Analysis and Research on LDA Model Focusing on Smart Logistics

Zhenyu Zhao[a*], Yuyuan Feng[b], Feiyang Zhang[c]

School of Transportation, Beijing Jiaotong University, Beijing, China

[a*]21251299@bjtu.edu.cn
[b]21231051@bjtu.edu.cn, [c]21251252@bjtu.edu.cn

Abstract. The rapid development of mobile Internet, so that a large number of logistics facilities to achieve interoperability, "smart logistics" concept is increasingly mentioned. The concept of "smart logistics" has been mentioned more and more. The development of smart logistics, combining emerging technologies with basic industries, is conducive to the realization of digitalization, intelligence, efficiency and sharing in the logistics industry. We have crawled the discussions of social groups about "smart logistics" on relevant platforms as samples, calculated the optimal number of topics through the calculation of confusion and consistency scores, and selected the LDA topic model for topic extraction to analyze the extracted data to find out the core hotspots of social groups' concerns about smart logistics.

Keywords: Intelligent logistics; LDA topic model; social platform; hotspots of concern.

1 Introduction

Entering the era of big data, along with artificial intelligence, the Internet of Things and other technologies and the deep integration of the transportation industry, "Internet + logistics" as the characteristics of the smart logistics came into being, smart logistics began to start the development. Currently, the development of "smart logistics" in China is mainly characterized by the following: 1) high synergy. In the logistics system into the big data, logistics network technology, so that the logistics system can receive and process data in real time, and make the optimal decision, to achieve maximum benefits; 2) has a broad market and development space. The new economic model spawned by the e-commerce economy promotes the continuous development of the logistics industry, and at the same time, it also puts forward higher requirements for the intellectualization of logistics, which provides a good environment for the development of smart logistics; 3) fast development speed. The development of smart logistics mainly relies on the development of new science and technology and equipment, artificial intelligence, big data, Internet of Things technology has been widely used in the

© The Author(s) 2024
T. Ramayah et al. (eds.), *Proceedings of the 2024 International Conference on Applied Economics, Management Science and Social Development (AEMSS 2024)*,
Advances in Economics, Business and Management Research 284,
https://doi.org/10.2991/978-2-38476-257-6_41

logistics industry, the future of emerging technologies will also serve as a key support to facilitate the various logistics links.

2 Literature review

2.1 Research on "Intelligent Logistics"

Research on smart logistics has been conducted in the following areas.Alexandra Lagorio, Chiara Cimini, Roberto Pinto, Sergio Cavalieri[1] study the main application areas of 5G in smart logistics and analyze the favorable impacts and challenges faced by 5G technology on smart logistics. Aidi Xu, Fangbin Qian, Huanhuan Ding, and Xuan Zhang[2] use autoregressive distributed lag model (ARDL) to analyze the impact of the development of smart logistics on the logistics industry. Hêriş Golpîra, Syed Abdul Rehman Khan, and Sina Safaeipour [3]propose a general architecture for the L-IoT. Potential goals for future research directions were presented. The impact of digital transformation on the development of the logistics industry is investigated, an assessment model for the application of digitalization in the logistics system is developed, and the model is tested using Salesforce company as an example.

2.2 Research on Text Mining Technology

Twil Ali, Bencharef Omar, Kaloun Soulaimane[4] in their study combined topic modeling and sentiment analysis with manual validation techniques for topic tagging to extract reviews of Marrakech from TripAdvisor, which were analyzed to derive the characteristics of each place. Chaorun Xie, Xiaolong Tian, Xiaochun Feng, Xiaoni Zhang, and Junhu Ruana[5] analyzed consumers' preference in purchasing fresh produce online during the epidemic, used Pycharm software to collect online review texts of fresh produce from online platforms, and pre-formatted the collected data into text preprocessing, which combined with the results of the LDA model to clarify the influence of consumers' Runbin Xie, Samuel Kai Wah Chu, Dickson Kak Wah Chiu, Yangshu Wang[6] analyzed the public's perception of the New Crown Pneumonia outbreak on microblogs using the LDA technique. Jaemin Chung, Jiho Lee, Janghyeok Yoon [7]analyzes social media data to investigate the determinants of customer satisfaction in music streaming services by retrieving reviews of online applications of five music streaming services through topic modeling and text regression.

3 LDA framework and modeling

3.1 Data collection and processing

In this paper, we use a crawler to get the discussions about "Intelligent Logistics" on public discussion platforms such as "People's Daily", "Baidu Post Bar" and "Zhihu", and obtain 6213 valid data. "We obtained 6213 pieces of valid data, and 5742 pieces of valid data by removing duplicates, useless data and incomplete data. The obtained text

is processed by jieba in python, and the deactivated words are removed for training and analysis, which is used for the subsequent work.

3.2 Models and formulas

3.2.1 LDA formula.

LDA model main equation:

$$P(\mu|d) = P(\mu|t) \cdot P(t|d)$$

formula.$P(\mu|t)$ Words under the theme tμ .$P(t|d)$ Topic t under article d. Multiplying the two gives the number of topics.

3.2.2 UCI Consistency Score Formula.

UCICalculation formula:

$$C\text{UCI}(\mu_i, \mu_j) = \log \frac{P(\mu_i,\mu_j)+1}{P(\mu_i)\cdot P(\mu_j)}$$

where$P(\mu)$ is the sliding window in which the μ the probability of occurrence of the word, and $P(\mu_i, \mu_j)$ is the probability that the wordμ_i andμ_j the probability of both appearing in the sliding window.

3.2.3 Confusion level formula.

Confusion level formula:

$$\text{perplexity} = \exp\left\{-\frac{\sum_{d=1}^{D} P(\mu_d)}{\sum_{d=1}^{D} N_d}\right\}$$

whereN_d is the length of the firstd length of the article, and$p(\mu_d)$ denotes the frequency of occurrence of words in the documentd the frequency of occurrence of words in the document.

4 Analysis of concerns

4.1 Theme Topic Identification

4.1.1 Analysis of UCI consistency score results.

The UCI consistency score increases as the number of topics increases. This increase becomes smaller as the number of themes increases. The appropriate number of themes is determined when there is an inflection point where there is no significant increase. Figure 1 allows us to determine that the number of themes is more appropriate at 6 or 11.

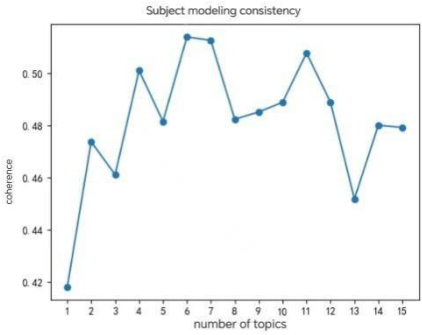

Fig. 1. Thematic Coherence and Number of Topics Curve

4.1.2 Confusion level result analysis.

The degree of confusion will decrease with the increase of the number of topics, and when there is a clear inflection point, it is selected as the appropriate number of topics. Calculating the degree of confusion through python, we can get the first inflection point when the number of topics is 11, and the curve fluctuates smoothly after that, so according to Figure 2, we can determine that the number of topics is 11 is more appropriate.

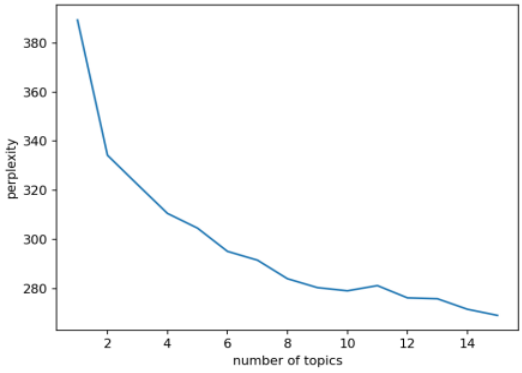

Fig. 2. Curve of Confusion and Number of Topics

4.2 Analysis of Thematic Results

4.2.1 LDA model results.

By analyzing for topic consistency and perplexity. The number of selected topics is 11. after proposing meaningless words. all topics are obtained through LDA modeling. the topics appearing under five topics (Table 1) are analyzed as an example. in giving the topic-keyword graph for Topic1 (Figure 3).

Table 1. Theme-Keyword Matrix

Topic1	Topic2	Topic3	Topic4	Topic5
platform	figure	enterprise	logistics	intelligence
manage	synergia	work	village	robot
data	upgrade	department	business	networking
system	territory	financial	supply chain	equipment
construction	science	industry	e-commerce	solution
regulation	service	talents	index	self-action
information	transition	investigate	transportation	AI
project	economics	time	mail	new energy
synthesize	tax	loans	construction	link
digitization	pilot	publicity	center	voluntarily

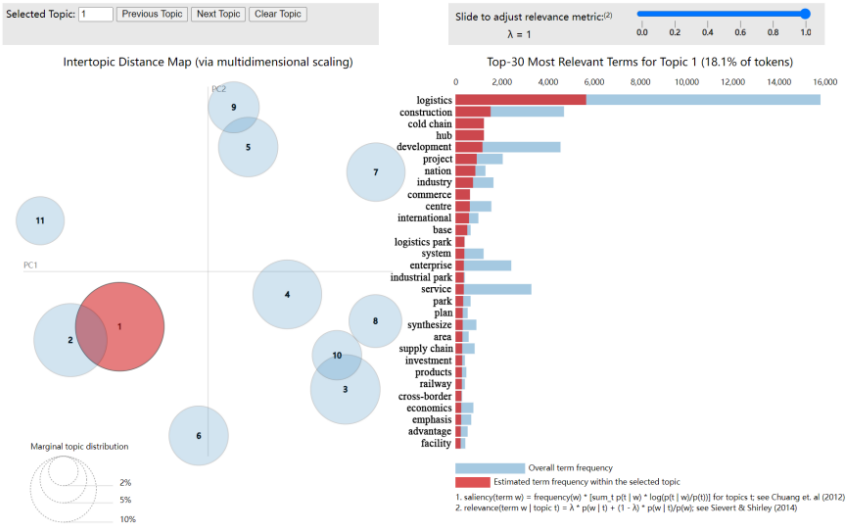

Fig. 3. Topic1 Theme-Keyword Map

4.2.2 Thematic Keyword Analysis.

Combining the subject words to describe the text, taking Topic0 as an example, the related questions mainly involve words such as SF, air, freight, express, etc., which illustrate the main transportation modes of fresh food logistics, including the transportation enterprises involved as well as the transportation modes. According to the same method, the topics of Topic1-Topic9 are described as the operation of logistics management with the words of operation, management, system, etc.; customer evaluation and feedback with the words of service, experience, evaluation, etc.; for the description of the special epidemic period with the words of epidemic, etc.; for the development mode of logistics with the words of mode, industry, platform, etc.; for the national, international and other words to show in the fresh food logistics Showing national policy measures; corporate policy measures mainly by words such as company, enterprise,

etc.; agricultural products logistics mainly by words such as agricultural products, food, etc.; logistics and distribution mainly by e-commerce platforms such as Tmall and Cainiao, etc.; and the theme of storage and transportation characteristics such as cold chain and temperature.

5 Presentation of results

5.1 Word cloud analysis

The LDA model is analyzed and the results are presented through word cloud maps (Figure 4).

Fig. 4. Word cloud analysis diagram

In the word cloud analysis, we can intuitively see that industry, service, logistics, system, structure and data are the focus of public groups for intelligent logistics. After understanding these focuses, we can foresee that the future development of intelligent logistics will be more and more structured and systematic, the industry behavior is constantly standardized, and the service level is constantly improved.

6 Conclusions

6.1 Increase the introduction of advanced talents to ensure the wisdom and stability of all aspects of logistics

Intelligent logistics puts forward higher requirements for all aspects of logistics, along with the continuous promotion of "Internet + logistics", a large number of talent reserves are needed to improve and optimize the intelligent logistics system through high-end technical personnel, as well as to ensure the daily operation and maintenance of the system.

6.2 Introducing smart logistics to multiple industries, realizing a "win-win" situation for both industries and smart technologies

Along with the penetration and development of logistics and digital intelligence in various fields, the introduction and development of smart logistics should also be done in the fields of transportation, medical care, agricultural products, and food freshness and cold chain, so as to realize the new development situation of "smart logistics + transportation", "smart logistics + agricultural products" and "smart logistics + cold chain".

6.3 Combine the public suggestions to continuously improve the new development mode of smart logistics

In the development process of smart logistics, it is necessary to combine the evaluation and opinions of social groups in a large number of ways, so that the combination of emerging technologies and logistics can be responsive to public opinion and convenient for the people. In the subsequent research, the LDA model can be continuously applied to investigate the information of social groups, and through a series of data analysis, the new strategy of industry development that best satisfies the people can be derived to promote the rapid development of smart logistics.

6.4 Strengthening regulatory efforts to purify the industry environment

According to the Topic-Keyword Matrix, the social group's views and future expectations of smart logistics can be derived. Taking Topic1 as an example, the relevant descriptions mainly include words such as platform, management, data, and information, implying that smart logistics is to build a digitalized comprehensive information management platform capable of autonomous supervision. Through generalization and integration, as well as keyword screening and amplification, Topic2-Topic11 can be described in turn as the development goal of smart logistics mainly in terms of synergy, upgrading, and high quality; the development foundation of the smart logistics system mainly in terms of talent, implementation, and training - talent cultivation; and the logistics, supply chain, market-based improvement of the basic areas of the logistics industry; intelligence, technology, robotics, networking-based technology core; agriculture, the elderly, water conservancy-based multidimensional development areas; national, era, people, culture-based development strategy and policy support; epidemic, prevention and control, materials as the core keywords, indicating the great role of the smart logistics in the key specific historical period; service, convenience, Function as the main direction and future expectation of the development of smart logistics, i.e. to improve the quality and level of service; Project, planning, investment as the main words reflecting the preliminary preparation for the development of smart logistics; Global, cooperation, the General Assembly and other key words reflecting the smart logistics can improve the international communication and connection.

References

1. Alexandra Lagorio, Chiara Cimini, Roberto Pinto, Sergio Cavalieri,(2023)5G in Logistics 4.0: potential applications and challenges, Procedia Computer Science, Volume 217, Pages 650-659, https://doi.org/10.1016/j.procs.2022.12.261.
2. Aidi Xu, Fangbin Qian, Huanhuan Ding, Xuan Zhang, (2023) Digitalization of logistics for transition to a resource-efficient and circular economy, Resources Policy,Volume 83, 103616,https://doi.org/10.1016/j.resourpol.2023.103616.
3. Hêriş Golpîra, Syed Abdul Rehman Khan, Sina Safaeipour,(2021) A review of logistics Internet-of-Things: Current trends and scope for future research, Journal of Industrial Information Integration, Volume 22, 100194, https://doi.org/10.1016/j.jii.2020.100194.
4. Twil Ali, Bencharef Omar, Kaloun Soulaimane,(2022) Analyzing tourism reviews using an LDA topic-based sentiment analysis approach, MethodsX, Volume 9, 101894, https://doi.org/10.1016/j.mex.2022.101894.
5. Chaorun Xie, Xiaolong Tian, Xiaochun Feng, Xiaoni Zhang, Junhu Ruana,(2022) Preference Characteristics on Consumers' Online Consumption of Fresh Agricultural Products under the Outbreak of COVID-19: An Analysis of Online Review Data Based on LDA Model, Procedia Computer Science, Volume 207, Pages 4486-4495, https: //doi. org/ 10. 1016/ j. procs.2022.09.512.
6. Runbin Xie, Samuel Kai Wah Chu, Dickson Kak Wah Chiu, Yangshu Wang,(2021) Exploring Public Response to COVID-19 on Weibo with LDA Topic Modeling and Sentiment Analysis, Data and Information Management, Volume 5, Issue 1, Pages 86-99, https: //doi. org/10.2478/dim-2020-0023.
7. Jaemin Chung, Jiho Lee, Janghyeok Yoon, (2022) Understanding music streaming services via text mining of online customer reviews, Electronic Commerce Research and Applications, Volume 53, 101145, https://doi.org/10.1016/j.elerap.2022.101145.

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.