



Digital Innovation and Business Models

Empirical Research Based on Text Mining of Annual Reports of 280 Listed Companies

Simin Zeng ^{1,a}, Weihong Xie ^{2,b}, Tiepeng Peng ^{3,c*}, Haorui Cheng ^{4,d}

¹*School of Economics and Commerce. Guangdong University of Technology, Guangdong Manufacturing Big Data Innovation Research Center, Guangdong Provincial Key Think Tank Guangzhou, China*

²*School of Economics and Commerce. Guangdong University of Technology, Guangdong Manufacturing Big Data Innovation Research Center, Guangdong Provincial Key Think Tank Guangzhou, China*

³*General Committee Office. Guangdong University of Technology, GDUT Guangzhou, China*

⁴*School of Economics and Commerce. Guangdong University of Technology, Digital Economy and Data Governance Laboratory, Guangdong Provincial Key Laboratory of Philosophy and Social Sciences Guangzhou, China*

^a*13167426208@163.com*

^b*940295712@qq.com*

^{c*}*Corresponding Author: ^c121723788@qq.com*

^d*947312320@qq.com*

Abstract

The era of the digital economy has promoted a new round of globalization driven by digital technology, and digital innovation will affect the business model of enterprises in all aspects. This paper uses computer language technology to mine the annual report texts of 280 listed companies in the information transmission, software and information technology service industries in Shenzhen and Shanghai, and conducts an empirical analysis of the association rule Apriori algorithm for digital innovation and business models. Two indicators, descriptive digital innovation and true digital innovation, are used to comprehensively evaluate the level of digital innovation of enterprises. It also portrays the business model of the enterprise from three dimensions: value creation, value transmission and value realization. The empirical results show that: digital innovation is an important factor of the business model; the value creation of the business model is significantly negatively correlated with the digital innovation; the value transmission and value realization of the business model is significantly positively correlated with the digital innovation.

Keywords-*digital innovation; business model; text mining; empirical research*

1. INTRODUCTION

The arrival of the fourth industrial revolution driven by digital technologies such as artificial intelligence and big data has brought a new round of major changes to the business model of companies, and companies are facing digital transformation and innovation. Digital innovation is a continuous activity. In this process, the value creation mechanism will suddenly reach the destructive boundary, which is accompanied by the emergence of new business models ^[1]. Although there are many factors that affect business models, such as corporate culture, market environment, executive support, organizational big data capabilities, corporate innovation capabilities, etc. ^{[2][3][4][5]}. But there are few studies on the direct relationship between digital innovation and business

models. To this end, this article starts with the method of mining industry enterprise data and text, exploring whether digital innovation is an important factor in the business model, and aims to provide theoretical and practical guidance for the optimization of the business model of the enterprise. The article is divided into four parts: (1) A review of the literature on digital innovation and business models; (2) A hypothesis based on the review, and an overall description of the research data source and indicator design; (3) The association rule Apriori algorithm and regression analysis crawled through python Data; four, draw conclusions.

2. THEORETICAL BACKGROUND

Digital innovation refers to the use and recombination of digital resources by different subjects to innovate new products, services, processes, business models and other digital or non-digital results [6]. Open innovation in the digital context can increase the growth performance of enterprises by influencing the business model innovation of enterprises [7]. From the perspective of output, the measurement indicators of digital innovation include the number of digital patents submitted by enterprises and the number of new products based on digital technology [8]. From a practical perspective, digital innovation can also be measured from two dimensions: descriptive innovation and authenticity innovation; descriptive innovation can positively predict future authenticity innovations to a certain extent, authenticity innovation is mainly measured by digital patent counts Data, and descriptive innovation is measured by the keyword count of text analysis [9]. After the emergence of the Internet, the business model has attracted attention as an emerging concept. Business models are more to explore the process of value creation rather than just to prove the existence of value [10]. Different scholars have put forward many different definitions based on different research purposes and perspectives on business models. This has also led to the development of business models. The process of being conceptualized tends to be diversified. Business models include key resources and processes, profit models, and customer value propositions [11]. The iceberg theory of business models has contributed well to domestic methods of business model evaluation [12]. The business model has four basic elements: value proposition, value production, value submission, and value acquisition [13]. It is also a strategic realization process from resource element input to product service. The practicality, value creation, and profitability of the business model will ultimately manifest in the form of a value chain.

3. METHOD

3.1. Data Preparation

This article takes 280 listed companies in the information transmission, software and information technology service industries in the Shanghai and Shenzhen stock markets as the research objects. The data includes: (1) Financial data and corporate governance data, (2) The crawler collects relevant indicators of digital innovation. All the data in this article are crawled from the 2019 annual reports of the sample companies. First of all, identify keywords about the Internet business model: digital technology, intelligent manufacturing, big data, artificial intelligence, cloud computing, platform, social media, mobile technology, digital drive, informatization, etc. Secondly, according to the

description information of these keywords in the annual report, the frequency of keywords multiplied by 100 is the ratio of the total number of words in the annual report as a measure of the company's descriptive digital innovation. Descriptive innovation is not only considered to be significantly positively correlated with the current real innovation activity level, but also can positively predict the future real activity innovation level and earnings sustainability [9].

Data processing: (1) The measurement of the level of digital patents, because each company has a wide variety of patents, the indistinguishable digital patent level and sample missing are replaced by 0. (2) Excluding the sample companies with severe data missing, there are 307 listed companies in the information transmission, software and information technology service industries in the original Shenzhen and Shanghai stock markets, and 280 remain after screening.

3.2. Propose hypotheses

Digital innovation is the endogenous source of power for companies to gain new competitive advantages on the basis of maintaining their original advantages. It can not only revitalize the adaptability of the organization, but also help improve organizational performance. Digital innovation is defined in this article as descriptive digital innovation and true digital innovation. Descriptive innovation is more about predicting the true innovation level of the enterprise in the future. Based on the iceberg theory of business model, this article evaluates the business model of an enterprise from three parts: creation, delivery and realization of business model value. Based on the above-mentioned literature review and definition, the following hypotheses are proposed.

H1: Digital innovation is an important factor in the business model;

H2: The business model of an enterprise has a significant impact on the development of digital innovation.

3.3. Indicator design

Digital innovation includes descriptive innovation and the true digital innovation level of enterprises. Descriptive digital innovation uses a large sample analysis of key word frequency. Traditional word frequency analysis is currently the most commonly used text indicator construction method, which is understandable and easy to copy. The characteristics have been widely concerned and applied. Through literature review, expert interviews, and team discussions, a set of ten measurable words about descriptive digital innovation is finally obtained. The true digital innovation level of a company is measured by the company's digital R&D investment ratio and digital-related patent level

counts. This article believes that the true digital innovation level of a company is equally important as descriptive digital innovation. Therefore, the data processing in the subsequent analysis is based on 0.5 the weights are integrated.

TABLE 1. INDEX DESCRIPTION

	LEVEL 1	LEVEL 2	DESCRIPTION
DIGITAL INNOVATION	Descr.	Related word set frequency	100 times the proportion of the total word frequency of the descriptive digital innovation keyword set to the total number of words in the financial report
	Real	Digital R&D investment ratio	Digital investment R&D expenses divided by operating income
		Digital patent level	Natural logarithm of the number of patents based on digital technology plus 1
BUSINESS MODEL	VC.	Current ratio	Ratio of current assets to current liabilities
		Debt protection ratio	The ratio of net cash flow from operating activities to total liabilities
		Debt-to-equity ratio	Ratio of owners' equity to total liabilities
	VT.	Inventory turnover	Ratio of cost of sales to inventory
		Accounts Receivable Turnover Rate	Ratio of sales revenue to accounts receivable
		Turnover rate of total assets	Ratio of sales revenue to average total assets
	VR.	Year-on-year growth rate of operating income	The ratio of the increase in the current operating income of the enterprise to the total operating income of the previous year
		Net profit growth rate	The ratio of net profit growth to last year's net profit
		OPE	The ratio of gross profit after deducting various operating expenses to sales revenue

The three dimensions of business model measurement: value creation, value delivery and value realization. Value creation is divided into three indicators: current ratio, debt protection ratio, and equity-to-debt ratio. It measures the company's working capital level, default risk, and financial leverage. The higher the value creation level, the more adequate the company's working capital, the smaller the default risk, and the development The greater the potential and the higher the long-term debt repayment ability, the smaller the business and production risks it bears. Value transfer is composed of the turnover rate of three indicators of inventory, accounts receivable and total assets. It measures the company's sales capacity, inventory turnover speed, realization speed and efficiency, and asset ratio. The higher the value transfer level, the company The inventory management efficiency is high, the bad debt rate is low, and the enterprise asset utilization level is higher. Value realization is composed of three indicators: the year-on-year growth rate of operating income, the growth rate of net profit, and the profit rate of main business. It measures the degree of change in business income, the growth rate of net profit and the profitability of operating activities during the fiscal year; the level of value realization The higher the market share, the faster the company's market share expansion, the more optimistic development prospects, and the stronger the profitability and management capabilities of the company. This article believes that the three dimensions of an enterprise's business model are equally important, so if subsequent data analysis is required, the data will be integrated with a weight of 1/3.

4. FINDINGS

4.1. Descriptive statistics

As shown in the descriptive statistics of each variable in Table 2, it is found that the standard deviation of accounts receivable is the largest at 10.2994, followed by the standard deviation of today's profit growth rate of 9.2994. This shows that the bad debt rate and profit creation ability of various information transmission, software and information technology service companies are quite different. The current ratio and debt-to-equity ratio above 4 indicate that there is a certain gap in the value creation capabilities of various companies in this industry. The standard deviation of inventory turnover rate above 1 also shows that the operating capabilities of various enterprises are also different. The total asset turnover rate, the year-on-year growth rate of operating income, and the standard deviation of the main business profit rate are all smaller than 1, indicating that the industry's asset utilization efficiency is high, and the growth prospects are more optimistic due to the rapid market share. The profitability of the company's main business is also relatively even.

TABLE 2. DESCRIPTIVE STATISTICS OF INDICATORS

NAME	MIN	MAX	MEAN	S
RELATED WORD SET FREQUENCY	0.0027	1.2705	0.2354	0.1888
DIGITAL R&D INVESTMENT RATIO	0.0000	0.7000	0.1155	0.0953
DIGITAL PATENT LEVEL	0.0000	4.0000	/	/
CURRENT RATIO	0.0212	44.2734	3.6767	4.6549
DEBT PROTECTION RATIO	-1.0177	10.4023	0.2561	0.7220
DEBT-TO-EQUITY RATIO	-0.7153	43.2734	3.5109	4.9214
INVENTORY TURNOVER	0.0000	10.1295	3.1948 5	1.5254
ACCOUNTS RECEIVABLE TURNOVER RATE	0.0002	105.768 0	5.6553	10.621 8
TURNOVER RATE OF TOTAL ASSETS	0.0001	11.3449	0.5439	0.7214
YEAR-ON-YEAR GROWTH RATE OF OPERATING INCOME	-0.9854	3.8410	0.1287	0.4325
NET PROFIT GROWTH RATE	-40.6809	131.527 4	- 0.2891	9.2994
OPE	-1.0636	0.9989	0.4263	0.2898

4.2. Association rule analysis

Association rules refer to the interdependence between data objects, and the task of discovering rules is to discover from the database those strong rules whose confidence (Conk-dente) and support (Support) are greater than a given value. The association rules found in the database have been studied the most in recent years. At present, it has developed from the discovery of association rules at a single concept level to the discovery of association rules at multiple concept levels. The continuous deepening at the conceptual level makes the information provided by the association rules of Faguan more and more specific. In fact, this is a process of gradually deepening the knowledge found. The following three indicators are generally used to measure relevance:

- Support: Indicates the probability that the itemset {X, Y} appears in the total itemset. Represents the probability that A and B occur simultaneously in the total number I, the formula where I represents the total transaction set. num() represents the number of times a specific item set appears in the transaction set.

$$\text{Support}(X \rightarrow Y) = \frac{P(X,Y)}{P(I)} = \frac{P(X \cap Y)}{P(I)} = \frac{\text{num}(X \cap Y)}{\text{num}(I)} \quad (1)$$

- Confidence: Represents the probability that Y is derived from the association rule "X→Y" when the prerequisite X occurs. Indicates the probability that Y will occur at the same time in the item set where X occurs, that is, the proportion of the number of X and Y occurring at the same time in the number of only X.

$$\text{Confident}(X \rightarrow Y) = P(Y|X) = \frac{P(X,Y)}{P(X)} = \frac{P(X \cap Y)}{P(X)} \quad (2)$$

- Lift: The degree of lift represents the ratio of the probability of including X and Y at the same time to the probability of only looking at Y. The promotion degree reflects the correlation between X and Y in the association rules. The promotion degree > 1 and higher indicates the higher the positive correlation, the promotion degree <1 and the lower indicates the higher the negative correlation, and the promotion degree =1 indicates no correlation Sex, that is, independent of each other.

$$\text{Lift}(X \rightarrow Y) = \frac{P(Y|X)}{P(Y)} \quad (3)$$

Set the digital innovation and business model as two frequent item sets respectively. Among them, those with a business model evaluation level less than 1 are classified as "0", those with greater than 1 are classified as "1", and those with digital innovation greater than 0.06 are classified as "0". 1", those less than 0.06 are classified as category "0". The former is because of "digital innovation" and the consequence is "business model". The Apriori algorithm of association rules is run with python to get the results shown in Table 3.

TABLE 3. ASSOCIATION RULES APRIORI ALGORITHM ANALYSIS RESULTS

Antecedents	Consequents	Support	Confidence	Lift
frozenset({0})	frozenset({1})	0.7179	0.8204	1.0120
frozenset({1})	frozenset({0})	0.7179	0.8855	1.0120

The data shows that when the level of digital innovation is high, it is more likely that the business model is not particularly good or mature, but it does not mean that digital innovation is not one of the important factors of the business model; because both digital innovation and business model are frequent The support of itemsets is above 71%, which also shows to some extent that digital innovation is an important factor in the business model. There may be two reasons for the results presented above: a. When the company is at a high level of digital innovation, the business model structure is changed, and traditional business model evaluation indicators have a certain practical gap with new business model innovation; b. Digitalization The input-output cycle of innovation is longer than that of traditional ones, and business model changes may lag behind. Therefore, limited by cross-sectional data, the results presented by correlation analysis can only be partially referenced.

4.3. regression analysis

In the relationship between digital innovation and business models, regression results show that R^2 is about 0.6335. The value transmission and value realization of the business model have a significant positive correlation with digital innovation, that is: if the value transmission and realization process of an

enterprise is more efficient, the context of its digital innovation will be able to promote digital innovation more, or in other words, the more digital innovation. For high-end enterprises, the value creation and delivery of their business models can be smoother. The negative correlation between value creation and digital innovation shows that the greater the investment and application of digital innovation, it will not only change the capital structure of the enterprise to a certain extent, but may also increase the risks faced by the enterprise.

TABLE 4. REGRESSION ANALYSIS RESULTS

Var.	VC.	VT.	VR.	Intercept
Re. Co.	-7.9679	5.0517	3.0644	3.5218

5. CONCLUSION

This article uses text mining to explore the relationship between digital innovation and business models. The results show that: digital innovation is an important factor in the business model; there is a strong positive correlation between the value transmission and value realization of the business model and digital innovation. The value creation of a business model is mainly composed of internal elements, including the core capabilities of the company's daily operations, capital conditions, etc.; it is more related to the cost structure of the company, and a series of business capabilities such as products and services to meet target customers. Due to the lag period in the specific application after digital innovation, it tends to be reflected in the way consumers accept products or services, that is, the value transmission of the business model, and the ability of enterprises to gain more profit points and profit methods. Realize the value of the business model. The main theoretical contribution of this research lies in: taking information transmission, software and information technology service industries as research samples, using empirical methods to support the important conclusion that digital innovation is an important factor in business models. Enterprises can optimize their business model structure and obtain business model innovation by planning the relevant investment and R&D degree of digital innovation.

Related fund projects:

1. National Natural Science Foundation of China, "Research on the Influencing Factors and Behavioral Effects of Internet Privacy Concerns in the Context of Big Data: Based on Multidimensional Development Theory", 71672043

2. The Natural Science Foundation of Guangdong Province-General Project, "Research on the Mechanism of the Impact of Digital Innovation on the Performance of Manufacturing Enterprises: Based on the Perspective

of Restructuring Innovation and Resource Orchestration", 2020A1515010971

REFERENCES

- [1] Skog D A, Wimelius H, Sandberg J. Digital disruption[J]. *Business & Information Systems Engineering*, 2018, 60(5): 431-437.
- [2] Di Lu, Xiaoming Miao. "Meta-analysis of the influencing factors of business model innovation from a multi-level perspective." *Science and Technology Progress and Countermeasures*. China, vol. 35, pp. 93-101, 2018.
- [3] Weihong Xie, Zhongshun Li, Fang Su, Yongjian Wang. "Executive Support, Big Data Capability and Business Model Innovation." *Research and Development Management*. China, vol. 30, pp. 152-162, 2018.
- [4] Baoliang Hu. "Study on the Influencing Factors of the Adoption of Business Model Innovation." *Science and Technology Management Research*. China, vol. 33, pp. 6-9, 2013.
- [5] Weifeng Yao. "Evaluation Research on Influencing Factors of Enterprise Business Model Innovation." *Journal of Harbin University of Commerce (Social Science Edition)*. China, vol. 02, pp. 92-95, 2013.
- [6] Weihong Xie, Peiwang Lin; Zhongshun Li; Haizhen Guo. "Digital Innovation: Connotative Features, Value Creation and Prospects." *Foreign Economics and Management*. China, vol. 42, pp. 19-31, 2020.
- [7] Hai Guo, Jiaping Han. "The Impact of Open Innovation on the Growth of New Ventures in the Digital Context: The Mediating Role of Business Model Innovation." *Management Review*. China, vol. 31, pp. 186-198, 2019.
- [8] André Hanelt, Firk S , Hildebrandt B , et al. Digital M&A, digital innovation, and firm performance: an empirical investigation[J]. *European Journal of Information Systems*, 2020(11):1-24.
- [9] Nan Hu, Ting Zhang, Xiaoyu Li. "Descriptive innovation, real innovation activities and continuity of earnings: based on text analysis and machine learning." *The 17th China International Conference on Empirical Accounting*. China, 2018, 12.
- [10] Zott, C., Amit, R., Massa, L.. *The Business Model: Recent Developments and Future Research*. *Journal of Management*, 2011, 37(4): 1019-1042.

- [11] Johnson, M. W., Christensen, C. M., Kagermann, H.. Reinventing Your Business Model. Harvard Business Review, 2008, 86(12): 57-68.
- [12] Shouyang Wang, Jingning Ao, Han Qiao, Yifan Yang, Yi Hu, Mao Jiang. “The iceberg theory of business models based on knowledge management.” Management Review. China, vol. 27, pp. 3-10, 2015.
- [13] Yongyan Shen, Yanjie Lv. “Rethinking the Basic Concepts of Business Model.” Management Modernization. China, vol. 01, pp. 29-31, 2011.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

