 0.46, 0.53)	(1.96, 2.44, 2.86)	(0.41, 0.89, 1.48)	(0.43, 0.66, 0.87)	(0.36, 0.49, 0.61)

(6)计算PIS和NIS, Z^+ 为PIS, Z^- 为NIS。

$$Z^+ = [(0.65, 0.76, 0.87) (0.44, 0.59, 0.71) (0.39, 0.46, 0.53) (2.32, 2.68, 2.98) (0.41, 1.00, 1.60) (0.62, 0.85, 1.01) (0.46, 0.60, 0.67)]$$

$$Z^- = [(0.58, 0.72, 0.85) (0.31, 0.47, 0.62) (0.28, 0.39, 0.51) (1.84, 2.32, 2.80) (0.11, 0.53, 1.12) (0.43, 0.66, 0.87) (0.36, 0.49, 0.61)]$$

(7)计算贴近度,并进行排序,结果如表6所示。

2.3 结果讨论

(1) 指标体系讨论

指标构建与权重确定是数据资产质量评价的关键步骤。本文在指标构建过程中,通过对观点的归纳总结,认为数据资产质量不仅包括数据质量特征,还应该体现数据作为资产的重要特性,如收益性和风险性。数据作为资产产生收益和价值已被广泛认可,如银行通过对客户数据分析,了解客户风险偏好、消费能力、信用状况等信息,帮助银行了解客户业务需求,从而帮助银行提供更精准的服务和控制风险。因此,收益性

表3 指标权重结果

Tab. 3 Index weighting results

指标	P_1	P_2	P_3	P_4	P_5	综合
C_1	0.099	0.079	0.092	0.095	0.082	0.089
C_2	0.071	0.118	0.079	0.044	0.070	0.076
C_3	0.049	0.068	0.069	0.059	0.039	0.057
C_4	0.165	0.367	0.184	0.391	0.382	0.298
C_5	0.369	0.237	0.398	0.237	0.246	0.297
C_6	0.165	0.095	0.138	0.079	0.099	0.115
C_7	0.083	0.037	0.041	0.095	0.082	0.067

表4 指标评价语言变量与三角模糊数的对照

Tab. 4 Index evaluation of linguistic variables against triangular fuzzy sets

指标评价语言变量		对应三角模糊数
效益型指标	成本型指标	
非常差	非常大	[0,0,1]
差	大	[0,1,3]
较差	较大	[1,3,5]
一般	一般	[3,5,7]
较好	较小	[5,7,9]
好	小	[7,9,10]
非常好	非常小	[9,9,10]

(5)根据指标权重,获得加权后的评价矩阵

表6 贴近度及排序

Tab. 6 Proximity and ranking

银行	D^+	D^-	D	排序
A_1	0.54	0.16	0.23	5
A_2	0.16	0.50	0.76	1
A_3	0.33	0.48	0.59	2
A_4	0.39	0.36	0.48	4
A_5	0.33	0.37	0.53	3

是数据资产质量的重要特性之一。数据资产的安全和风险也是质量评价的重要维度,比如银行在收集、存储过程中可能会丢失或者泄露客户的信息;运营过程中滥用行为可能触发法律风险,数据资产一旦触发上述

风险,将会对数据的质量和和价值产生影响。

从表3中可以看出,数据资产的完整性和收益性权重分别为0.298和0.297,占总权重的近60%。其次是数据资产的安全性和准确性指标,权重分别为0.115和0.089。从数据治理的角度,数据成为资产的首要条件是保证数据的完整性和准确性。对于商业银行而言,拥有完整、准确的数据才能对客户风险或需求进行分析,从而帮助银行控制风险或提高收入。由于银行数据中,包含大量客户隐私信息,一旦数据泄露将对客户造成严重损失。因此,完整性、收益性、安全性和准确性是数据资产的重要指标。本文所构建的数据资产质量评价指标体系也可以为相关评价问题提供指标参考。需要注意的是,不同于传统实体资产,大数据背景下的数据资产具有规模大、流转快、类型多等特点,不同场景、不同主体对数据资产的质量要求各不相同,数据资产的质量评价指标具有相对性和动态性。因此,数据资产质量评价指标的确定应在特定场景下进行。

(2) 模型有效性讨论

为进一步说明本文提出的质量评价框架的有效性和与优点,首先计算专家评价初始模糊值的期望值,并用图1表示。图1中可以看出,各家银行数据资产质量在各个指标上得分差异较小,直接采用专家评价语言方法难以有效地区分不同银行的数据资产质量。随后,采用AHP方法获得评价指标权重,并结合评价价值对质量进行排序,获得如表7所示的评价结果对比情况。从对比结果来看,本文提出的基于MCDM的数据资产质量评价框架与传统的AHP方法的评价结果整体具有一致性,也说明了本文所提出模型的有效性。但相较于专家评估和AHP方法的,本文所提出的框架操作简单易于实现,且评价结果具有较高的区分度。

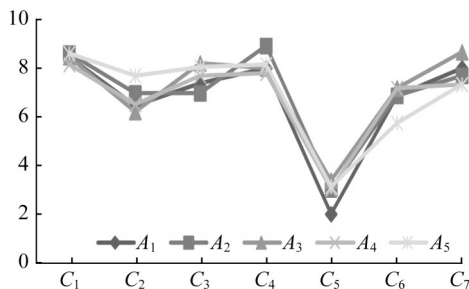


图1 各银行在各指标上的评价得分期望值

Fig.1 Expectation of banks' evaluation scores on each index

(3) 数据资产质量讨论

根据表6结果,5家银行数据资产质量排序为 $A_2 >$

表7 MCDM与AHP的质量评价结果对比

Tab.7 Comparison of the evaluation results between the MCDM-based method and AHP method

银行	基于MCDM的方法		AHP方法	
	质量得分	排序	质量得分	排序
A ₁	0.23	5	6.01	5
A ₂	0.76	1	6.58	1
A ₃	0.59	2	6.54	2
A ₄	0.48	4	6.25	4
A ₅	0.53	3	6.31	3

$A_3 > A_5 > A_4 > A_1$ 。为进一步探究数据资产质量差异原因,本文对各指标评价情况进行讨论。从各指标情况来看,目前所有银行在数据资产的准确性和完整性方面表现较好,得分情况明显高于其他指标。该得分较高的主要原因是银行业是我国较早开展数据治理的行业,2018年5月银保监会发布《银行业金融机构数据治理指引》,明确指出数据治理应该纳入公司治理的范畴,监管层的要求有效提升了商业银行的数据治理水平。在数据资产的完整性指标上,银行A2的得分高于其他银行。完整性指标的权重占比最高,使得在整体质量排序中A2要高于其他银行。相对于其他指标,所有被评价银行的资产收益性得分均较低,其中银行A1的得分情况最低。各单位在收益性指标中的得分说明尽管银行的数据管理水平较高,但如何进一步利用数据,让数据为银行产生更多的收益仍存在一定挑战。从银行业实际情况看,对于数据资产的价值发挥仍在业务数据化阶段,相关数据资产还没有形成产品以产生更多收益。从安全性指标来看,得分情况相对较低,尤其是A5在该项的得分低于平均水平,说明该银行需要进一步规范数据的收集、存储流程,确保数据的安全性。

(4) 管理建议

通过对商业银行数据资产质量的评估,得出管理建议如下:①提升数据完整性、准确性。数据的准确性和完整性是数据资产化的基础,也是影响数据资产质量的关键指标。数据存在错误或缺失,将对企业的运营决策产生不利影响,尤其是金融机构,可能会产生一定风险;②梳理数据资产及其应用场景。当前多数银行对数据资产的应用仅限于为当前业务辅助决策,从案例评价结果来看,数据资产的收益性得分仍然较低,如何发挥数据本身的价值需要企业进一步根据业务需求进行拓展;③保障数据安全。银行的相关数据资产往往涉及多方信息,包括客户隐私数据。相关数据泄露、流失也将会触及法律风险。监管机构需要银行进一步规范数据的收集和存储流程,从银行层面也需要完善相关的流程与规定,保障数据安全。

3 结语

数据已经成为新的生产要素和战略资产,为技术创新和产业升级提效增能。评价和提升数据资产质量,对发挥数据资产价值具有重要意义。本文结合BWM、TOPSIS和三角模糊数等方法,提出一种用于评估数据资产的质量的MCDM框架,为数据资产质量评价问题提供模型参考。通过案例验证与对比分析,本文提出的数据资产质量评价框架操作过程清晰、结果科学准确、易于实现,且排序结果的区分度高于传统方法。通过将提出的方法应用于商业银行的数据资产质量评估,提出商业银行的数据资产质量评价指标体系,并从数据完整准确性、梳理数据资产与应用场景和加强数据安全等角度为商业银行数据资产质量提升提供相应管理建议。

作者贡献说明:

尤建新:提出选题,设计论文框架。
徐涛:负责整理文献,开展调研,论文撰写与修订。

参考文献:

- [1] 中国信息通信研究院. 数字经济治理白皮书(2019年)[R]. 北京: 中国信息通信研究院, 2019.
China Academy of Information and Communication Research. White paper on digital economy governance (2019) [R]. Beijing: China Academy of Information and Communication Research, 2019.
- [2] MAYER-SCHONBERGER V, CUKIER K. Big data: a revolution that will transform how we live, work, and think[M]. New York: Houghton Mifflin Harcourt, 2013.
- [3] 尤建新. 大数据时代呼唤数据质量治理[J]. 上海质量, 2019(10):20.
YOU Jianxin. Calling for data quality governance in the era of big data[J]. Shanghai Quality, 2019, (10):20.
- [4] 中国信息通信研究院. 数据资产管理实践白皮书[R]. 北京: 中国信息通信研究院, 2019.
China Academy of Information and Communication Research. White paper on data asset management practices [R]. Beijing: China Academy of Information and Communication Research, 2019.
- [5] 王汉生. 数据资产论[M]. 中国人民大学出版社. 北京, 2019
WANG Hansheng. Theory of Data Assets [M]. Beijing: People's University of China Press. 2019
- [6] 虞正, 高红冰, 徐盈, 等. 数据资产化之路-数据资产的估值和行业实践[R]. 上海: 德勤&阿里研究院, 2019.
YU Zheng, GAO Hongbin, XU Ying, *et al.* The Road to data assetization-valuation of data assets and industry practices [R]. Shanghai: Deloitte & Ali Research Institute, 2019.
- [7] 孙俐丽, 吴建华, 袁勤俭. B2C企业数据资产质量影响因素研究[J]. 情报理论与实践, 2017(7):99.
SUN Lili, WU Jianhua, YUAN Qinjian. Research on factors influencing data asset quality of B2C enterprises[J]. Intelligence Theory and Practice, 2017(7):99.
- [8] OMARA E, SAID T, MOUSA M. Employing neural networks for assessment of data quality with emphasis on data completeness [J]. International Journal on Artificial Intelligence and Machine Learning, AIML, 2011, 11(1): 21.
- [9] 黄莺. 元数据质量的定量评估方法综述[J]. 图书情报工作, 2013(4):145.
HUANG Ying. A review of quantitative assessment methods for metadata quality[J]. Library Intelligence Work, 2013(4):145.
- [10] PELTIER J, ZAHAY D, LEHMANN D. Organizational learning and crm success: a model for linking organizational practices, customer data quality, and performance[J]. Journal of Interactive Marketing, 2013, 27(1): 1.
- [11] JORGE M, ISMAEL C, BIBIANO R, *et al.* A data quality in use model for big data[J]. Future Generation Computer Systems, 2016, 63(10):123.
- [12] 宋立荣. 基层科技报告资源建设中元数据质量评估研究——以中国科学技术信息研究所为例[J]. 中国科技资源导刊, 2016, 48(1):57.
SONG Lirong. Research on metadata quality assessment in the construction of grassroots science and technology reporting resources: the case of China Institute of Scientific and Technological Information [J]. China Science and Technology Resource Guide, 2016, 48(1):57.
- [13] YOON K P, HWANG C L. Multiple attribute decision making: an introduction[M]. Thousand Oaks: Sage publications, 1995.
- [14] HWANG C L, MASUD A S M. Multiple objective decision making-methods and applications: a state-of-the-art survey[M]. New York: Springer Science & Business Media, 2012.
- [15] SAATY T L. What is the analytic hierarchy process? In mathematical models for decision support[M]. Berlin/Heidelberg: Springer, 1988.
- [16] FENG C, MA R. Identification of the factors that influence service innovation in manufacturing enterprises by using the fuzzy DEMATEL method[J]. Journal of Cleaner Production, 2020, 253. DOI:10.1016/j.jclepro.2020.120002
- [17] REZAEI J. Best-worst multi-criteria decision-making method[J]. Omega, 2015, 53: 49.
- [18] LIANG D, LIU D, PEDRYCZ W, *et al.* Triangular fuzzy decision-theoretic rough sets [J]. International Journal of Approximate Reasoning, 2013, 54(8):1087.
- [19] 邵国霞, 曹政国. 基于三角模糊数-TOPSIS的既有高铁近接地基加固方案优选[J]. 铁道标准设计, 2019, 63(2):40.
SHAO Guoxia, CAO Zhengguo. Preferred solution for strengthening the proximity foundation of existing high-speed rail based on triangular fuzzy number-TOPSIS [J]. Railway Standard Design, 2019, 63(2):40.