

## 基于 FISM-ANP 新零售数据质量评价体系研究

闫鑫泉

北方民族大学数学与信息科学学院 宁夏银川

**【摘要】目的** 新零售利用数据的优势，充分挖掘其丰富的商业价值，以利用数据资产实现商业利益最大化。而数据质量的优劣是挖掘有实用价值的信息的关键所在。为了推动数据质量系统评估与规范化管理，需要建立相应的新零售数据质量评价指标体系。**方法** 本文以新零售和数据质量等相关概念为基础，通过国内外文献梳理，借鉴已有的数据质量评价体系，最终以数据固有质量、用户适用度、数据监管水平和数据技术水平 4 个数据质量评价维度和 18 个二级数据质量评价指标。通过模糊解释结构模型确定评价指标因素之间的存在的直接与间接关系，将得到的关系作为网络层次分析的网络结构的输入，最终构建评价指标体系。**结论** 在新零售数据质量评价体系中，数据准确性是决定新零售数据质量的关键因素。此外，在整体数据质量评估中，用户适用度是新零售数据质量的重要维度，因此在数据业务发展成熟阶段，应该专注于提升用户适用度。

**【关键词】** 新零售数据；质量评价；模糊解释结构模型；网络层次分析

**【基金项目】** 北方民族大学研究生创新项目“大数据背景下新零售行业数据质量评价模型的建立及实证研究”（YCX22090）

**【收稿日期】** 2023 年 10 月 20 日 **【出刊日期】** 2023 年 12 月 28 日 **【DOI】** 10.12208/j.jmba.20230007

### Research on the evaluation system of data quality in new retail based on FISM-ANP

Xinquan Yan

(School of Mathematics and Information Science, North Minzu University, Yinchuan)

**【Abstract】Objective** New retail leverages the advantages of data to fully explore its rich commercial value and maximize business benefits through the utilization of data assets. The quality of data is crucial in extracting useful and valuable information. To promote the evaluation and standardized management of data quality systems, it is necessary to establish a corresponding evaluation index system for data quality in new retail. **Method** Based on the concepts of new retail and data quality, this paper conducts a review of domestic and foreign literature and references existing data quality evaluation systems. Ultimately, it identifies four dimensions of data quality evaluation, including intrinsic data quality, user applicability, data governance level, and data technology level, along with 18 secondary data quality evaluation indicators. The fuzzy interpretive structural model is used to determine the direct and indirect relationships between the evaluation indicator factors. These relationships are then used as inputs for the network structure of the analytic hierarchy process to construct the evaluation index system. **Conclusion** In the evaluation system for data quality in new retail, data accuracy is identified as the key factor determining data quality. Additionally, user applicability is an important dimension in overall data quality assessment. Therefore, in the mature stage of data business development, efforts should be focused on enhancing user applicability.

**【Keywords】** New retail data; Quality evaluation; Fuzzy interpretive structural model; Analytic hierarchy process

### 1 引言

近年来，随着新零售时代的不断推进，其加速

推动了零售业的数字化转型和发展。为了充分利用数据价值，首先需要使用专业的方法和数据分析软

件对新零售数据进行处理和分析。如今,数据分析处理软件不断更新迭代,同时各种机器学习、深度学习算法也得到应用,并且专注于改进数据分析的方式和手段。但与此用时也需要关注新零售数据自身的质量问题。通过评估数据质量指标,可以识别和解决数据质量问题,提高数据质量,从而提供更准确、可靠和有用的信息,为业务决策提供支持和保障。此外,评价数据质量指标还可以帮助组织制定数据质量管理计划,并监测其实施效果,以确保数据质量的持续改进和优化。

为了避免企业因使用劣质数据进行分析而做出错误决策并蒙受经济损失,对新零售数据质量的研究变得尤为迫切。然而,目前研究者们对线下新零售数据质量的研究较为缺乏。因此,本文旨在构建新零售数据质量评价指标体系,以对国内新零售平台提供的数据质量进行评估,推动相关领域的研究进展。

## 2 数据质量评价体系研究视角

为更综合的找出新零售数据质量指标因素,在选择指标时,要考虑评估指标体系的完整性,需尽可能的找出数据质量及新零售行业数据质量的所有指标,为了实现这一研究目标,本文在多个文献数据库平台,按照“数据质量”“数据质量 企业零售行业”“数据质量管理”等关键字进行搜索,汇总整理了国内外数据质量的研究成果,汇总了研究范围内的指标因素如附录 1。

通过查阅大量的国内国外文献,学者们从两个方面构建数据质量评价体系,即基于数据质量的固有属性和数据质量的广泛性,以对数据质量进行综合评估。数据质量固有属性重视数据自身质量,如数据完整性、数据准确性、数据一致性、数据及时性等,而数据质量广泛性考虑数据在特定行业环境中的质量特征<sup>[1]</sup>。总体而言,可为以下角度分别是数据视角、用户视角、监管视角和技术视角。

### 2.1 数据视角

在早期的研究中,研究人员通常将数据质量仅限于准确性,即数据要么正确,要么错误。他们将数据质量视作一个单一维度的概念,定义相较狭隘。然而,随着研究的进一步深入,学者们逐渐认识类似于产品质量,数据质量也涉及多个维度,其质量受多个因素的影响,数据质量是一个多方面的概念。

因此,通过将数据质量与产品质量进行类比,学者们尝试将数据看作产品来进行质量评估。如 Wang 将数据看作产品,并提出 TDQA (Total Data Quality Assessment) 的框架,用于管理数据质量。该框架下固有质量维度下将其细分为准确性、客观性、可信性、可访问性 4 个方面<sup>[2]</sup>。Behkama 等人将数据质量分为语义准确性、句法准确性、唯一性、完整性及一致性 5 个维度来说明数据固有特征质量能有效反映真实的数据信息。刘寒基于基本的数据质量原则,提出数据质量指标包括数据完整性,准确性,一致性,精准度,时效性,唯一性,及时性,合理性,隐私性,参照完整性,有效性<sup>[3]</sup>。在这种观点下建立的数据质量评价体系着重于数据本身的内在价值,忽略了数据用户对数据质量的需求,更多的是关注数据生产者的利得。

### 2.2 用户视角

为了使得数据能够满足用户人群的需要,学者们已经开始以数据用户的视角出发,来研究数据质量的评价方法。如杨青云等学者强调了考虑用户需求的重要性,提出了一种基于用户需求的定量数据质量评价模型,以更精确地评估数据质量,确保数据满足实际应用的要求<sup>[4]</sup>。王今等构建的政府开放数据质量指标评价体系以用户满意度为基础,其中包括政府准备、公众期望、感知质量、数据易用和公众信任 5 个方面,共 14 个指标,构建出一个科学定量的数据质量评价模型<sup>[5]</sup>。王瑞云等人基于用户适应度,划分为数据内在质量维度和用户行为外在质量维度两个方面的数据质量框架,这种方法充分考虑了用户行为对数据质量的影响。其中,外在质量维度中通过考量浏览次数、下载次数、用户关注度、时段下载比和浏览下载比这五个指标,对用户行为的外在质量进行评价<sup>[6]</sup>。王力等基于数据内在质量维度下,提出提高用户满意度的质量维度包括数据可得性、数据可用性、数据可靠性、数据相关性和数据可理解性等数据质量评价维度<sup>[7]</sup>。用户视角下构建的数据质量评价体系充分考虑用户需求,旨在确保数据满足用户的实际需求,并提供有用的结果。然而,相对对于数据的价值方面,该评价体系给予的关注相对较少。

### 2.3 监管视角

孙俐丽等人在研究影响数据资产质量的支撑与

保障因素中指出，在数据质量管理方面，大多企业缺乏公司级独立的监管部门和完善的监管体系，数据质量监管体系的不完善，会导致数据质量的置信度不高，严重影响数据价值<sup>[8]</sup>。此外，张敏指出，在我国的大数据交易中，缺乏规范的交易规则，存在多种交易风险。因此，迫切需要建立相关的法律监管制度。针对该情况她提出了从主体、范围、价格和质量四个方面进行监管<sup>[9]</sup>。周艳红在构建电商大数据质量评价体系中，从监管视角提出了法律法规数量、资金投入、数据监管范围和监管部门级别跨度 4 个评价指标<sup>[10]</sup>。安小米等人提出的 HBDA 视角下的大数据评估指标框架下，针对“管理-全局视图”维度提出了包含安全性、大数据标准体系、法律法规数量和管理制度 4 个指标<sup>[11]</sup>。

### 2.4 技术视角

安小米等人提出数据技术视角指的是大数据基础设施设备、数据环境、算法和技术对数据质量的影响。在精准治理的基础上，对数据技术基础设施进行评估，特别关注数据的 IT 治理和技术融合的有效性<sup>[11]</sup>。其中在机理-技术视角总结了可恢复性、硬件可靠性、硬件可扩展性、系统易用性、存储利用实时性、合理性、可维护性等 11 个指标。辛金国等人在构建统计数据质量影响因素模型中，从流程、技术和管理三个维度分析影响数据质量的因素，其中技术维度方面包括数据库技术、数据质量检测识别和数据分析 3 个影响因素<sup>[12]</sup>。Mohammad A 等人也在研究大数据质量因素及框架中，在技术层面上提出了数据可扩展性，集成性，容错性、异构性、完整性等评估指标<sup>[13]</sup>。

## 3 研究方法

### 3.1 模糊解释结构模型

为分析复杂系统和问题美国系统科学家 James Grier Warfield 提出了一种用于分析复杂系统和决策的方法——解释结构模型 (Interpretive Structural Modeling, ISM)。ISM 的目标是协助人们理解和应对复杂问题，尤其是那些牵涉到多个因素和相互关系的问题。ISM 采用图形化和结构化的方法，以揭示问题内在结构和变量之间的相互作用。

#### 3.1.1 解释结构模型相关原理

(1) ISM 的构建步骤如下：

① 组建 ISM 专家小组，共同确定问题的因素

集  $N = \{S_i, i = 1, 2, \dots, n\}$ ；

② 由因素间的独立及依赖关系，构建出邻接布尔矩阵 A，设 A 的因素  $a_{ij}$  定义为  $A = (a_{ij})_{n \times n}$ ，其中因素间的关系由以下规则确定：

$$a_{ij} = \begin{cases} 1 & \text{因素 } S_i \text{ 与因素 } S_j \text{ 存在直接二元关系} \\ 0 & \text{因素 } S_i \text{ 与因素 } S_j \text{ 不存在直接二元关系} \end{cases}$$

③ 根据邻接布尔矩阵 A 计算得出可达矩阵 M；

④ 对上述得到可达矩阵 M 进行层级划分；

⑤ 构建元素递阶结构有向图。

(2) 可达矩阵的计算：

若有初始系统对应的原始布尔矩阵 A，为了得到可达矩阵 M，首先将单位矩阵 I 与原始布尔矩阵 A 相加，即将原始布尔矩阵 A 的对角线元素都加上 1，得到一个乘积矩阵。然后连续地将这个乘积矩阵与自身相乘，直到矩阵不再改变。此过程用以下公式表示：

$$(A + I)^{k-1} \neq (A + I)^k = (A + I)^{k+1} = M$$

除了前面提到的连续相乘的方法，可达矩阵也可以使用 Warshall 法和逐次平方法进行计算。

(3) 因素的层级划分：

可达集合 (或结果集合)：对于可达矩阵中的每个元素  $e_i$ ，在该元素所在行的数值为 1 的集合中，并记作  $R(e_i)$ 。这表示元素  $e_i$  能够到达的所有其他元素所组成的集合。

先行集合 (或前因集合)：对于可达矩阵中的每个元素  $e_i$ ，在该元素所在列的数值为 1 的集合中，并记作  $Q(e_i)$ 。这表示可以到达元素  $e_i$  的所有其他元素所组成的集合。

然后由  $R(e_i) \cap Q(e_i) = R(e_i)$  或者  $R(e_i) \cap Q(e_i) = Q(e_i)$  进行层级划分，这两种方法在层级划分时会得到不同的结果。根据问题的要求和具体情况，我们可以选择适合的方法进行因素的层级划分。

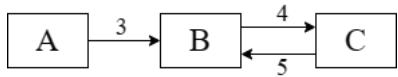
#### 3.1.2 模糊解释结构模型相关理论

FISM 与 ISM 的基础理论和主要思路是相同的，但它们在布尔矩阵和模糊邻接矩阵运算规则上存在差异。FISM 模型通过模糊邻接矩阵运算来处理关系，将模糊邻接矩阵转为布尔矩阵然后进行计算。在计算过程中，涉及到以下概念和计算方法的差别。

(1) 模糊邻接矩阵

FISM 使用模糊矩阵来表达因素之间的关系。在

模糊矩阵中，元素可以取任意介于 0 到 1 之间的实数，用于表示因素之间的模糊程度或强度，如下举例：



得到上述系统的模糊邻接矩阵 B（对其进行归

一化，以 5 为单位），其形式如下： $B = \begin{bmatrix} 0 & 0.6 & 0 \\ 0 & 0 & 0.8 \\ 0 & 1 & 0 \end{bmatrix}$

在这个系统中，因素 A 影响因素 B 的模糊程度为 0.6，而因素 B 影响因素 C 的模糊程度为 0.8。

(2) 模糊算子

在 FISM 中，使用模糊算子对模糊矩阵进行运算。模糊算子可以是最大最小算子（Zadeh 算子）、概率模糊算子、有界算子、乘积算子等或其他合适的模糊运算规则。这些算子能够处理模糊矩阵中的模糊度和不确定性。

在关于不同的模糊算子在求解模糊可达矩阵时，黄炜<sup>[4]</sup>在其研究的论文中得出一个重要结论。当在选取了 Zadeh 算子计算时，模糊可达矩阵的取值范围才不会在计算过程中无限趋向于 0 或 1。基于此，本文选择了 Zadeh 算子来计算模糊可达矩阵。

(3) 求解模糊可达矩阵及 λ 划分值域集合

通过使用 Zadeh 算子和公式 2-1，求解出模糊可达矩阵。

定义模糊矩阵  $\tilde{R} = (r_{ij})_{n \times n}$ ，并记  $R_\lambda = (r_{ij}(\lambda))$ ，其中

$$r_{ij}(\lambda) = \begin{cases} 1, & (r_{ij} \geq \lambda) \\ 0, & (r_{ij} < \lambda) \end{cases}$$

即  $R_\lambda$  为  $\tilde{R}$  的 λ 截矩阵，λ 为  $\tilde{R}$  的划分截距。

由模糊可达矩阵中 λ 的值域集合  $\lambda_i (i \in (0,1))$ ，通过 Zadeh 算子计算  $\lambda_i$  截矩阵，然后得到 i 个布尔矩阵，将模糊信息转化为了布尔形式，实现了从 FISM 到 ISM 的计算过程。

3.2 网络层次分析法

1996 年，教授 Thomas L. Saaty 提出了网络分析法（ANP），旨在克服 AHP 假设元素间相互独立的限制，与系统元素间的相互作用不符合的客观现实。ANP 用于分析具有相互交织的复杂网络结构系统，可看作是 AHP 的一个扩展。

下图（图 1）为常见的 ANP 模型。网络层次分析法的基本步骤如下所示。

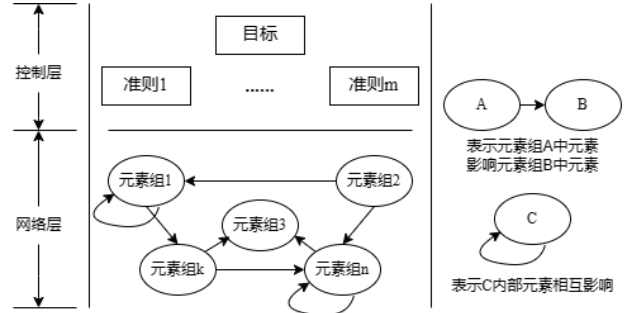


图 1 常见的 ANP 网络结构图

(1) 元素相关性分析

首先，明确决策问题的目标，将决策问题分解成不同层次，然后，通过对进行决策的问题系统分析，将其分为各个独立的元素及相对应的元素集。随后，需要评估这些分解出的不同元素层次之间的独立性，同时考虑不同元素层次之间可能存在的彼此依赖关系和元素间的反馈效应。在此过程中，可以采用相关领域的专家打分或根据经验进行 Brainstorming 等方法进行评估。

(2) 构建 ANP 网络结构

在 yaanp 软件中构建 ANP 网络结构时，需要考虑元素之间的相互作用关系。构建网络结构时，首先建立控制层，其中，需要确定决策问题的目标和相对应的各个准则，并使用层次分析法来得到每个准则对最终目标的权重。接下来在构建网络层，在网络层中对元素进行分类，形成各个元素组。在 yaanp 软件中，这些元素组可以表现为簇。然后，由第一步元素的相关性分析，将 ANP 网络层中的元素组间和元素组内部的关系进行关联，从而构建出网络层次分析的网络结构。

(3) 进行两两比较

通过元素相关性分析，确定元素及元素组之间的相互影响关系。构造元素判断矩阵和全部判断矩阵。完成矩阵构建后，再进行成对的比较，确定它们之间的相对重要性。

(4) 计算超矩阵及权重

由上一步结果，构建初始未赋权超矩阵，再通过未赋权超矩阵计算出赋权超矩阵和极限超矩阵，最终确定各元素组及元素的权重。

4 新零售数据评价指标体系构建

### 4.1 筛选数据质量指标

通过上述查阅文献，将整理出的 242 评价指标进行梳理，整合相似或者相同的数据质量评估指标，得到以下 65 个数据质量指标，如清单如下表 1。

以上 65 个数据质量指标建立初始指标清单表，为了进一步筛选出主要的新零售数据质量评价指标，对业内人士进行问卷访问调查（问卷见附录 2），参与调研的人员主要有：数据技术主管、部门项目经理、数据分析师、数据运营、程序员和数据库管理人员等，通过线上问卷发放形式，一共发放 30 份问卷，整理问卷删除 4 份无效问卷，最终收回 22 份有效问卷。对所有问卷结果整合取平均值。由整理出的结果，最终筛选出 18 个新零售数据质量评价指标，结合相关文献及新零售数据质量内涵，分别从数据固有质量、用户适用度、数据监管水平，数据技术水平 4 个维度对新零售数据质量进行描述。

对 18 个数据质量评价指标分类，得到新零售数据质量评价指标体系，包含 4 个维度，18 个指标，详见表 2。

### 4.2 基于 FISM 数据质量网络层次结构构建

#### 4.2.1 评价指标模糊邻接关系矩阵构建

组建了 FISM 工作小组，对新零售行业领域专业人士进行调研问卷（附录 3）。通过计算问卷数据的平均值，建立模糊邻接关系矩阵  $A_0$ 。

#### 4.2.2 确定模糊可达矩阵与 $\lambda$ 划分

根据 Zadeh 算子计算模糊可达矩阵，通过矩阵连乘运算得  $k=5$  时：

$$(A_0 + I)^4 \neq (A_0 + I)^5 = (A_0 + I)^6$$

即数据质量评价指标的模糊可达矩阵  $M_0$ ：

$$M_0 = (A_0 + I)^5$$

根据模糊可达矩阵  $M_0$  得出：划分  $\lambda$  的值域集合  $\lambda_i = (0.2, 0.3, 0.5, 0.6, 0.7, 0.8, 0.9)$ 。

#### 4.2.3 确定 $\lambda$ 水平截矩阵及布尔可达矩阵

根据 FISM 专家小组分析讨论，最终选取  $\lambda$  水平为 0.7 的截矩阵计算布尔可达矩阵。计算水平截

距阵  $A_{0.7}$ 。取  $\lambda = 0.7$  时， $A_{0.7} = \lambda_{d_{ij}} = \begin{cases} 0, & d_{ij} < 0.7 \\ 1, & d_{ij} \geq 0.7 \end{cases}$ ，

得到  $\lambda = 0.7$  截距阵  $A_{0.7}$ ，即得到布尔邻接矩阵，然后计算可达矩阵  $M_{\lambda=0.7}$ 。

#### 4.2.4 截矩阵层次结构

按照结果优先的方法，进行层级划分。基于可达矩阵的结果，绘制了截距  $\lambda = 0.7$  时对应的数据质量评价指标网络层次结构图。

最终得到数据质量评价指标因素的评价指标因素分层情况（表 3），数据质量评价体系被划分为五个层级，自上而下分别为第一层到第五层。各数据质量评价指标通过不同的途径对数据质量评价体系的网络结构产生影响，具体的影响关系如图 2 所示。

表 1 语义重复及相似的数据质量评价指标合并降维后的指标清单

一级指标	二级指标
数据固有质量	完整性、准确性、一致性、安全性、精准性、时效性、冗余性、数据规模、规范性、解释性、可比性、全面性、真实性、可靠性、客观性、数据到达速率、数据专业度、透明性、唯一性、语法质量、可信度、多样性、数据清洗粒度、合法性、有效性、增值性
用户适用度	有用性、可访问性、可获得性、可理解性、浏览次数、下载次数、浏览下载比、时段下载比、来源权威性、易于操作性、相关性、敏感性、及时性、数据列表现出的数据属性丰富度、用户关注度
数据监管水平	数据质量管理目标、数据质量管理机制、法律法规数量、资金投入、数据监管范围、监管部门级别跨度、共享性、可衔接性、大数据标准体系、企业管理者认识
数据技术水平	可恢复性、复杂性、硬件可靠性、硬件可扩展性、存储利用实时性、系统易用性、合理性、可维护性、安全性、数据分析质量、数据建模质量、可变性、复制性、存储格式多样性

表 2 新零售数据质量评价指标体系

一级指标	二级指标	说明
数据固有质量	完整性	完整性=（非空数据数量/数据规模）×100%
	准确性	准确性=（内容无误的数据质量/数据规模）×100%
	一致性	指同一个字段信息在不同数据集中的语义逻辑是否相同
	及时性	数据延迟时间=数据发布时间-数据产生时间，较短的延迟时间代表数据及时性，以年为时间单位
	冗余性	冗余度 =（非必要重复字段/总字段数）×100%
	可访问性	指数据在特定条件下被合法、便捷地获取和利用的程度
	可理解性	指数据的呈现形式、结构和信息内容是否易于被用户理解和解释
用户适用度	来源权威性	根据零售平台所在地区的行业排名，评估和比较各个零售平台的行业地位，将其分为 5 个段位，分别为[1,3]、[4,10]、[11,50]、[51,100]和其他，分别赋值 5、4、3、2、1 进行评估
	价值性	采用浏览下载比表示数据价值性，反映用户在浏览数据集内容选择下载的概率，代表用户根据数据集与自身需求进一步匹配选择的概率 浏览下载比=（下载次数/浏览次数）×100%
	数据属性丰富度	数据集的字段数
	易操作性	根据关联开放数据质量的五星标准，基于数据发布格式和符合标准程度，按 1、2、3、4、5 进行评估
	法律法规数量	根据零售数据及零售平台制定的相关政策法规数量
	监管规范性	指数据源的规范性，将平台、范围、定价、质量四个方面的监管情况分别用 $\gamma_1$ 、 $\gamma_2$ 、 $\gamma_3$ 、 $\gamma_4$ 表示。
	数据监管水平	数据监管范围= $((\gamma_1+\gamma_2+\gamma_3+\gamma_4)/4) \times 100\%$ ， $\gamma_i=0$ 或 1
数据技术水平	监管部门级别跨度	监管部门的级别等级值按照零售平台所在国家的行政级别由高到低依次赋值。由（最高监管部门等级-最低监管部门等级）表示监管部门级别跨度对其进行评估
	大数据标准体系	数据标准化体系数量
	可恢复性	指数据是否具备备份恢复的功能
数据技术水平	异构性	指数据结构上的差异性，分为非结构化，半结构化，结构数据，按 1、2、3 进行评估
	储存安全性	指确保数据在存储过程中免受未经授权的访问、修改、泄露、破坏或丢失的能力

表 3 评价指标因素分层情况

层级	要素
第 1 层	完整性、准确性、一致性、可访问性、易操作性、储存安全性
第 2 层	冗余性、数据属性丰富度、可恢复性
第 3 层	及时性、可理解性、来源权威性、价值性、监管规范性、异构性
第 4 层	大数据标准体系
第 5 层	法律法规数量、监管部门级别跨度

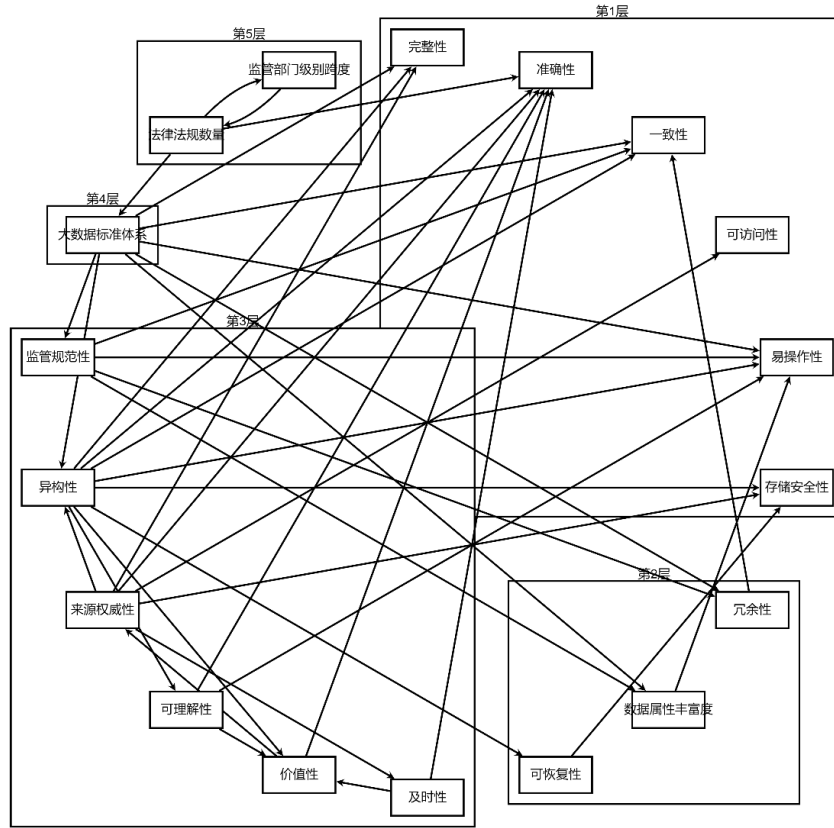


图 2 评价指标因素网络层次结构图

### 4.3 基于 FISM-ANP 的评价指标权重计算

#### 4.3.1 建立新零售数据质量的 FISM-ANP 评价指标网络结构

由 4.2 节得出的评价指标之间的关系，通过 yaanp 软件建立 ANP 网络结构图，如下图 3 所示：

#### 4.3.2 构造比较判断矩阵，计算评价指标权重

确定评价指标因素的 ANP 网络结构关系后，使用 1-9 标度法对评价指标因素两两相对重要性比较，构造判断矩阵，运用 yaanp 软件构造未赋权超矩阵、计算赋权超矩阵与极限超矩阵，最后整理得出评价指标权重结果（表 4）如下所示。

由上表 4 的重要性权重，得出基于固有属性质量、用户适用度、数据监管水平和数据技术水平四个维度，在应用新零售数据质量评价时，用户适用度在整体数据质量评价中的影响程度为 34.61%，充分体现了以人为中心，数据应与用户需求相匹配，从而数据用户提供更高的利用价值。其中，可访问性和易操作性在用户适用度中占比权重为 24.89%和 32.74%，反映出面对海量的新零售数据，用户获取和使用操作数据的难易程度是影响用户适用度质量

最关键的因素，只有当数据获取和操作变得简单、高效，才能够充分挖掘数据中蕴含的价值，实现对数据的深入分析和有效应用。

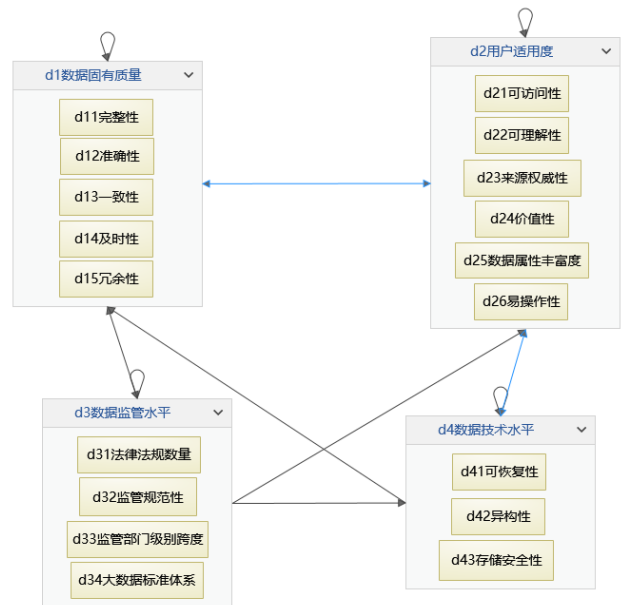


图 3 新零售数据质量 FISM-ANP 网络结构图

表 4 评价指标权重结果

一级指标维度	权重 $W_0$	二级评价指标	局部权重 $W'$	全局权重 $W$
数据固有质量 $d_1$	0.2966	$d_{11}$ 完整性	0.0751	0.0223
		$d_{12}$ 准确性	0.8645	0.2564
		$d_{13}$ 一致性	0.0476	0.0141
		$d_{14}$ 及时性	0.0045	0.0013
		$d_{15}$ 冗余性	0.0083	0.0025
用户适用度 $d_2$	0.3461	$d_{21}$ 可访问性	0.2489	0.0861
		$d_{22}$ 可理解性	0.0409	0.0141
		$d_{23}$ 来源权威性	0.1562	0.0541
		$d_{24}$ 价值性	0.0890	0.0308
		$d_{25}$ 数据属性丰富度	0.1376	0.0476
数据监管水平 $d_3$	0.2935	$d_{26}$ 易操作性	0.3274	0.1133
		$d_{31}$ 法律法规数量	0.5831	0.1711
		$d_{32}$ 监管规范性	0.0221	0.0065
		$d_{33}$ 监管部门级别跨度	0.2632	0.0772
数据技术水平 $d_4$	0.0639	$d_{34}$ 大数据标准体系	0.1316	0.0386
		$d_{41}$ 可恢复性	0.1076	0.0069
		$d_{42}$ 异构性	0.3377	0.0216
		$d_{43}$ 存储安全性	0.5547	0.0354

在数据固有性质量维度中，占其最高权重系数为数据的准确性，说明数据准确性的重要程度仍然无法被低估，对于企业来说，数据准确性对于建立和维护客户信任至关重要，并且准确的数据对于统计分析、研究和调查都至关重要。在数据监管水平方面，权重系数最高的是法律法规数量，它是新零售数据质量评价体系中的根本深层原因，是提升数据质量优劣的源动力。对于数据技术水平，数据存储安全性是该方面占比最高的影响因素，也反映出对个人、企业和组织的关键数据保护具有重要意义，可通过采取适当的物理安全措施、数据加密、访问控制和监控、定期备份和灾难恢复计划等措施，确保数据保持完整、可用、可靠和一致，从而提高数据质量。

### 5 结论

(1) 本文基于新零售数据质量，结合文献数据及问卷调研，分析整理出了影响新零售数据质量主要的评价指标，共分 4 大维度及 18 个评价指标因素，构建新零售数据质量评价指标体系。

(2) 借助 ANP 在确定权重方面考虑评价指标间的相对重要性，将 FISM 分析得出的评价指标相互影响关系作为 ANP 网络结构的输入，可以在整体上综合评估和分析指标之间的相互作用，避免了简单地将指标单独考虑的局限性，同时也能解决专家难以判断评价指标间接影响关系的问题。

### 参考文献

- [1] 刘桂锋, 聂云贝, 刘琼. 数据质量评价对象、体系、方法与技术研究进展[J]. 情报科学, 2021, 39(11): 13-20.
- [2] WANG RY. A product perspective on total data quality management[J]. Communications of the ACM, 1998, 41(2): 58-65.
- [3] 刘寒. 大数据环境下数据质量管理、评估与检测关键问题研究[D]. 吉林大学, 2019.
- [4] 杨青云, 赵培英, 杨冬青, 唐世渭, 童云海. 数据质量评估方法研究[J]. 计算机工程与应用, 2004(09): 3-4+15.
- [5] 王今, 马海群. 政府开放数据质量的用户满意度评价研



- 究[J]. 现代情报, 2016, 36(09):4-9.
- [6] 王瑞云, 贾君枝. 基于用户适用度的开放数据质量提升研究[J]. 数字图书馆论坛, 2018(12): 18-26.
- [7] 王力, 周晓剑. 大数据质量评估的标准及过程研究[J]. 经营与管理, 2018(04): 84-88.
- [8] 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, 40(07): 99-102+98.
- [9] 张敏. 交易安全视域下我国大数据交易的法律监管[J]. 情报杂志, 2017, 36(02): 127-133.
- [10] 周艳红. 电商大数据质量评价模型的建立及实证研究[D]. 重庆工商大学, 2021.
- [11] 安小米, 黄婕, 许济沧, 王丽丽, 洪学海, 王志强, 韩新伊. 全景式大数据质量评估指标框架构建研究[J]. 管理科学学报, 2023, 26(05): 138-153.
- [12] 辛金国, 张亮亮. 大数据背景下统计数据质量影响因素分析[J]. 统计与决策, 2017(19): 64-67.
- [13] Mohammad A, Mohammad M, Ahmad A, Ayman A. Big data quality factors, frameworks and challenges[J]. CompuSoft, 2020. 9(8): 3785 -3790.
- [14] 黄炜. 黑客与反黑客思维研究的方法论启示—解释结构模型新探[D]. 华南师范大学, 2003
- 版权声明:** ©2023 作者与开放获取期刊研究中心 (OAJRC) 所有。本文章按照知识共享署名许可条款发表。  
<http://creativecommons.org/licenses/by/4.0/>



**OPEN ACCESS**