

Article

Not peer-reviewed version

Exploring the Features and Trends of Industrial Products E-Commerce in China by Using Text Mining Approaches

Zhaoyang Sun , [Yuxin Mao](#) ^{*} , Qi Zong , Gongxing Wu

Posted Date: 26 September 2024

doi: 10.20944/preprints202409.2080.v1

Keywords: industrial products e-commerce; e-commerce platform; text mining; policy research; standard analysis



Preprints.org is a free multidiscipline platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This is an open access article distributed under the Creative Commons Attribution License which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Article

Exploring the Features and Trends of Industrial Products E-Commerce in China by Using Text Mining Approaches

Zhaoyang Sun ¹, Qi Zong ², Yuxin Mao ^{2,*} and Gongxing Wu ²

¹ High-Tech Standardization Research Institute, China National Institute of Standardization, Beijing 100191, China

² School of Management and E-Business, Zhejiang Gongshang University, Hangzhou 310018, China

* Correspondence: maoyuxin@zjgsu.edu.cn

Abstract: Industrial products e-commerce refers to the specific application of the e-commerce concept in the field of industrial products transactions. It enables industrial enterprises to conduct transactions via Internet platforms and reduce circulation and operating costs. Industry literature such as policies, reports and standards related to industrial products e-commerce contains abundant and crucial information. Through systematical analysis of the literature information, it is beneficial to explore and grasp the development characteristics and trends of the industrial products e-commerce. To study the characteristics and development status of industrial products e-commerce in China, literature comprising 18 policy documents, 10 industrial reports and 5 standards is deeply analyzed by employing text mining methods. Firstly, natural language processing (NLP) technology is utilized to pre-process the text data related to industrial products commerce. Then, word frequency statistics and TF-IDF keyword extraction are performed, and the results of word frequency statistics are visually displayed. Subsequently, the feature set is obtained by combining the manual screening method. The original text corpus is used as the training set by employing the skip-gram model in word2vec, and the feature words are transformed into word vectors in the multi-dimensional space. The k-means algorithm is used to cluster the feature words into groups. Moreover, the latent dirichlet allocation (LDA) method is also utilized to further group and discover the features. The research results based on text mining provide supportive decisions for uncovering the development characteristics and trends of industrial products e-commerce in China.

Keywords: industrial products e-commerce; e-commerce platform; text mining; policy research; standard analysis

1. Introduction

Industrial products refer to products and services that are mainly used in industrial production or for maintaining the operation of an enterprise. These products are typically not directly sold to individual consumers, but are procured and used by enterprises as means of production or maintenance, repair, and operation materials. According to the purpose of use by the purchaser and the original nature of the product, they are divided into two categories: one is non-production materials purchased for maintenance, repair, and operation, and the other is items directly incorporated into the production process, namely production materials. Industrial products e-commerce refers to the specific application of the e-commerce concept in the field of industrial products trading and circulation. It entails the online display, sales, procurement, and related service activities of industrial products via Internet platforms [1]. In contrast to traditional offline transaction methods, industrial products e-commerce employs digital technology [2] to realize the online and automated processes of information exchange, transaction matching, contract signing, payment settlement, and logistics distribution between supply and demand parties. Industrial enterprises

utilize e-commerce technologies, and both parties of transactions complete the entire transaction process through real-time interaction, timely grasping various market information, and reducing the implementation costs [3].

In recent years, the development momentum of the industrial Internet has been strongly influenced by policy promotion, market demand and digital technology progress. As a result, many countries and regions around the world attach great significance to the development of industrial products e-commerce. For instance, the EU has continuously strengthened its digital development strategy and issued a series of industrial policies to promote the implementation of the "Industry 5.0" strategy and promote the comprehensive digital transformation of the industry in the region. With the vigorous promotion of the industrial Internet and the "Industry 5.0" strategy by the EU and its member states, digital transformation has become the core driving force of the European industrial products market [4]. Meanwhile, Europe's leading industrial companies, such as Siemens, Dassault and SAP, are leading the industry's innovation and transformation by building advanced digital service platforms. These platforms not only optimize the management efficiency of the supply chain, but also promote the personalization and onlineization of products and services, inject high value-added products and technical support into industrial e-commerce, and further catalyze the prosperity and diversity of the market. In China, the digital economy has continuously accelerated the development of the B2B market for industrial products. To address the pain points of industrial development, the government has successively issued relevant policies to integrate the Internet economy with industrialized products. China's Ministry of Industry and Information Technology (MIIT) has also issued a series of policies related to e-commerce of industrial products, clearly stating that the application of industrial e-commerce should be further popularized and deepened, and encouraging and promoting industrial product B2B platforms to jointly hold online trading activities for industrial products.

Although the field of industrial products e-commerce has exhibited remarkable growth momentum in recent years, the industry itself is still in a crucial stage of transformation from traditional models to digitalization, and its scale benefits and systematic construction still need to be strengthened. The progress of industrial products e-commerce requires the dual roles of policy guidance and independent promotion by the industry. In-depth exploration of relevant literature such as policy documents, industry reports and standards will assist us accurately grasp the current status and characteristics of the industry, thereby better promoting the development of the industry and related enterprises. Based on this principle, this study aims to take industrial products e-commerce related literature including policies, reports and standards in China as research data, employ text mining approaches to conduct quantitative research and analysis, and minimize the bias caused by pure manual analysis as much as possible. Based on the process of text mining and analysis, targeted countermeasures and suggestions are put forward for the high-quality development of industrial products e-commerce. It has been proven that the text mining process used in this study can help reveal the topic features of industrial products e-commerce in China, help industrial enterprises better understand the policies related to industrial products e-commerce and key issues in the industry, and also provide an empirical basis and quantitative reference for the government to improve the support policy framework for industrial products e-commerce and optimize the construction of the standard system.

2. Related Works

Researchers have employed various theories to study the impact of e-commerce on the field of industrial product. According to different research focuses, existing research efforts can be divided into the following categories:

(1) Research on e-commerce and its role in the transformation of traditional industrial enterprises. Zhang et al. [5] proposed the concept of e-commerce embeddedness for the integration trend of e-commerce and manufacturing industry, analyzed the role of e-commerce on industrial research and development innovation mechanism, and the results showed that e-commerce embeddedness can significantly affect the R&D investment and R&D results of manufacturing

enterprises, and also revealed its role in the innovation and upgrading mechanism. Second, the obstacles faced in the development process of industrial e-commerce model. Claycomb et al. [6] empirically tested different models using overall B2B e-commerce use as the dependent variable and industrial firms' innovation characteristics, environment, channel factors, and organizational structure as predictor variables. The researchers found that factors such as compatibility with existing systems, technological specialization and information technology decision-making facilitated the overall use of B2B e-commerce in industrial firms, which in turn enhanced the value of the firm's performance.

(2) Research on the obstacles faced in the development process of industrial products e-commerce model. Chen et al. [7] believe that the key factor restricting the large-scale growth of industrial e-commerce is its backward development strategy, such as the conflict between the e-commerce model and the industrial sales model, as well as the lack of professional technical and service support. Thirdly, it is the study of the mechanism of how e-commerce promotes industrialization. Waithaka and Mnkanlla [8] identified technical, security, cost, lack of computer knowledge, and environmental issues as obstacles for Kenya's manufacturing industry to adopt the B2B market.

(3) Research on the mechanism of how e-commerce promotes industrialization. Tang and Wu [9] establish a power model, use the principle of system dynamics to discuss the various factors of e-commerce in the process of promoting industrial development, establish a causal loop diagram and a system flow diagram, and analyze the mechanism of e-commerce's impact on industrial development. Wang [10] explored how cloud computing and e-commerce affect industrial companies and industries in terms of technology architecture, service model and industry chain.

(4) General research about text mining on policy or other texts in other fields. Juventia et al. [11] used text mining to quantify countries' commitments towards safeguarding and using agrobiodiversity. The study extracted and scored relevant sections of official documents, revealing varying levels of commitment among countries. Puri et al. [12] applied commonsense knowledge to text mining of urban policy documents and social media postings. The approach uses reasoning based on commonsense knowledge to better account for pragmatics and semantics in the text, providing insights into public satisfaction and policy effectiveness. Tobback et al. [13] proposed an improved method for measuring economic policy uncertainty using text mining techniques. The study compared traditional keyword-based methods with modality annotation and support vector machines (SVM) classification. Rao and Dey [14] discussed how text mining techniques can assist in decision support for e-governance by retrieving and analyzing information from textual data sources. The study presents an integrated text mining-based architecture to help policymakers discover associations between policies and citizens' opinions expressed in electronic public forums and blogs.

A review of the literature reveals that there is relatively little work using text mining approaches to study and analyze industrial product e-commerce related literature. Existing research efforts mostly focuses on the basic concepts of industrial products e-commerce, the sorting and analysis of model characteristics, and the study of impact mechanisms. There are few quantitative analyzes of industrial products e-commerce from the perspective of text mining. However, for the industrial products industry with many characteristics, there is a gap between the current research results and the actual development needs, and these studies fail to fully combine the characteristics of the industry to further propose solutions, and lack of specific analyses and guidance to support the development of the industry.

Text mining is an increasingly prominent research tool in the field of data mining technology, which aims to reveal hidden patterns and regularities in large-scale text data [15]. Policy documents and industry reports contain a huge amount of information. Compared with traditional manual analyzing methods, text mining technology can effectively deal with problems such as large amounts of text and strong subjectivity. It has more advantages than other data mining techniques, during the process of analyzing the policies, reports and standards related to industrial products e-commerce. Therefore, from the perspective of text mining, research related to the industrial products e-

commerce industry has a large amount of explorable space, and future research should dig deeper into the industry’s needs.

3. Methodology

3.1. Data Collection

This study adopts literature including policies, reports and standards as research data, which presents the information about industry development from different dimensions. Policies are typically formulated and issued by governments at all levels. Thus, we download the related policies mainly from different government websites. The reports are collected from different industry research institutes or consulting companies. The standards mainly originate from various industry organizations and relevant government departments. After collection and screening, the research data finally gathered mainly include 18 policy documents from 2015 to 2023, 10 research reports, as well as 5 national standards in China. Some sample policies, reports and standards are given in Tables 1–3. All these texts are written in Chinese.

Table 1. Sample policies related to industrial products e-commerce.

NO	Policy Title	Issuing Department	Date
1	Guiding Opinions on Deepening the Integration and Development of Manufacturing and the Internet	State Council (PRC)	2016.05
2	Three-Year Action Plan for the Development of Industrial E-Commerce	MIIT of China	2017.09
3	Guiding Opinions on Promoting the Orderly Reopening of Industrial Communications Enterprises	MIIT of China	2020.02
4	Development Plan for the Deep Integration of Informatization and Industrialization under the 14th Five-Year Plan	MIIT of China	2021.11

Table 2. Sample reports related to industrial products e-commerce.

NO	Report Title	Research Organization	Date
1	China Manufacturing Industry Internet C2M E-Commerce Industry Research Report	iResearch	2019.05
2	Industrial E-Commerce White Paper	Department of Information Technology and Software Services, MIIT of China	2019.07
3	2020 China Industrial Products B2B Industry Research Report	36Krypton Research Institute	2020.12
4	Research on the Development and Investment Value of China’s Industrial E-commerce in 2020	CCID Consultants	2020.12

5	2022 China Industrial Products B2B Industry Research Report	iResearch	2022.07
---	----------------------------------------------------------------	-----------	---------

Table 3. Sample standards related to industrial products e-commerce.

NO	Standard Title	Standard Number
1	E-Commerce Supplier Evaluation Criteria Quality Manufacturers	GB/T 30698-2014
2	Industrial Internet Platform Application Implementation Guide Part 2: Digital Management	GB/T 23031.2-2023
3	Industrial Internet Platform Selection Requirements	GB/T 42562-2023
4	Industrial Internet Platform Application Implementation Guide Part 5: Personalized Customization	GB/T 23031.5-2023
5	Industrial Internet Platform Application Implementation Guide Part 6: Service Extension	GB/T 23031.6-2023

3.2. Text Mining Framework for Literature Analysis

Text mining approaches have been widely used in analyzing industrial literature to uncover underlying patterns. In this regard, the design of an effective text mining framework becomes essential. The design of the framework entails the planning and organization of the entire task, and a well-designed framework can enhance efficiency and ensure the achievement of text mining goals. In this study, the literature from industrial products e-commerce will be dealt with in detail by the framework illustrated in Figure 1. The framework mainly comprises the following modules.

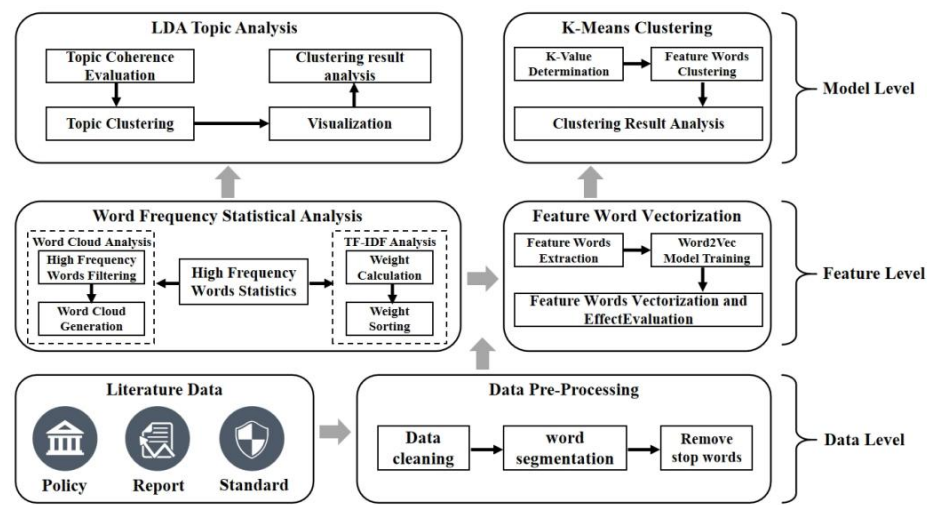


Figure 1. Framework of text mining on industrial products e-commerce.

(1) Data Pre-Processing: Data pre-processing is an important module for subsequent data analysis, which includes two key tasks: data cleaning and word segmentation. Firstly, we will deal with the data with different formats from different sources and try to convert it into a unified format. We also remove irrelevant information from the original data. Then, we split the text data into independent words. Subsequently, optimize and adjust the segmentation results by. stop words filtering, retaining only key information. We can add new vocabulary through a custom dictionary to maintain word integrity.

(2) Word Frequency Statistical Analysis: We perform word frequency statistics to find out the keywords with high frequency in the text corpus, which helps to reduce data dimensions and thus alleviate the burden of subsequent model training. In this study, the TF-IDF [16] method is used to calculate the weights of the feature words in the text corpus, and these feature words are ranked according to the magnitude of their weights, and the words with a TF-IDF value higher than a specified threshold are selected as the final feature words. This step significantly reduces the dimensions of the text model, providing a suitable model foundation for subsequent semantic calculations. Moreover, we can also build word cloud according to the results of word frequency statistics to highlight high frequency words further.

(3) Feature Word Vectorization: By learning the distributed representation of words in the context, words with similar semantics are kept close to each other in the vector space. Word2Vec is a word embedding technique used to convert words in text into vector representations, which has advantages such as simple models, fast training speed, and the ability to express similarities and analogical relationships between different words effectively. Two possible models, Skip-Gram [17] and CBOW can be used to train the Word2Vec model. Selecting the Skip-Gram model for word vectorization training results in more predictions, but through optimization of multiple parameters, the final word vector obtained is more accurate. Therefore, Skip-Gram model is chosen for vectorization training of feature words in this study.

(4) K-Means Clustering: K-means is a simple and classic clustering algorithm with simple algorithm ideas and fast calculation speed, and is suitable for text clustering tasks on large-scale data sets. First, the appropriate k value is determined using the profile coefficient, the excellent samples are selected through the individual profile coefficients of the data objects, and the initial clustering centre is adaptively selected for k-means clustering [18] that optimizes the selection of initial clustering centres. The calculation formula of the profile coefficients is as follows:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (1)$$

The calculated sample $a(i)$ is the average distance from sample i to other samples in the same cluster, and $b(i)$ is the average distance from sample i to other clusters. The value of the profile coefficient is between -1 and 1. The closer it is to 1, the better the cohesion and separation are. The closer it is to -1, the worse the cohesion and separation are. The average of the profile coefficients of all data points is the total profile coefficient of the clustering result.

Second, each sample is assigned to the cluster to which the nearest clustering centre belongs. Thirdly, the clustering centers are updated by calculating the average value of all the samples in each cluster and using it as the new clustering centre. Fourthly, the second and third steps are repeated and if the distance between the newly calculated clustering centre and the original centre is less than a set threshold, the clustering is considered to reach the desired result and the algorithm stops.

(5) LDA Topic Analysis: LDA is a generative model for text topic modeling. It assumes that each document is generated by a mixture of multiple topics, and each topic is generated by a group of words. This study constructs a topic model and performs topic clustering on the pre-processed text data. The coherence calculation method is used to determine the number of topics, and then the topic hotspots are identified. Finally, the results of the LDA model analysis are visualized.

Topic coherence measures are used to evaluate the quality of topics in the LDA model. A common topic coherence measure is the topic coherence score, the basic idea of which is to measure the co-occurrence of words in the topic. One of the formulas for topic coherence score is based on mutual information (PMI):

$$\text{Coherence}(V) = \sum_{m=2}^M \sum_{l=1}^{m-1} \log \frac{D(v_m, v_l) + \epsilon}{D(v_l)} \quad (2)$$

Topic strength is also called topic popularity or topic attention. It can show the degree of attention paid to a specific topic within a specific time period and is a quantitative indicator for judging the attention allocation of a specific topic. The formula for calculating topic strength is:

$$P_k = \frac{\sum_i^N \theta_{ki}}{N} \tag{3}$$

P_k represents the strength of the k th topic, N is the number of texts, and θ_{ki} represents the probability of the k th topic in the i th text.

4. Findings

Based on the text mining framework for industrial products e-commerce, we can explore the development features and trends of industrial products e-commerce in depth.

4.1. Keyword Frequency Analysis

Keyword analysis plays an important role in revealing the core content of industrial literature. In this study, the Jieba library [19] was used to perform the word-splitting process of text data to obtain the keywords and their word frequency distribution. Subsequently, the keyword frequencies of all the documents were summarized and calculated, and the word segmentation results were ranked according to the frequency to generate a word cloud map of the industrial products e-commerce industry, as shown in. Since the samples are industry support policies and market research reports, the document set after word separation may contain high-frequency words such as pronouns, quantifiers, convergent verbs, etc., which are not substantially helpful for text characterization, so these words are eliminated. Finally, the top 20 effective high-frequency words are organized, as shown in **Error! Reference source not found.**.

Table 4. Frequency of top 20 keywords.

Keyword	Word Frequency	Keyword	Word Frequency
industry	2542	data	728
enterprise	2068	ability	702
platform	1908	industry	695
service	1468	management	666
e-commerce	1198	production	648
Internet	1189	demand	633
development	1119	digitalization	624
product	897	B2B	610
procurement	841	technology	603
industrial products	758	manufacturing	568

In terms of keyword frequency distribution, the terms “industry”, “enterprise”, “platform” “service” and “e-commerce” have the highest frequency. It is inferred that industrial products e-commerce and platforms are important engines for promoting the innovation and development of industrial enterprises. By supporting the key cultivation platforms of industrial products e-commerce, the government and the industry have pushed enterprises to accelerate platform construction. The frequency of terms “digitalization”, “data”, and “technology” arranged from high to low shows that the industry is increasingly focused on achieving cost reduction and efficiency improvement in digital transformation, in order to build a sound service system.

To further show the results of keyword frequency analysis, we generate word cloud (see Figure 2) based on the keyword with high frequency. In the word cloud, keywords with larger font sizes appear more frequently in the documents. The keywords in the word cloud are all in Chinese as they are in the original texts.



Figure 2. Industrial products e-commerce word cloud (in Chinese).

4.2. Feature Extraction and Vectorization

In the process of feature word extraction, we set the threshold value to 0.1, and only feature words with TF-IDF value greater than 0.1 will be selected. **Error! Reference source not found.** shows some feature words with high TF-IDF values. On the basis of performing feature word extraction, we further train the model by using Word2Vec to obtain word vectors. By processing the pre-processed text corpus, a corpus word list is obtained, in which each word corresponds to a 200-dimensional space vector. The industrial products e-commerce feature words correspond to 17 200-dimensional word vectors in this word list.

Table 5. Feature word screening and corresponding TF-IDF values.

Characteristic Word	TF-IDF Value
industry	0.404026256
enterprise	0.347823617
platform	0.283990818
e-commerce	0.222205136
development	0.205269903
service	0.203222787

The effectiveness of the trained word vectors can be evaluated by looking at the similarity between words and the list of related words for a single word. The four word pairs extracted from policy documents are “digitalization” and “technology”, “manufacturing” and “production”, “material” and “platform”, and “user” and “Internet” in the policy document, for example, the similarity of each word pair is calculated, and the results are shown in **Error! Reference source not found.** Observing the word pair similarity results, it is found that the processing results are consistent with people’s daily cognition, which indicates that the Word2Vec training model can generate reasonable and effective word vectors. Therefore, the same operation can be taken for other feature words in the corpus, so as to verify the effectiveness of the word vectors obtained by training.

Table 6. Similarity of related word pairs.

Pair of Words	Similarity
<Digitalization, Technology>	0.83570686
<Manufacturing, Production>	0.91326806
<Material, Platform>	0.33396235

<User, Internet>	0. 26313045
------------------	-------------

4.3. K-Means Clustering

The contour coefficient is a metric employed to evaluate the quality of clustering, integrating the tightness within clusters and the separation between clusters. After obtaining the feature vectors, the contour coefficients are calculated for different values of k. By observing the trend of the contour coefficients as k values change, it is found that the contour coefficients gradually decrease with the increase of k values and fluctuate within the range from 0.06 to 0.44. In the cases where k is equal to 3, 4, and 5, the corresponding contour coefficients are the three largest values among all the contour coefficients corresponding to different k values, which are 0.4377, 0.2673, and 0.2922. Since a contour coefficient closer to 1 indicates a better clustering effect, we choose to set the value of k to 5 considering that dividing into 3 and 4 clusters is not practical.

After determining the value of k, the feature words are clustered using the k-means clustering algorithm. The details of the algorithm are given below. When k=5, the feature screening and the corresponding categories are shown in Table 7. According to the clustering results in Table 7, each cluster represents a specific topic or related field within industrial products e-commerce. Cluster 1 is mainly about the products and services provided by industrial products e-commerce to meet user needs. Cluster 2 mainly reflects the close connection of industrial products e-commerce to the development of the Internet and e-commerce. Cluster 3 mainly indicates that the B2B model is the most common operation mode in industrial products e-commerce. Cluster 4 mainly involves industrial enterprise development, platform construction, technology application, and digital transformation. Cluster 5 is mainly associated with the manufacturing production of industrial products.

Algorithm: K-Means Clustering
Input: $D = \{x_1, x_2, \dots, x_n\}$ (set of n data points), k (number of clusters)
Output: A set of k clusters
1: # Step 1: Initialization
2: Choose k initial centroids z_1, z_2, \dots, z_k from D randomly.
3: Initialize a set of clusters $C = \{C_1, C_2, \dots, C_k\}$, where each C_i is initially an empty cluster.
4: # Step 2: Assignment
5: Repeat until convergence (centroids do not change or a maximum number of iterations is reached):
6: For each data point x_i in D:
7: Find the closest centroid z_j to x_i (the one with the smallest distance $d(x_i, z_j)$).
8: Assign x_i to the cluster C_j corresponding to the closest centroid z_j .
9: # Step 3: Update
10: For each cluster C_j :
11: Compute the new centroid z_j as the mean of all the data points currently assigned to C_j .
12: If any of the centroids have changed, go back to the Assignment step; otherwise, the algorithm has converged.
13: # Step 4: Termination
14: The algorithm stops when there is no change in the centroids between successive iterations or when a predefined stopping criterion is met.
15: Return the final set of clusters C.

In summary, these clusters involve several aspects such as product, service, Internet, B2B, enterprise digitalization, technology construction, and manufacturing. It can be observed that the construction and policy development of China’s industrial products e-commerce are gradually maturing. At present, industrial products e-commerce has advanced from the initial stage of platform construction to the stage of deep integration of industrial products trading, production, and services, while the policy focus has gradually shifted to ensuring the smooth flow of all aspects of the supply chain through industrial products e-commerce.

From the perspective of the characteristics within the clusters, Cluster 2 and Cluster 3 involve the Internet, industry, and e-commerce, indicating that these two clusters are concerned with the interaction between the Internet and industrial products e-commerce in the development of e-commerce. National policies continuously promote the combination of the Internet and the industrial sector. This not only creates a favorable environment for the promotion and development of industrial products e-commerce but also prompts the practical application of the industrial Internet platform. The two promote each other to form a virtuous circle. Cluster 4 contains the most keywords, signaling the further extension of technology application to the industrial side of industrial products e-commerce. The government and the industry are jointly committed to promoting the digital transformation of the supply chain. The transformation of enterprise management through digital technology has become an important direction in the industrial products e-commerce market competition. Cluster 1 reflects the deepening of services becoming the new focus of industrial products e-commerce. With the continuous development of industrial products e-commerce, the improvement of its service system has become an inevitable requirement. Due to the diversity and complexity of industrial products, both the supply and demand sides are facing certain problems. To solve these problems, the service system must be continuously improved. Cluster 5 is mainly associated with the manufacturing production of industrial products. Due to the relative maturity of manufacturing policies and industry research, in relation to the industrial Internet and industrial products e-commerce, there is not much need for extensive text narrative.

Table 7. Feature screening and corresponding categories.

Cluster	Featured Keywords
1	product, service
2	internet, development, industry, e-commerce
3	b2b, industry, industrial
4	enterprise, platform, construction, technology, digitalization, data
5	fabrication

4.4. LDA Topic Modeling and Visualization

This study draws a curve based on the number of optimal topics through topic coherence, as shown in Figure 3. Generally speaking, higher topic coherence indicates stronger topic internal connections and higher interpretability.

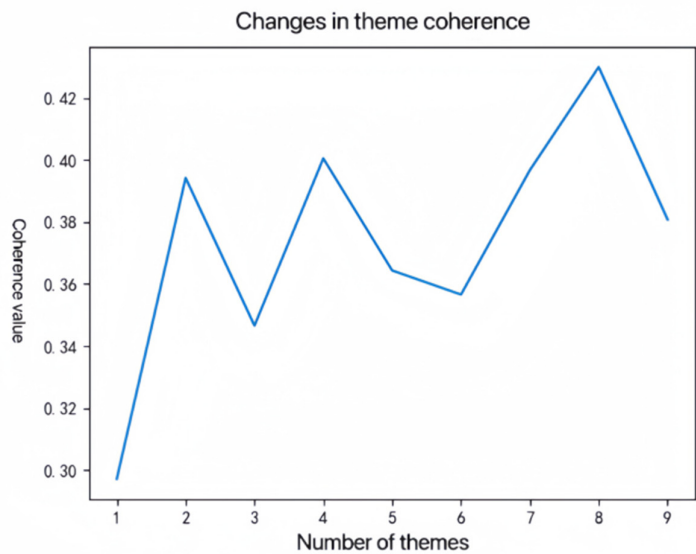


Figure 3. Topic coherence for LDA clustering.

The topic coherence is used to identify that 8 is the appropriate number of topics in this job. Based on this, the number of model topics is set to 8, and the model is rerun to obtain the topic category to which each document most likely belongs. Through LDAvis [20], the visual documents of topic clustering, the visualization results, and the frequency distribution of feature words under each topic are obtained, as shown in Figure 4. The document results show the distribution and probability ratio of each of the 8 topics. When a topic is selected, 30 representative hot keywords under the topic will emerge, which can be used to determine the topic content. The size of the circle intuitively reflects the significance and coverage of each topic in the entire text corpus, that is, the topic intensity. The larger the area of the circle, the higher the proportion of the topic in the entire corpus and the more significant its importance. The distance between circles reveals the degree of correlation between topics, the closer the distance, the higher the correlation between topics. Figure 5 shows the visualization results of word frequency distribution of topic 1.

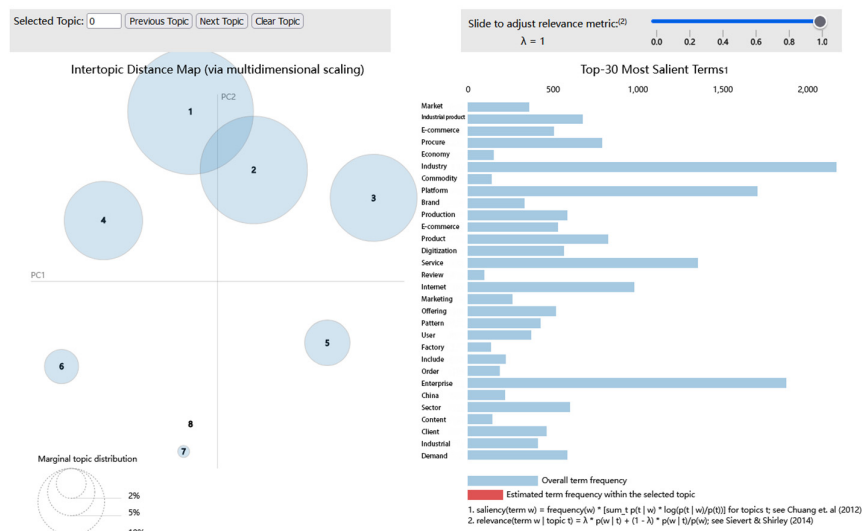


Figure 4. Frequency distribution of feature words under each topic.

The results of the LDA topic model analysis of the industrial products e-commerce related texts are shown in Table 8. This study manually summarized the general titles of each topic based on the text modeling results. According to the data on the proportion of topics, we can find that Topic 1 occupies the highest proportion, reaching 35.6%. These words cover core concepts such as industry, platform, service, enterprise, and e-commerce, indicating that these are the hot topics from the industrial products e-commerce field in China, and also show the importance of industrial products e-commerce platforms and related corporate management and market demand.

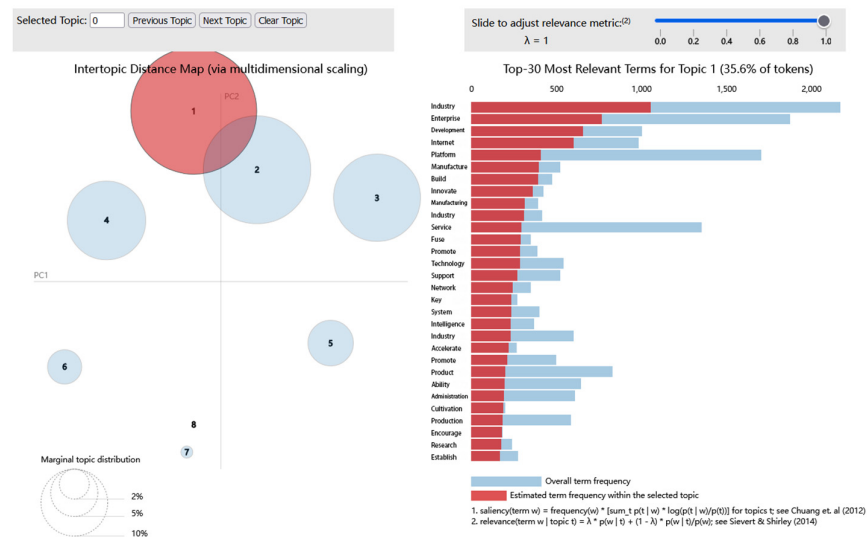


Figure 5. Frequency distribution of feature words under topic 1.

The other topics also have their own features. Topic 2 reflects the transformation and upgrading process of industrial product enterprises in the digital era and the changes in consumer behavior; Topic 3 emphasizes the impact of the macroeconomic environment on the circulation of industrial products e-commerce market and development trends in China; Topic 4 covers and emphasizes the changes brought about by Internet technology to traditional manufacturing and the exploration of new manufacturing models; Topic 5 reflects the role of industrial products e-commerce service providers in market competition and the impact of customer evaluation on corporate image; Topic 6 explores the impact of changes in consumer behavior on the operating model of industrial products e-commerce platforms and the importance of data analysis; and the scope of Topic 7 and Topic 8 is too small and can be ignored.

Table 8. Results of LDA topic model for industrial products e-commerce related texts.

NO	Topic	Topic Feature Words (Top 8)	Topic Ratio (%)
1	Industrial products e-commerce platform	industry, platform, service, enterprise, e-commerce, demand, Internet, management	35.6
2	Digital transformation	enterprise, e-commerce, product, digitalization, industry, industrial product, marketing, user	25.9

3	Market circulation	market, commodity, economy, circulation, entity, consumption, development, production and sales	17
4	Manufacturing innovation	industry, development, Internet, platform, manufacturing, construction, innovation, manufacturing industry	13.8
5	Service review	review, e-commerce, product, manufacturer, module, enterprise, provide, content	4.7
6	Consumption model	brand, factory, consumption, platform, model, order, demand, data	2.6

5. Conclusions

This study uses text mining methodology to explore the policies, reports and standards related to industrial products e-commerce and draws the following conclusions:

(1) Platforms are the core of the industrial products e-commerce ecosystem. In the policy-based, report-based and standard-based text data, the rank of keywords such as “platform” and “enterprise” shows their high importance, indicating that policies and industries actively promote the innovative development of platforms and continue to empower industrial enterprises. For example, the industrial products e-commerce development program of China’s Jiangsu Province emphasizes encouraging large enterprises to build their own collection and marketing platforms, cultivating third-party industrial e-commerce service platforms, upgrading the service capacity of key cultivation platforms, and forming a transparent, efficient, and cost-effective centralized purchasing system on the Internet. The industrial products e-commerce platform is also constantly improving the supply chain finance, warehousing and logistics services and other diversified functions, to meet the needs of enterprises online sales at the same time also bring more value-added benefits for the industrial enterprises to take the initiative to integrate into the new e-commerce business to provide good conditions.

(2) Policy and industry focus on technology application and landing of industrial products e-commerce. From the keyword extraction and clustering results, we preliminarily conclude that the focus of policies and industries has shifted to the combination of technologies with industries and applications, and more attention has been paid to the landing of industrial products e-commerce related technologies. For example, the Ministry of Industry and Information Technology of China and many other departments mentioned improving the level of intelligent manufacturing, encouraging the innovation of industrial Internet, 5G, artificial intelligence and industrial APP integration application modes and technologies, accelerating the application innovation and landing, and guiding more enterprises to apply industrial products e-commerce platforms to carry out business. In the traditional industrial supply chain channel, various roles including agents, distributors and retailers are actively transforming themselves to adapt to the trend of digitalization, and comprehensively using e-commerce platforms to promote the digitalization process of the entire supply chain.

(3) The industrial products e-commerce has not yet formed a sound policy framework. From the process of collecting policy documents in this study and the number of valid texts eventually collected, the strategic design and institutional system related to the industrial products e-commerce in China has not yet been perfected, and the degree of protection of the institutional and digital ecosystem is relatively low, and the institutional support provided to the digital transformation of small and medium-sized enterprises (SMEs) is insufficient. The progress of regional and district governments in the implementation of policies involving industrial products e-commerce is relatively slow, and in some areas there is even a lag. Although the country has issued relevant

policies, different regions have not yet formulated corresponding implementation rules or policy opinions, resulting in the digital transformation of industrial SMEs lacking a good cluster effect and the formation of a relatively weak e-commerce ecological environment.

(4) The degree of standardization in the industrial products e-commerce industry is low. Although industrial products e-commerce has made significant progress in technology application and platform construction, there is a relative lack of discussion on “standards”, “norms” or “certification systems”, highlighting the lag in standardization construction in the current industrial products e-commerce field. Standardization is not only related to basic aspects such as product description, quality control, and logistics packaging, but also involves a series of core aspects of industrial product e-commerce operations such as service quality and transaction processes. The lack of unified standards leads to serious information asymmetry problems, increases transaction costs, and limits the efficient operation and resource integration capabilities of the industrial products e-commerce market.

Based on the analysis and findings of this study, government departments can further improve policies, while enterprises can further adjust and optimize their business strategies to jointly promote the healthy development of the industrial products e-commerce.

Author Contributions: Conceptualization, Zhaoyang Sun and Yuxin Mao; methodology, Qi Zong and Yuxin Mao; formal analysis, Qi Zong; writing—review and editing, Gongxing Wu and Zhaoyang Sun. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Acknowledgments: This research was funded by Basic Scientific Research Business Fees Project “Research on Machine-Readable Standards and Intelligent Application Technologies for the Clothing Industry” (NO. 532023Y-10393), and Major Humanities and Social Sciences Research Projects in Zhejiang Higher Education Institutions (NO. 2023QN077).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Ben Youssef, A.; Dahmani, M. Examining the Drivers of E-Commerce Adoption by Moroccan Firms: A Multi-Model Analysis. *Information*. 2023, 14(7), 378.
2. Zhang, C.; Yang, Q.; Zhang, J.; Gou, L.; Fan, H. Topic Mining and Future Trend Exploration in Digital Economy Research. *Information*. 2023, 14(8), 432.
3. Martín-Gómez, A.M.; Agote-Garrido, A.; Lama-Ruiz, J.R. A Framework for Sustainable Manufacturing: Integrating Industry 4.0 Technologies with Industry 5.0 Values. *Sustainability*. 2024, 16(4), 1364.
4. Ocloo, C.E.; Xuhua, H.; Akaba, S.; Shi, J.; Worwui-Brown, D.K. The Determinant Factors of Business to Business (B2B) E-Commerce Adoption in Small-and Medium-Sized Manufacturing Enterprises. *Journal of Global Information Technology Management*. 2020, 23(3), 191-216.
5. Zhang, Y.H.; Zhuang, Z.Z.; Li, Z.W. Can E-Commerce Promote Innovative Behavior in Traditional Manufacturing? [J]. *The Journal of Quantitative & Technical Economics*. 2018, 35 (12), 100-115. (in Chinese)
6. Claycomb, C.; Iyer, K.; Germain, R. Predicting the Level of B2B E-Commerce in Industrial Organizations. *Industrial Marketing Management*. 2005, 34(3), 221-234.
7. Chen, M.L.; Chen, Y.F.; Lin, Q.Y. Research on Industrial Control Mode of Electronic Commerce under Industry 4.0 Background. *Manufacturing Automation*. 2015, 37 (4), 146-147+150. (in Chinese)
8. Waithaka, S.T.; Mnkanjala, E. Challenges Facing the Use of Mobile Applications for E-Commerce in Kenya's Manufacturing Industry. *The Electronic Journal of Information Systems in Developing Countries*. 2017, 83(1), 1-25.
9. Tang, P.P.; Wu, L. A Research on the Effect Mechanism of New Electronic Commerce to Theunderdeveloped Areas of China. *Chinese Journal of Management*. 2014, 11(08), 1143-1149. (in Chinese)
10. Wang, D. Influences of Cloud Computing on E-Commerce Businesses and Industry. *Journal of Software Engineering & Applications*. 2015, 6(6), 313-318.
11. Juventia, S.D.; Jones, S.K.; Laporte, M.A.; Remans, R.; Villani, C.; Estrada-Carmona, N. Text Mining National Commitments Towards Agrobiodiversity Conservation and Use. *Sustainability*. 2020, 12(2), 715.

12. Puri, M.; Varde, A.S.; de Melo, G. Commonsense based text mining on urban policy. *Language Resources and Evaluation*. 2023, 57(2), 733-63.
13. Tobback, E.; Naudts, H.; Daelemans, W.; de Fortuny, E.J.; Martens, D. Belgian Economic Policy Uncertainty Index: Improvement Through Text Mining. *International Journal of Forecasting*. 2018, 34(2), 355-65.
14. Rao, G.K.; Dey, S. Decision Support for E-Governance: A Text Mining Approach. *International Journal of Managing Information Technology (IJMIT)*. 2011, 3(3), 73-91.
15. Talib, R.; Hanif, M.K.; Ayesha, S.; et al. Text Mining: Techniques, Applications and Issues. *International Journal of Advanced Computer Science and Applications*. 2016, 7(11), 414-418.
16. Qaiser, S.; Ali, R. Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications*. 2018, 181(1), 25-29.
17. Jin, X.; Zhang, S.; Liu, J. Word Semantic Similarity Calculation Based on Word2Vec. In *Proceedings of the 2018 International Conference on Control, Automation and Information Sciences (ICCAIS)*, IEEE, 2018, 12-16.
18. Krishna, K.; Murty, M.N. Genetic K-Means Algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*. 1999, 29(3), 433-439.
19. Ding, Y.; Teng, F.; Zhang, P.; Huo, X.; Sun, Q.; Qi, Y. Research on text information mining technology of substation inspection based on improved Jieba. In *Proceedings of the 2021 International Conference on Wireless Communications and Smart Grid (ICWCSG)*, IEEE, 2021, 561-564.
20. Sievert, C.; Shirley, K. LDavis: A Method for Visualizing and Interpreting Topics. In *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, 2014, 63-70.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.