



# An Analysis of the Hot Topics and Evolution of Chinese E-commerce Policy Based on LDA-Word2vec Model

Xiaolu Xia\*, Jieqin Jin, Fang Zhang

School of Economics and Management, Nanjing University of Technology, Nanjing 210000, China

\*2767891007@qq.com

**Abstract.** This paper uses the LDA-Word2vec model to identify the hot topics and analyze the evolution of 665 e-commerce policies published in China from 1999 to 2023. It is found that there are obvious life cycle characteristics in the evolution process of China's e-commerce policy, which can be divided into three stages: germination stage, development stage and slowdown stage. There are some differences in the hot topics of e-commerce policy in different periods, but at the same time, there are some evolutionary correlations among the hot topics. In the course of theme evolution, there are five kinds of evolutionary relations: birth, extinction, inheritance, merger and split. In the future, China's e-commerce policy will continue to focus on the new theme of "e-commerce poverty alleviation", and focus on promoting the sustainable development of themes such as "rural e-commerce" and "supervision and management".

**Keywords:** Electronic commerce, Policy, Text mining, Subject recognition, Thematic evolution.

## 1 Introduction

E-commerce is a new business form, which plays an increasingly important role in promoting economic growth and creating job opportunities. In order to standardize the e-commerce market order, various countries have issued a series of policies and regulations. With the rapid development of e-commerce in China, China has continuously introduced relevant policies and regulations to promote the healthy development of e-commerce. E-commerce policy contains important information such as the development trend of the industry, so it is of great practical significance to analyze it deeply. In this paper, LDA-Word2vec model is used to identify and analyze the topic and hotspot of Chinese e-commerce policy text, and provide new ideas for e-commerce policy text analysis.

## 2 Literature Review

At present, existing researches mainly focus on e-commerce frontier<sup>[1]</sup>, e-commerce trial platform<sup>[2]</sup>, new farmers' live e-commerce<sup>[3]</sup>, ICT application and e-commerce development<sup>[4]</sup>, the influence and mechanism of rural e-commerce clusters on farmers' entrepreneurial behavior<sup>[5]</sup>, factors restricting the development of rural e-commerce in China and countermeasures<sup>[6]</sup>. E-commerce logistics service quality evaluation<sup>[7]</sup>. To sum up, the existing researches mainly use statistical analysis and other methods to investigate the content and impact of e-commerce policies, and lack a systematic investigation of the overall development context and evolutionary characteristics of China's e-commerce policies. The key information of policy text can be revealed accurately by using text mining technology. Among them, LDA topic model is one of the commonly used models for topic recognition, and Word2vec<sup>[8]</sup> model is a neural network-based word vector model. In view of this, this paper combines LDA and Word2vec two models, taking into account the language model and semantic structure of the topic, so as to better grasp the development context and evolution process of China's e-commerce policy.

## 3 Research Design

### 3.1 Research Idea

In order to quantitatively analyze the hot topics of China's e-commerce policy and their evolution, firstly, the collected and pre-processed e-commerce policy text data are divided into stages. Then, LDA<sup>[9]</sup> and Word2vec<sup>[10]</sup> were used to construct the LDA-Word2VEC topic model, analyze the topic evolution of the e-commerce policy text corpus under different time Windows, and identify the hot topics. The theme similarity is used to reveal the theme evolutionary correlation and evolutionary relationship, and to find the theme evolution path of China's e-commerce policy.

### 3.2 Data Specification

Peking University Talisman and the Chinese government website were used as data sources. Specific search process: The search time is limited to 1999-2023. Search terms include "e-commerce", "e-commerce", "agricultural business", "agricultural products Internet", etc., for full text accurate search. The policy texts were de-emphasized and screened in stages, and 665 policy texts were finally determined as sample data. You can use the jieba library in python to segment words, clean them using regular expressions, and load a user-defined dictionary to prevent some words from being cut incorrectly. At the same time, in order to improve the efficiency and quality of the theme model, the "Chinese stop word table", "HIT stop word table", "Baidu stop word table", and "Sichuan University Machine Intelligence Laboratory stop word database" are comprehensively adopted in the stop word table.

## 4 Empirical Analysis and Results

### 4.1 Topic Hot spot Analysis

Through the statistics of policy release time, it is found that China's e-commerce policies are divided into three distinct periods: a total of 119 policies were issued during the policy germination period (1999-2014), with a small number of policies and a slow growth trend; During the policy development period (2015-2018), there were 298 policies, and the number of policies issued increased rapidly, reaching a peak in 2018; There are 248 policies in the policy slowdown period (2019-2023), and the number of policy releases tends to be stable. According to the division of e-commerce policy stages, the confusion degree of different stages is calculated, and the optimal number of topics in the three stages is determined to be 4, 6 and 6, respectively, as shown in Figure 1.

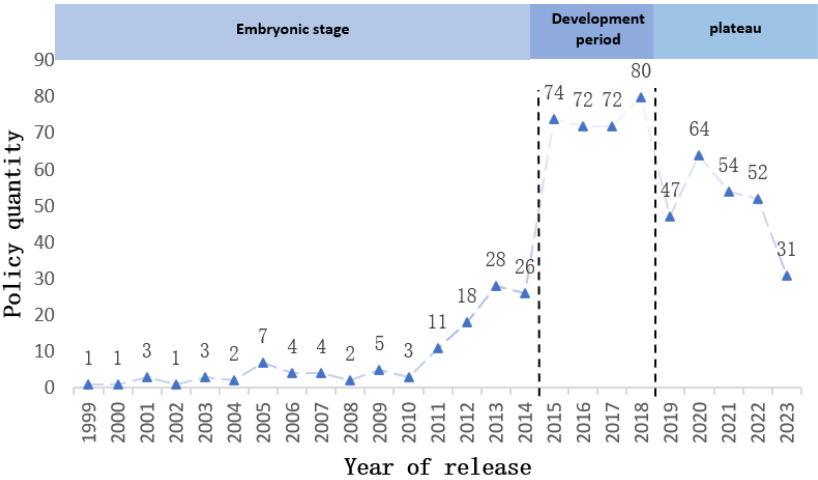


Fig. 1. E-commerce policy release time stage division.

### 1. Topic clustering

Based on the time staging of China's e-commerce policy release and the calculation of confusion degree, theme clustering was carried out to identify the themes of China's e-commerce policy in the three stages of infancy, development and slowdown, as well as the hot topic description (see Table 1).

Table 1. Topics and topics describe hot topics.

| Embryonic stage |                                 | Development period |                                 | plateau |                        |
|-----------------|---------------------------------|--------------------|---------------------------------|---------|------------------------|
| Topic0          | Internet in-formatization       | Topic0             | E-commerce demonstration county | Topic0  | Market super-vision    |
| Topic1          | Information online transac-tion | Topic1             | Product quality                 | Topic1  | Digital industry chain |

|        |                     |        |                                |        |                                |
|--------|---------------------|--------|--------------------------------|--------|--------------------------------|
| Topic2 | Rural e-commerce    | Topic2 | E-commerce import and export   | Topic2 | Plastic product                |
| Topic3 | Balance of payments | Topic3 | Personnel training             | Topic3 | Rural e-commerce               |
|        |                     | Topic4 | Rural e-commerce               | Topic4 | Supervision and administration |
|        |                     | Topic5 | E-commerce poverty alleviation | Topic5 | E-commerce poverty alleviation |

2. Thematic popularity

In order to intuitively show the concern of different topics of e-commerce policy, the popularity of different topics is visualized. In the topic heat map, the topic heat is represented by the depth of color, with dark colors being the more hot topics and light colors being the less hot topics. As can be seen from Figure 2, the heat value of each theme in the germination stage is relatively average, with the highest Topic0 heat value being 0.32 and the lowest Topic1 heat value being 0.22. In the development period and the slowdown period, except for a few topics, the theme popularity is basically below 0.20. This indicates that the research heat of e-commerce policy in the embryonic stage is relatively high. As time goes by, the policy concerns of e-commerce in the development period and the slowdown period continue to expand, the theme heat gradually diverges, and the policy orientation shifts to explore the new direction derived from the embryonic theme.

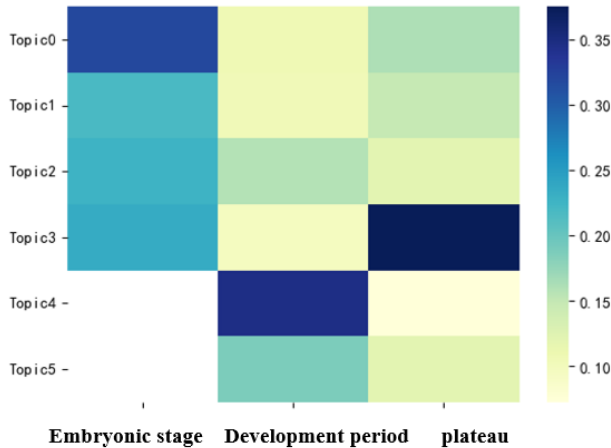


Fig. 2. E-commerce policy topic heat map.

4.2 Thematic Evolution Analysis

1. Topic evolution correlation

Through the function of Word2Vec library<sup>[11]</sup> in Gensim package, the subject words are mapped to the unified vector space, and the subject vector matrix of different stages

is generated. Through the similarity method<sup>[12]</sup>, cosine similarity between all topic vectors in each stage is calculated, and the forward and backward correlation of topics in each stage is established. Select the time window of the development period, and calculate the similarity between each theme in the development period and each theme in the germination period and the slowing period respectively. The corresponding theme with the greatest similarity is taken as the forward theme and the backward theme of each theme in the development period. For example, the similarity of topic associations in the development period is shown in Table 2.

Table 2. Developmental theme association.

| prior  | sim   | Topic  | post   | sim   |
|--------|-------|--------|--------|-------|
| Topic3 | 0.713 | Topic0 | Topic0 | 0.790 |
| Topic3 | 0.594 | Topic1 | Topic2 | 0.632 |
| Topic1 | 0.642 | Topic2 | Topic2 | 0.665 |
| Topic2 | 0.702 | Topic3 | Topic2 | 0.701 |
| Topic2 | 0.720 | Topic4 | Topic3 | 0.920 |
| Topic2 | 0.629 | Topic5 | Topic3 | 0.921 |

2. Thematic evolution relationship

It can be found that there are five types of evolutionary relationships in the evolution process of e-commerce policy, namely, new birth, extinction, inheritance, merger and split of themes, by discriminating theme evolution relationships through topic association filtering, as shown in Figure 3.

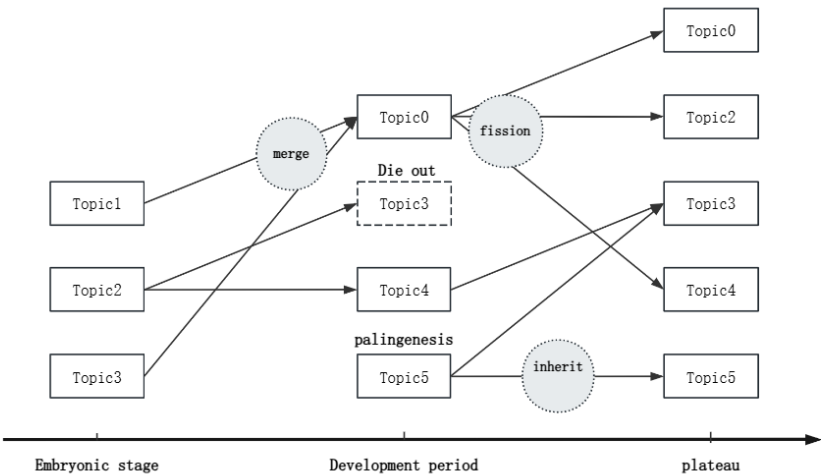


Fig. 3. Thematic evolution relationship.

New theme: The theme of Topic5 in the development period is "E-commerce poverty alleviation", which mainly focuses on the development and application of e-

commerce in poverty alleviation and agricultural assistance, and has become a new topic in the development period of e-commerce.

**Extinction theme:** e-commerce has entered the development period, and the contradiction between talent supply and demand and unbalanced development have become increasingly prominent, which has become the focus of attention in the development period. The theme of "talent training" came from the split of "rural e-commerce" in the embryonic period, but it evolved into the theme of extinction in the slow period

**Inheritance theme:** Inheritance theme: There are four inheritance leading roles in the evolution process of the three stages of rural electricity supplier, the development period Topic4 inheritance embryonic period Topic2, focusing on the development of "rural electricity supplier", and the slowing period Topic3 inheritance development period Topic5, which is the expression of "electricity supplier poverty alleviation" at different stages. Similarly, the theme of inheritance relationship is Topic0 in the development period, Topic3 in the germination period, and Topic0 in the slow period, Topic0 in the development period.

**Merger theme:** The slow phase Topic3 is a combination of the development phase Topic4 and Topic5. Topic4 "rural e-commerce" in the development period has a higher weight of "agricultural products", "Internet", "informatization", and Topic5 "e-commerce poverty alleviation" in the development period has a higher weight of "poor areas", "poor counties", "agricultural products". The subject words of these two themes and the Topic3 "rural e-commerce" have a high coincidence degree, and the similarity is 0.920 and 0.921 respectively.

**Split theme:** the development period Topic0 "e-commerce demonstration county" split into the slow period Topic0 "market supervision", Topic2 "plastic products", Topic4 "supervision and management". Similarly, Topic2 "rural electricity supplier" in the embryonic stage is divided into Topic3 "talent training" and Topic4 "rural electricity supplier" in the development stage, and Topic5 "electricity supplier poverty alleviation" in the development stage is divided into Topic3 "rural electricity supplier" and Topic5 "electricity supplier poverty alleviation" in the slow stage.

## 5 Research Conclusion

According to the topic hot spot identification and evolution analysis of e-commerce policy text, it is found that there are three main theme evolution paths of China's e-commerce policy.

First, rural e-commerce pays attention to rural development. "Rural e-commerce" runs through China's e-commerce policy. Second, e-commerce poverty alleviation promotes rural revitalization. "E-commerce poverty alleviation" is the key theme of e-commerce development period and slowdown period. Third, supervise and regulate the development of e-commerce. "Supervision and management" runs through the whole evolution process of the theme. In addition, there may still be some limitations in the research process of this paper: first, the collection of relevant policy texts may not be perfect, and second, the selection and evaluation of thematic models are insufficient. In

future research, we will expand a wider range of data, achieve full coverage of the research field, and optimize model selection through more innovative methods.

## References

1. Gray P.: Frontiers of electronic commerce. *INFORMATION SYSTEMS MANAGEMENT* 16(1), 84-87 (1999).
2. Wittgreffe P J, Dix C.: The electronic commerce trials platform. *BT TECHNOLOGY JOURNAL* 17(3), 78-86 (1999).
3. Duan S, Lin J, van Dijck J.: Producing new farmers in Chinese rural live E-commerce: platformization, labor, and live E-commerce sellers in Huaiyang. *CHINESE JOURNAL OF COMMUNICATION* 16(3), 250-266 (2023).
4. Huang J, Su L, Huang Q, et al.: Facilitating inclusive ICT application and e-Commerce development in rural China. *AGRICULTURAL ECONOMICS* 53(6), 938-952 (2022).
5. Mei Y, Mao D, Lu Y, et al.: Effects and mechanisms of rural E-commerce clusters on households' entrepreneurship behavior in China. *GROWTH AND CHANGE* 51(4), 1588-1610 (2020).
6. Ning S.: A Study on the Factors that Restrict the Development of China's Rural E-commerce and Countermeasures. *AGRO FOOD INDUSTRY HI-TECH* 28(3), 3664-3667 (2017).
7. Wang L, Dong M.: An Entropy Weight-TOPSIS Based Method for e-Commerce Logistics Service Quality Evaluation. *TEHNICKI VJESNIK-TECHNICAL GAZETTE* 30(4), 1253-1256 (2023).
8. Alshari E M, Azman A, Doraisamy S, et al.: SENTI2VEC: AN EFFECTIVE FEATURE EXTRACTION TECHNIQUE FOR SENTIMENT ANALYSIS BASED ON WORD2VEC. *MALAYSIAN JOURNAL OF COMPUTER SCIENCE* 33(3), 240-251 (2020).
9. Rasiwasia N, Vasconcelos N.: Latent Dirichlet Allocation Models for Image Classification. *IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE* 35(11), 2665-2679 (2013).
10. Church K W.: Emerging Trends Word2Vec. *NATURAL LANGUAGE ENGINEERING* 23(1), 155-162 (2017).
11. Di Gennaro G, Buonanno A, Palmieri F.: Considerations about learning Word2Vec. *JOURNAL OF SUPERCOMPUTING* 77(11), 12320-12335 (2021).
12. Alguliev R M, Aliguliyev R M.: Experimental investigating the F-measure as similarity measure for automatic text summarization. *APPLIED AND COMPUTATIONAL MATHEMATICS* 6(2), 278-287 (2007).

**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.

