Constructing a Digital Capability Evaluation Framework for Manufacturing Enterprises in the Context of Digital Economy: Based on LDA, Entropy Weight and TOPSIS Model

Yang Lu[1,2,*,a], Jinliang Wang[1,2,b]

{[a]luyang1@gdust.edu.cn, [b]wangjinliang@gdust.edu.cn}

Guangdong University of Science and Technology, No.99, Xihu Road, Nancheng District, Dongguan City, Guangdong Province, China[1]
City University of Macau, Avenida Padre Tomás Pereira Taipa, Macau SAR, China[2]

Abstract. The rapid advancement of the digital economy has significantly influenced the manufacturing sector, necessitating the development of robust digital capabilities for sustained competitiveness. This study addresses the critical gap in evaluating digital capabilities post-digital transformation in manufacturing enterprises. To avoid subjective evaluation, the research adopts text data mining and integrates Latent Dirichlet Allocation (LDA) Topic Modelling, Entropy Weight Method, and the Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) Model by Python programming to construct a comprehensive digital capability evaluation framework. The findings reveal key indicators of digital prowess in manufacturing enterprises, with a particular focus on R&D investment, IT expenditure, and smart manufacturing levels. This study contributes to the field by providing a systematic evaluation tool, aiding manufacturing enterprises in navigating the complexities of the digital economy and enhancing their digital transformation strategies.

Keywords: Digital Economy, Manufacturing Enterprises, Digital Capability, LDA Topic Modelling, Entropy Weight Method, TOPSIS Model

1 Introduction

The advent of the digital economy has ushered in a transformative era, characterized by the rapid integration of digital technologies across various sectors [1]. This paradigm shift has profound implications for manufacturing enterprises, which are increasingly compelled to adapt and evolve in this digital landscape. The significance of the digital economy cannot be overstated, as it fundamentally alters the way businesses operate, compete, and deliver value [2]. In this context, the development of digital capabilities within manufacturing enterprises emerges as a critical factor for sustaining competitiveness and fostering innovation.

However, the journey towards digital transformation is fraught with challenges, particularly in the construction of a robust digital capability evaluation framework. While numerous manufacturing enterprises are actively engaging in digitalization efforts, there is a conspicuous gap in the systematic assessment of the outcomes of these transformations [3]. The lack of a comprehensive and tailored evaluation framework impedes the ability of these enterprises to

IEDM 2024, February 23-25, Kuala Lumpur, Malaysia

DOI 10.4108/eai.23-2-2024.2345917

accurately gauge their digital capabilities, thereby hindering informed decision-making and strategic planning.

This research aims to address this gap by constructing a digital capability evaluation framework specifically designed for manufacturing enterprises in the context of the digital economy. The significance of this study lies in its potential to provide a structured and empirical approach to assess digital capabilities, thereby facilitating the strategic development of these capabilities in manufacturing enterprises. This framework is expected to serve as a valuable tool for enterprises to benchmark their digital maturity and identify areas for improvement.

The structure of this paper is as follows: Initially, the research methodology is outlined, encompassing the data collection and the use of LDA topic modelling, entropy weight method, and TOPSIS model. This is followed by the construction process of the evaluation framework and the explication of the indicators. Subsequently, the entropy weight method is employed to determine the weights of these indicators, and the TOPSIS model is applied for a comprehensive evaluation of the selected enterprises. The paper concludes with a summary of the findings and their implications for the digital transformation of manufacturing enterprises.

2 Methodology

2.1 Data collection

The data collection process is twofold. Firstly, policy texts concerning "Digital Economy" and "Intelligent Manufacturing" from central, Guangdong Province, and Dongguan City authorities were gathered. These texts, amounting to over 530,000 Chinese characters, provide a macro-level understanding of the policy environment shaping digital transformation in manufacturing. Secondly, on the micro-level, abstracts of articles from core journals in the China National Knowledge Infrastructure (CNKI) database, using "Digitalization" AND "Manufacturing" as keywords, were collected. For enterprise-level data, listed manufacturing companies in Dongguan City were selected as the study subjects. This selection is based on the availability of data and the representativeness of Dongguan's manufacturing industry as a microcosm of China's broader manufacturing sector [4].

2.2 Text analysis by LDA topic model

To explore the latent themes in policy texts and literature to establish evaluation indicators, and to reduce the subjective bias in the classification process, this paper combines the application of LDA document topic generation model with manual coding. LDA is a common unsupervised topic modeling method, aimed at uncovering hidden structures and latent semantics in documents [5]. As a three-layer Bayesian model, it encompasses a document layer, a topic layer, and a word layer, with the topic layer being the hidden layer [6]. Each document is considered a random mixture of various topics, and each topic is considered a random combination of words [7]. Hence, by estimating the probability of each topic in a document and the probability of each word in a topic, documents are classified into different clusters [8]. Therefore, based on probabilistic distributions, LDA objectively reflects the latent topics of documents, reducing the subjective errors inherent in direct manual tagging.

Subsequently, by selecting the most probable topic words and combining them with manual coding, the key terms identified by LDA can be coded into more abstract research themes. Figure 1 demonstrates the clustering process based on LDA for the construction of digital capability assessment indicators in manufacturing enterprises, which includes data collection, data preprocessing, topic modeling and clustering, visualization and manual coding.

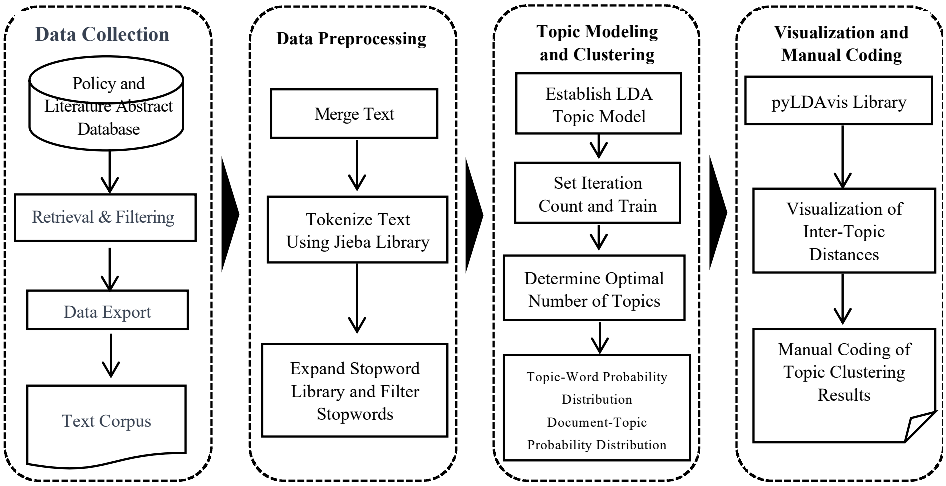

Figure 1. LDA topic model analysis flow chart

2.3 Entropy weight method

Following the identification of potential indicators, the entropy weight method is applied to determine their respective weights. This method quantifies the disorder or randomness in the system, with lower entropy indicating less uncertainty and thus a higher weight for the indicator [9]. In determining the weight of an indicator, its entropy value is pivotal. A lower entropy value indicates more significant differences in evaluation indicators across various companies, thus denoting increased importance and weight in the assessment process. The entropy weight method for weight assignment involves these steps:

Step 1: Normalize the raw data matrix, including both positive and negative indicators, before data analysis. Use these formulas for normalization:

$$y_{ij} = \frac{x_{ij} - x_{min(j)}}{x_{max(j)} - x_{min(j)}} \tag{1}$$

$$y_{ij} = \frac{x_{max(j)} - x_{ij}}{x_{max(j)} - x_{min(j)}} \tag{2}$$

Step 2: For the i-th item under the j-th indicator, compute the weight p_{ij} of the indicator value using the formula below:

$$p_{ij} = r_{ij} / \sum_{i=1}^{m} r_{ij} \tag{3}$$

(Wherein, the variance matrix of the indicators is comprised of p_{ij}.)

Step 3: Determine the entropy value e_j for the j-th indicator using the formula provided below:

$$e_j = -k \sum_{i=1}^{m} p_{ij} \cdot \ln p_{ij} \tag{4}$$

(Wherein, $k = 1/\ln m$)

Step 4: Compute the weight for the j-th indicator using the specified formula below:

$$\omega_j = (1 - e_j) / \sum_{i=1}^{n} (1 - e_j) \tag{5}$$

(In this context, 1-e_j is indicative of the redundancy in information entropy.)

2.4 TOPSIS model

The final evaluation of digital capabilities is conducted using the TOPSIS Model. This model identifies an ideal (best) and a negative-ideal (worst) solution and assesses each enterprise's relative closeness to these solutions [10]. The ideal solution is defined by the best possible attribute values across all alternatives. Conversely, the negative ideal solution signifies the least favorable option, characterized by the worst attribute values among all choices. To identify the most advantageous alternative, one must find the option nearest to the ideal solution and farthest from the negative ideal solution. The TOPSIS method is implemented through the following practical steps:

Step 1: Standardize the pre-normalized matrix to counteract differing data indicator dimensions, using this formula:

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{n} x_{ij}^2}} \tag{6}$$

Step 2: Use the standardized, maximized scoring matrix Z to determine the best and worst digital capability solutions for companies, with the highest and lowest values in each column representing these solutions, respectively.

Step 3: Determine the distances (d_i^+ and d_i^-) between the weighted, standardized evaluation vector for companies' digital capability and the identified the best and worst solutions. Use the following formula for the calculation:

$$d_i^+ = \sqrt{\sum_{j=1}^{m} \omega_j (z_j^+ - z_{ij})^2} \tag{7}$$

$$d_i^- = \sqrt{\sum_{j=1}^{m} \omega_j (z_j^- - z_{ij})^2} \tag{8}$$

Step 4: Compute the relative closeness degree (S_i) between each company's digital capability evaluation vector and the best solution. Subsequently, derive the comprehensive capability evaluation value for the companies based on this relative closeness degree. The relevant formula is as follows:

$$S_i = \frac{d_i^-}{d_i^+ + d_i^-} \tag{9}$$

3. Evaluation framework construction

3.1 Integration of text analysis results

The text analysis, comprising word frequency statistics and LDA Topic Modelling, yielded significant insights into the key themes and concepts prevalent in the discourse on digital transformation in manufacturing. For instance, terms like "manufacturing", "technology", "innovation", and "data" were frequently mentioned, indicating their centrality in the digital economy context, as depicted by Table 1. LDA Topic Modelling further distilled these themes into coherent topics, revealing the multifaceted nature of digital transformation, encompassing aspects such as smart manufacturing, technological innovation, and environmental sustainability, as shown in Table 2.

Table 1. Partial demonstration of word frequency statistics

Word	Frequency	Word	Frequency	Word	Frequency
Manufacturing	2136	Technology	2044	Innovation	1670
Data	1436	Industry	1240	Enhancement	1103
Intelligence	1008	Advancement	989	Construction	902

Table 2. Partial demonstration of LDA topic modelling results

Topic No.	Topic keywords and weight
1	0.021*"Manufacturing" + 0.018*"Technology" + 0.016*"Employment" + 0.013*"Aviation" + 0.012*"Machine Tool" + 0.012*"Aircraft" + 0.011*"Labor" + 0.011*"Processing" + 0.010*"07" + 0.009*"Society"
2	0.037*"Data" + 0.022*"Service" + 0.020*"Construction" + 0.014*"Support" + 0.012*"Platform" + 0.011*"Advancement" + 0.010*"Intelligence" + 0.010*"Province" + 0.009*"Region" + 0.008*"Resources"
3	0.026*"Green" + 0.021*"Quality" + 0.012*"Management" + 0.011*"Advancement" + 0.011*"Level" + 0.009*"Enhancement" + 0.009*"System" + 0.008*"Service Industry" + 0.008*"High Quality" + 0.007*"Integration"
4	0.040*"Technology" + 0.032*"Global" + 0.023*"Value Chain" + 0.015*"Industrial Chain" + 0.015*"Supply Chain" + 0.011*"Product" + 0.008*"Design" + 0.008*"International" + 0.007*"Status" + 0.007*"Mold"

Subsequently, through manual coding of the results from word frequency statistics and LDA topic modeling, we further constructed specific evaluation indicators. Keywords such as "Technology," "Innovation," and "R&D" constitute the R&D Investment indicator, reflecting the degree of emphasis a company places on technological innovation. Keywords like "Data," "Informatization," and "Network" form the IT Expenditure indicator, measuring the company's investment in digital infrastructure. Terms such as "Intelligence," "Intelligentization," and "Automation" make up the Smart Manufacturing Level indicator, reflecting progress in production automation and intelligence. Keywords "Product," "Design," and "Innovation" create the Product Innovation indicator, assessing the company’s capability in launching new products and services. Terms like "Management," "System," and "Process" comprise the Digital Management Level indicator, evaluating the level of digitalization in a company's internal management. Lastly, keywords such as "Green," "High Quality," and "Environment" form the Environmental Sustainability indicator, reflecting the company's efforts in achieving sustainable development.

3.2 Development of the evaluation framework

The evaluation framework is developed by aligning the themes and topics identified through text analysis with measurable indicators. This alignment ensures that the framework is grounded in the empirical realities of the digital economy and reflects the specific challenges and opportunities faced by manufacturing enterprises. The framework comprises primary indicators, each broken down into secondary indicators, as detailed in Table 3. This evaluation framework represents a systematic approach to quantifying digital capabilities in manufacturing enterprises. It is a direct outcome of a rigorous analysis of relevant texts, ensuring that the framework is both comprehensive and contextually relevant. This framework is expected to serve as a valuable tool for enterprises and policymakers alike, aiding in the assessment and enhancement of digital capabilities in the manufacturing sector.

4. Empirical analysis

4.1 Indicator weights determination using entropy weight method

The first step in the empirical analysis was to determine the weights of the digital capability indicators using the entropy weight method. This method, which objectively reflects the information provided by each indicator, was applied to both primary and secondary indicators. Table 3 presents the calculated weights. The weight allocation highlights the relative importance of various aspects of digital capability, such as R&D investment and environmental sustainability. Notably, IT Expenditure and Smart Manufacturing Level emerged as significant factors, emphasizing the critical role of technological investments and intelligent manufacturing practices in the digital transformation of manufacturing enterprises.

Table 3. Indicator weights based on entropy weight method

Primary Indicator	Primary Indicator Weight	Secondary Indicator	Secondary Indicator Weight
R&D Investment	0.2025	R&D Expenditure (CNY)	0.0597
		R&D Expenditure as % of Operating Income (%)	0.0626
		Number of R&D Personnel	0.0491
		% of R&D Personnel	0.0311
IT Expenditure	0.2120	Digital Equipment Expenditure (CNY)	0.0913
		Software Purchase Expenditure (CNY)	0.1206
Smart Manufacturing Level	0.1848	Mentions of Smart Manufacturing in Annual Reports	0.1158
		Production Growth Rate (%)	0.0690
Product Innovation	0.0688	Sales Growth Rate of New Products/Processes	0.0199
		Patents/Inventions Filed or Granted	0.0488
Digital	0.1798	Mentions of Digitalization in Annual Reports	0.0457

Management Level		Inventory Turnover Rate (%)	0.0521
		Days of Inventory Turnover (Days)	0.0377
		Working Capital Turnover Rate (%)	0.0443
Environmental Sustainability	0.1522	Mentions of Environmental Protection in Annual Reports	0.0276
		Environmental Protection Tax (CNY)	0.1246

4.2 Analysis of the TOPSIS model results

The TOPSIS model, an effective decision-making tool that identifies solutions closest to the ideal, was then applied. The best vector solution was identified as $X^{+} = (0.7598, 0.5799, 0.5068, 0.4293, 0.7876, 0.4830, 0.8516, 0.8701, 0.4417, 0.5625, 0.3963, 0.6046, 0.4053, 0.3779, 0.6046, 0.1602)$. This solution represents the optimal digital capability profile, serving as a benchmark for the evaluated enterprises.

Using the TOPSIS model, a comprehensive evaluation ranking of Dongguan's listed manufacturing enterprises was derived. Table 4 shows the overall indicator rankings. The ranking provides a view of each enterprise's strengths and weaknesses in digital capabilities. Enterprises like Shengyi Technology and Yiheda demonstrate strong performance across multiple indicators, indicating a well-rounded digital capability.

Table 4. Evaluation ranking of Dongguan's manufacturing enterprises based on TOPSIS model

Overall Ranking	1	2	3	4	5
Manufacturing Enterprise	Shengyi Technology	Yiheda	Optorun	Muse Share	Jianlang Hardware

The empirical analysis underscores the heterogeneity in digital capabilities among manufacturing enterprises in Dongguan. The application of the entropy weight and TOPSIS model provides a comprehensive and objective assessment, essential for enterprises seeking to benchmark and enhance their digital transformation strategies. This analysis not only facilitates a deeper understanding of the current state of digital capabilities in manufacturing but also aids in identifying key areas for improvement and investment.

5. Conclusion and implications

This research significantly enhances the understanding of digital capabilities in manufacturing enterprises. By providing a detailed evaluation framework, it helps enterprises identify their strengths and areas needing improvement, thus supporting strategic decision-making in the rapidly changing digital economy.

The findings from this study have several implications for manufacturing enterprises. Firstly, the importance of R&D Investment and IT Expenditure in the framework suggests that companies should prioritize technological innovation and the development of digital infrastructure. This strategic focus is essential for enhancing competitive advantage in the digital economy. Secondly, the results indicate that smart manufacturing practices are key drivers of digital capability. Therefore, enterprises should incorporate smart technologies like IoT, AI, and automation to streamline production and increase efficiency. Additionally, the study highlights the importance of enhancing digital management capabilities, implying that enterprises should digitalize their management processes, including the adoption of digital tools for inventory management, data analysis, and decision-making. Lastly, the inclusion of environmental sustainability as a primary indicator reflects the increasing importance of sustainable practices in the manufacturing sector, aligning with the global sustainability trends.

For future research, there is an opportunity to expand the study's scope to include a wider variety of manufacturing enterprises across different regions. Longitudinal studies would also be valuable in understanding how digital capabilities evolve over time due to technological advancements and policy changes.

Acknowledgments. This work was supported by Guangdong University of Science and Technology 2023 Higher Education Teaching Reform Project: Research on the Integration of Cross-Border E-commerce Practice Course and AIGC Applications (GKZLGC2023090) and Guangdong University of Science and Technology 2023 Research Project: Construction of a Digital Capability Evaluation System for Dongguan's Manufacturing Industry and Research on Performance Enhancement (GKY-2023KYYBW-6).

References

[1] Wehartaty T, Ellitan L, Wehartaty T, et al. Building a digital workplace in the era of industry 4.0 and the digital economy. World J Adv Res Rev 2023; 18: 207–217.
[2] Peters MA. Digital trade, digital economy and the digital economy partnership agreement (DEPA). Educ Philos Theory 2023; 55: 747–755.
[3] Meng F, Wang W. The impact of digitalization on enterprise value creation: An empirical analysis of Chinese manufacturing enterprises. J Innov Knowl 2023; 8: 100385.
[4] He H, Zuo Z. Supply chain concentration and enterprise financialization: Evidence from listed companies in China’s manufacturing industry. PLOS ONE 2023; 18: e0285308.
[5] García-Pablos A, Cuadros M, Rigau G. W2VLDA: Almost unsupervised system for Aspect Based Sentiment Analysis. Expert Syst Appl 2018; 91: 127–137.
[6] Blei DM, Ng AY, Jordan MI, et al. Latent Dirichlet Allocation. J Mach Learn Res 2003; 3: 993–1022.
[7] Blei D, Carin L, Dunson D. Probabilistic Topic Models. IEEE Signal Process Mag 2010; 27: 55–65.
[8] Marshall P. A Latent Allocation Model for Brand Awareness and Mindset Metrics. Int J Mark Res 2022; 64: 526–540.
[9] Shannon CE. A mathematical theory of communication. Bell Syst Tech J 1948; 27: 379–423.
[10] Olson DL. Comparison of weights in TOPSIS models. Math Comput Model 2004; 40: 721–727.