Economics & Business Management
ISSN Print: 3079-5214
ISSN Online: 3079-5222
https://doi.org/10.63313/EBM.9178

# Research on Book Topic Planning Based on Online Review Mining

**Bo Zhang*. Can Lei**
Shanghai University of Science and Technology, School of Publishing, Shanghai 20082, China
**Author Contact:**
Zhang Bo, Room 412, Excellence Building, 516 Jun Gong Road, Shanghai, 20093, China
Email: zhangbo@usst.edu.cn

**How to cite this paper:** Zhang, B., & Lei, C. (2026). Research on Book Topic Planning Based on Online Review Mining. *Economics & Business Management*, 5(2), 107-110. ISSN Print: 3079-5214; ISSN Online: 3079-5222.
https://doi.org/10.63313/EBM.9178
**Published: 2026-04-23**

Copyright © 2026 by author(s) and Erytis Publishing Limited.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/



## Abstract

In the context of digital reading and the growing importance of online reviews for publishing decisions, traditional topic planning faces challenges such as information overload and delayed feedback. This study analyzes over 59,000 valid reviews from Dangdang.com's TOP15 best-selling literary books (April 2025). Using a "text mining + multi-dimensional comparison" approach and an optimized DTC framework with a topic mapping layer, we identify 17 core features of literary books via TF-IDF and K-means. Findings show that keywords like "classic," "moving," and "healing" significantly influence topic decisions and can be directly translated into content, author, and format strategies. The study validates the DTC model's applicability in publishing and provides data-driven support for topic planning.

## Keywords

Online Reviews; Book Topic Planning; Text Mining; Literary Publishing

## 1. Introduction

### 1.1. Research Background and Significance

With the rapid development of digital reading, online reviews have become a crucial reference for publishing topic planning. As of June 2025, China had 1.085 billion online shoppers. However, only 12.7% of literary book reviews contain valid features (Yan Weihua, 2025). Traditional topic planning relies heavily on editor experience and market research, which are subjective and slow. This study explores a data-driven approach to topic planning using text mining.

### 1.2. Research Questions

1. Which online review features are most valuable for topic planning, and how can they be translated into specific strategies ?
2. How do review features differ across sales rankings, authors, and publishers ?
3. How can an analytical framework transform review data into actionable topic

insights ?

### 1.3. Methods

We collected 59,650 valid reviews from Dangdang.com's April 2025 TOP15 literary bestsellers. Using TF-IDF and K-means clustering, we built a feature wordbank and applied a three-stage "feature interpretation→ strategy translation → decision alignment" framework.

## 2. Theoretical Framework and Methods

### 2.1. The Optimized DTC Model

We optimized Yan Weihua's (2025) DTC model by adding a "topic mapping layer," creating a four-layer framework:

**Table 1.** Functional Levels and Analytical Methods of DTC Model Corresponding to Topic Selection Directions

| administrative levels | function | Analysis method (tool) | Topic alignment direction |
|---|---|---|---|
| Data Layer | Raw data collection and screening of topic-related data | Octopus Collector | Focus on data relevant to your topic and exclude irrelevant information such as logistics |
| Label Layer | Feature Extraction and Classification | Word2Vec, K-means clustering | Classify features into three categories: content, author, and format, then align them with the content direction, author selection, and layout design aspects of the selected topic. |
| Comment Layer | Semantic Deep Mining | LDA topic model, context analysis | Extract specific requirements from comments (e.g., "add workplace content") and convert them into topic optimization directions. |
| Topic Mapping Layer | Characteristics-Strategy-Phase Response | Association Analysis, Case Matching | Define the topic selection strategies corresponding to each feature and match them to the specific stages of topic planning. |

### 2.2. Data Preprocessing and Feature Extraction

We cleaned text, performed word segmentation, and filtered parts of speech. TF-IDF was used to weight terms, and K-means clustering (k=80) generated 17 feature clusters, labeled as "classic," "moving," "healing," etc.

## 3. Empirical Results

### 3.1. Feature Wordbank and Topic Mapping

We identified 17 core features across five categories:

| Category | Examples | Topic Strategy |
|---|---|---|
| Content value | Classic, moving, healing | Reprint classics, emotional topics |
| Author/IP | Mo Yan, Shi Tiesheng | IP development, author signing |

| | | |
|---|---|---|
| Audience/scenario | Children, word-of-mouth | School collaboration, cross-media |
| Product format | Binding, printing, price | Deluxe editions, affordable versions |
| Service | Logistics, packaging | Optimize packaging, supply chain |

Each feature maps directly to a phase of topic planning (content direction, author selection, format design).

### 3.2. Sales Ranking Differences

| Rank Group | Key Features | Strategy |
|---|---|---|
| 1–5 (Top) | Classic (3.23%), moving (0.80%) | Classic replication + emotional reinforcement |
| 6–10 (Mid) | No dominant feature (>0.1%) | Differentiation breakthrough |
| 11–15 (Low) | Healing (1.62%), classic (1.08%) | Niche specialization + low-cost trial |

### 3.3. Author-Based Differences

Shi Tiesheng's works dominate the list. Key features: classic/required reading (0.0216%), emotional resonance (0.0064%), philosophical depth (0.0011%). Strategy: IP series development, style-based author selection.

### 3.4. Publisher-Based Differences

| Publisher | Key Features | Strategy |
|---|---|---|
| Hunan Literature & Art | Balanced (classic 1.74%, healing 0.53%) | Multi-value positioning |
| Jiangsu Phoenix Literature & Art | Healing (9.52%) | Niche brand deepening |
| Guangming Daily Press | Balanced but flat (0.87% each) | Mass-market coverage |

## 4. Conclusion and Implications

### 4.1. Key Findings

1. Online review features such as “classic,” “moving,” and “healing” can be directly translated into topic strategies.
2. Top-selling books rely on “classic” features; mid-tier books lack focus; lower-tier books succeed via niche values like “healing.”
3. Publishers exhibit distinct feature profiles, from balanced to specialized.
4. The optimized DTC model effectively supports data-driven topic planning.

### 4.2. Practical Implications

1. Topic direction: Prioritize high-TF-IDF features like “classic and “healing.”
2. Author strategy: Develop IP series for classic authors; assign clear tags to new authors.
3. Format optimization: Use hardcover for “collectible” appeal; adjust for “child-friendly.”
4. Brand positioning: Align publisher strategy with dominant features (e.g., Jiangsu Phoenix → healing).
5. Human-AI collaboration: Use AI for feature extraction, editors for cultural value review.

### 4.3. Limitations and Future Work

Limitations include single-platform data, short time frame, and limited semantic depth. Future research should integrate multi-platform data, apply deep learning (e.g., BERT), and explore AIGC-assisted topic generation.

## Acknowledgements

Project Title: General Project of the National Social Science Fund of China – "Research on Big Data Mining and Utilization in Publishing Based on Knowledge Graphs" (Project Number: 23BXW101)

## References

[1] China Internet Network Information Center. The 47th Statistical Report on the Development of China's Internet Network [R]. Beijing: CNNIC, 2025.

[2] Yan Weihua. Big Data Analysis and Precise Planning of Bestseller Topics [J]. Editor's Friend, 2025,41(3):45-52.

[3] Wang Yi, Cao Yue. Technical Logic and System Construction of Intelligent Book Topic Planning System Empowered by AIGC Technology [J]. Publishing and Distribution Research, 2025,36(5):67-74.

[4] Sang Yun. Research on the Integration and Development of Artificial Intelligence and Book Publishing [J]. Science and Technology Publishing, 2025,44(2):89-96.

[5] Chi Yun. Research on Publishing Topic Decision Model Based on User Profiles [J]. China Publishing, 2025,38(7):32-38.

[6] Luo Xiaowei, Li Xiaoming. An Empirical Study on the Application of Reverse Planning Model in Book Publishing [J]. Publishing Science, 2025,33(4):112-120.

[7] Wang Yujia. Research on the Ethical Dilemmas of Generative Artificial Intelligence in the Publishing Industry [J]. Modern Publishing, 2025,27(3):78-85.