

Research on Overseas Dissemination of the English Translation of *To Live* Based on Sentiment Analysis and Topic Mining

Jinjuan Li

Xi'an University of Technology, Xi'an, Shaanxi, 710000

Abstract: Against the background of today's deepening global literary exchange and translation field, a thorough investigation into the international readers' responses has become an important task. In this context, this study takes one of Yu Hua's masterpieces, *To Live*, as a case study, and systematically examines overseas English-speaking audiences' reviews of the English translation of *To Live* on major international book sales and review platforms. The study covers more than 900 English readers' comments, including multiple steps of data preprocessing, data description, sentiment analysis and topic mining to obtain comprehensive and in-depth conclusions. The findings of the study show that overseas English readers generally hold a positive attitude towards the English translation of *To Live*. The comments focus on various levels such as cultural background, writing style, translation, wisdom and emotional experience conveyed by the work, reflection on the novel's main themes, and recognition of the author. This finding reflects the wide resonance of *To Live* among international readers, and also provides strong empirical support for the global dissemination of Chinese literature.

Keywords: Sentiment analysis; Topic mining; *To Live*; Overseas communication

Introduction

The study of the dissemination of Chinese literature overseas has become a key area of Chinese culture's influence internationally, and is an integral and important part of practice and research. Yu Hua is one of the contemporary Chinese writers with international influence. One of his masterpieces, *To Live*, as a literary masterpiece full of humanistic reflections, profoundly reveals the resilience of human beings in the face of suffering and their desire for survival. Translated into English by American sinologist Michael Berry, the work was published in 2003 by Random House, a renowned American publisher, and has achieved great success all over the world. So far, the translation has been published for twenty years and has accumulated numerous reader reviews on Amazon and Goodreads, accumulating precious corpus for the empirical examination of the translation's overseas dissemination. At present, there are relatively few studies on the effect of overseas dissemination of the English translation of *To Live* focusing on the study of the text, and these papers have explored the influence of factors external to the language on the translation and discussed the influence of ideology on the English translation of *To Live*^[1]. Meanwhile, some scholars have pointed out that when examining the effect of Chinese literature dissemination overseas, there is an obvious lack of research on dissemination efficacy in terms of audience feedback and behavioural effects^[2]. In view of this, this paper collects readers' comments, which are easily neglected in traditional research, as the research object, and based on natural language processing technology, carries out sentiment analysis and theme classification of readers' comments in three different dimensions, namely, time and number of comments, sentiment polarity, and topics of comments, objectively presents readers' acceptance, and explores overseas readers' concern for the English translation of *To Live*, thus enriching the research content and methods of overseas communication for English translated works of literature.

1. Related Research

1.1 Status of research on the English translation of *To Live*

Current research on the English translation of *To Live* focuses on translation strategies or translation comparison studies^{[3][4]}. There are studies on the overseas acceptance of the translation mainly examining it in terms of the number of collections, readers' evaluations, international awards and sales^[1]. At the same time, some scholars have studied the impact of the English translation of *To Live* as part of their research on the English translations of Yu Hua's novels. They looked at the overall effect of translating Yu Hua's novels into English, while also showing how well *To Live* was received in its English version^[5]. Examining from many aspects can indeed show the acceptance effect of

the translated work in an all-round way, but it neglects the attitude and real thoughts of the widest acceptance terminal of the literary work, i.e. the mass readers^[6]. Although the overall examination of multiple works by the same author can examine the overall acceptance of the author's works, it does not give enough examination to a single work. And each translated work will cause different reactions from readers because of different translators or different themes. Therefore, this paper chooses one of Yu Hua's masterpieces, *To Live*, to exclude the influence of theme and translator on the acceptance effect of translation, and to examine in depth the overseas acceptance of the English translation of *To Live*.

1.2 Online comment mining

Online reviews, an important part of online platforms such as social media and e-commerce websites, reflects users' views and attitudes towards the object being evaluated. It can be examined from two dimensions, sentiment analysis and themes^[7]. Sentiment analysis refers to the process of analysing, processing, inducting and reasoning about subjective texts^[8], which is tasked with determining the sentiment polarity of a text, which is generally classified as positive, negative and neutral. Topic mining is the process of discovering topics by using the relationship between text feature items in a text collection^[9]. It is usually done by calculating the probability to mine text topics. At present, there are scholars who have applied relevant techniques to the study of translation effects. Most scholars mainly use sentiment analysis to present the acceptance of translations^{[10][11][12]}. There are relatively few studies that apply both sentiment analysis and topic mining to the effect of translation communication. Among them, Dai, Jianhua, and Minyun Xu et al^[13] examined the overall acceptance of the English translation of *Wandering Earth* and the topics that the comments focused on using a word-based sentiment analysis method and a statistical and word vector-based topic mining method. These studies inspired this paper in terms of research methodology. Therefore, this paper adopts a combination of sentiment analysis and topic mining to carry out this study of overseas readers' comments on the English translation of *To Live*, to visually examine the overseas dissemination effect of its translation.

2. Study Design and Process

2.1 Data acquisition

This study collects online readers' English review of the translation of *To Live* by Michael Berry, from the overseas book sales and review platforms, Amazon.com and Goodreads.com, respectively, by the Supinfor scraper (<https://scraper.supinfor.com/#/>), and saves them in the form of a table. A total of 952 English reviews were collected as of 15 November 2023. By eliminating duplicate and irrelevant comments, a total of 944 valid comments were obtained, totalling more than 92, 000 words.

2.2 Data preprocessing

The following pre-processing was carried out on the data: simple treatment of cases in the tables that would interfere with the subsequent study, deletion of columns and non-English characters that were not valuable for the subsequent study, unification of the time format, etc. The NLTK was used for stemming extraction and lexical reduction of the text, and the stop words corpus was cited for stop word deletion. It is the common corpus used in this type study for English text(<http://www.lextek.com/manuals/onix/stopwords1.html>), which consists of 429 words, is the most widely used English stop word corpus, and finally the data are saved as a whole for subsequent analysis.

2.3 Sentiment analysis

Load the TextBlob library and call the sentiment analysis function to analyze the sentiment of the text. TextBlob is a simple and powerful Python library based on the natural language processing library NLTK. It provides an easy-to-use interface for analyzing the sentiment of text(<https://pypi.org/project/textblob/>). Considering the number of comments collected, this study mainly uses the rule-based method of TextBlob for sentiment analysis. The resulting sentiment polarity values are usually between [-1, 1], with negative values indicating negative sentiment, positive values indicating positive sentiment, and 0 indicating neutrality.

2.4 Topic mining

Text Feature Extraction. The preprocessed text data are converted into vectors using the Term Frequency-Inverse Document Frequency (TF-IDF) method. It is a statistical method for natural language data processing. TF refers to the frequency of a word in an article, while IDF refers to the reciprocal of the frequency of a word in all articles in the entire text corpus^[14]. Using TF-IDF, we can effectively evaluate the importance of the words in the whole text corpus, and filter out the common and irrelevant words, while retaining the important words that affect the whole text.

Topic number determination. The choice of the number of topics directly affects the accuracy and interpretability of LDA topic modeling results^[15]. This study used empirical methods to determine the number of topics, based on existing literature or personal experience, through repeated debugging and observation of the effects displayed by the pyLDAvis visualization tool, followed by manual judgment to determine the number of topics. After several attempts, the number of topics was finally determined to be 4.

LDA Topic Mining. Latent Dirichlet Allocation (LDA) is a technique proposed by D.M. Blei et al. in 2003 to determine the common

topics of a set of documents^[16]. It is believed that each word in an article is classified into a topic by selecting a topic with a certain probability and selecting a word from that topic with a certain probability^[17]. It is the most widely used topic model in current text modeling, which can generate document topics. This procedure uses Python version 3.8 and the results are visualised using the pyLDAvis tool.

3. Findings and Analyses

3.1 Annual publication of reader comments

Figure 1 shows a graph of the trend over time in the number of reviews posted by overseas readers. Since the translation was published in 2003, there has been an overall upward trend in the number of reviews over the years. The number hovered in single digits in the five years just after publication, with a slowdown in the incremental number between 2009 and 2011 and between 2015 and 2018. The number over the years reflects the level of attention paid by overseas readers to *To Live*. The overall increasing trend in the number of reviews shows that it has come into the mainstream English-speaking world to a certain extent, and also has a tendency to keep having an impact on English-speaking readers in the future.

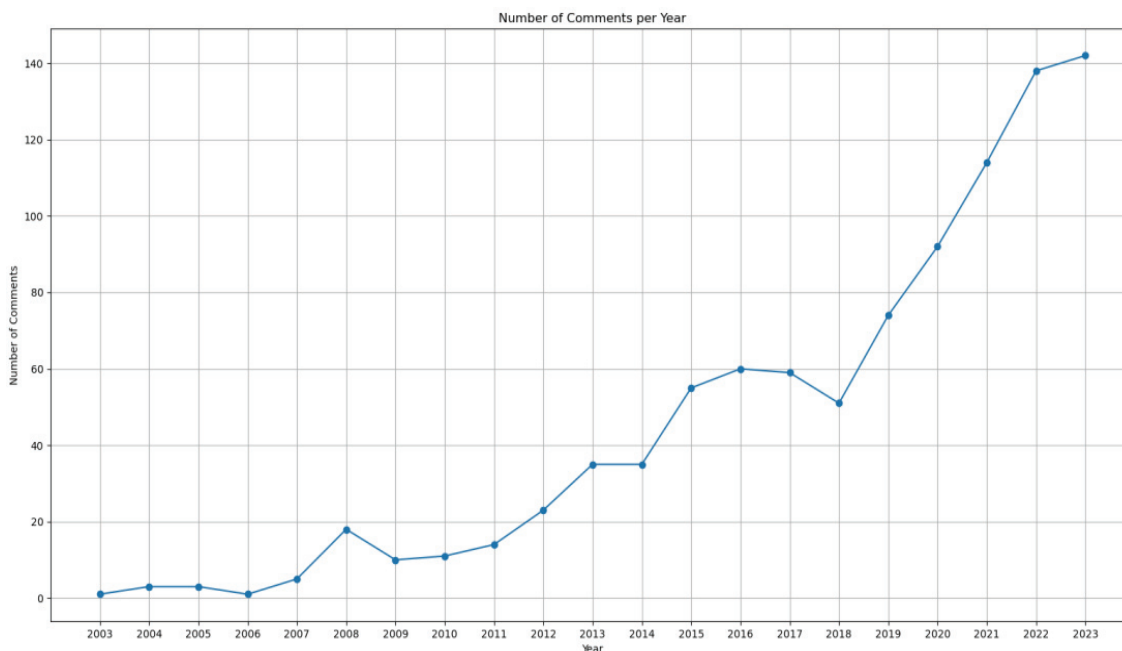


Fig. 1. Annual Trend of Comment Volume

3.2 Sentiment analysis of reader comments

After analysing 944 reader comments for sentiment polarity through sentiment analysis and statistics, the data in Table 1 can be statistically derived. Among the 944 reviews, 720 were positive, 78 were neutral, and 156 were negative. The mean value of positive sentiment evaluation is 0.23, and the mean value of negative sentiment is -0.17. Most of the comments are dominated by positive sentiment, and negative comments only account for 16.53% of the total comments. From the data derived from the sentiment analysis, it can be seen that the English translation of *To Live* has been well accepted by overseas readers, and most of the readers hold positive attitude of the work.

Table 1: Number, Proportion, and Mean of Different Types of Comments

Comment/indicator	Positive	Neutral	Negative
Number	710	78	156
Percentage (%)	75.21	8.26	16.53
Mean value of emotional polarity	0.23	0	-0.17

4. Comment Topic Extraction Results

The results of the visualisation of the topic model are shown in Figure 2. Each of the four circles represents four topics, and the area occupied by the circle represents the importance of each topic in the entire corpus, while the difference in the spacing between the dots of each circle represents the difference between topics. Setting the parameter $\lambda=1$, the top thirty key words with the highest relevance to the topic will appear in the right bar graph in descending order. Taking topic 1 as an example, the most relevant key words include read, book, love, life, and chinese, etc., as shown in Fig. 2.

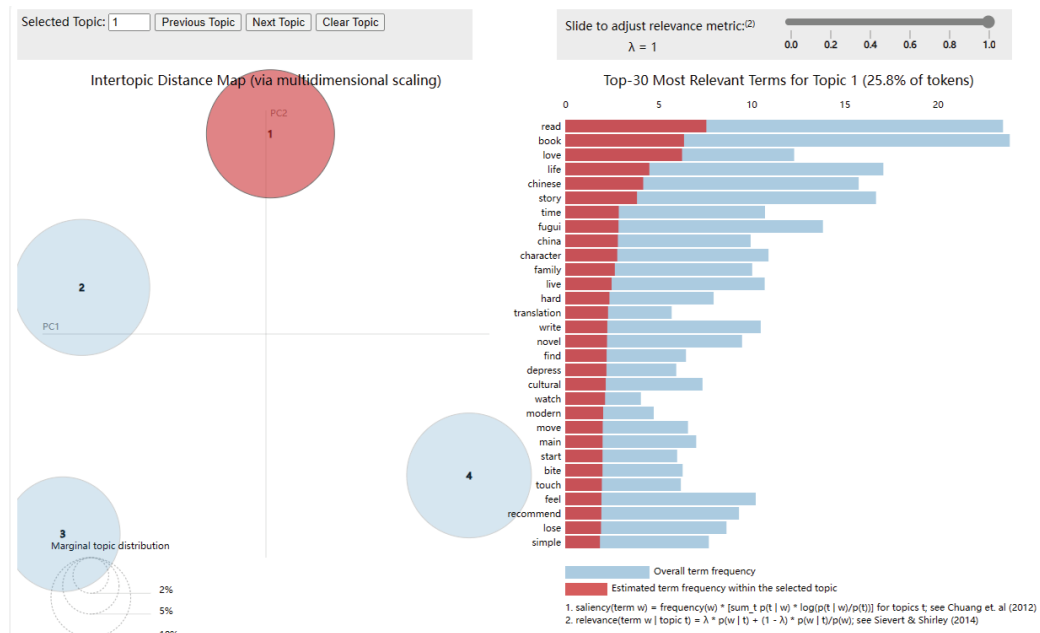


Fig. 2. Inter-topic Distance Chart (Left) and Keyword Bar Chart for Topic 1 (Right)

4.1 Topic mining results

After the model runs, the topics obtained do not have a clear topic name, so it is necessary to combine the keywords of each topic and their logical relationships to name them. Upon careful observation of the running results, it was found that there were duplicates among the keywords of each topic, such as book, read, chinese, story, life, write, live, fugui, character, and family. Other keywords outside of these 10 words demonstrate the uniqueness of each topic. After removing the 10 repeated words, the keywords of each topic are shown in the second column of Table 2.

In Topic 1, words such as cultural, translation, and modern reflects the readers’ attention to the novel’s cultural background and the quality of translation.

In Topic 2, words such as enjoy, sad, and tear indicates that readers’ comments put more emphasis on the emotional experience during the reading process, including joy, sadness, and moving.

In Topic 3, words such as death, powerful, die, bad, and heartbreaking highlights the commenters’ concerns about life and death, strong emotions, and heartbreaking plots.

In Topic 4, words such as tolive, yuhua, tragic, and experience reflect readers’ concerns related to the author as well as the novel’s main theme.

Eventually, four theme names, culture and translation, emotional experience, life and strong emotions, and author and theme, were abstracted, as shown in the first column of Table 2. Table 2 demonstrates the topic names, the corresponding 20 key words, and the number and distribution probability of texts under each topic.

Table 2: Theme-Keyword Table

Topic name	Topic keywords	Text number	Distribution probability (%)
Culture and translation	love, time, china, hard, translation, novel, find, depress, cultural, watch, modern, move, main, start, bite, touch, feel, recommend, lose, simple	252	26.69
Emotional experience	enjoy, hard, movie, novel, time, sad, finish, people, beautiful, feel, tear, simple, lose, main, short, china, like, tell, move, love	267	28.28
Life and strong emotions	sad, feel, death, recommend, favorite, time, china, touch, powerful, remember, die, simple, cry, happen, emotional, star, bad, heartbreaking, love, tell	215	22.78
Author and theme	lose, cry, novel, recommend, people, feel, tolive, china, time, cultural, yuhua, star, beautiful, love, tragic, tell, experience, easy, finish, start	210	22.25

4.2 Analysis of topics

Topic 1 - Culture and translation. Some of the comments under this topic express readers' appreciation for the cultural background of

the novel. These commenters believe that the cultural background elements are a highlight in the story and are deeply engaging. Some readers demonstrated their views on the novel's writing style, and were disappointed by the simplicity and incongruity of the text they read, stating that it was difficult to tell whether the incongruity was caused by the culture or the translation. There were also comments expressing disapproval of the quality of the translation. This group of readers felt that there were some emotions and details missing in the translation. Although some readers said that the translators did a good job in conveying the tone of voice, they felt that there were some problems in the translation of some words. Taken together, comments under this topic expressed love for the cultural elements of the novel, but criticized and worried about the writing style and translation quality.

Topic 2 - Emotional experience. Some of the texts under this topic expressed appreciation of the wisdom contained in the novel, as well as satisfaction at being able to learn about Chinese history through reading the book. Some of the comments emphasized the sadness experienced during the reading process, which is closely related to life itself, with one comment even describing the novel as one of the saddest works ever read. In addition, some of the reviews highlight the moving and complimentary nature of the novel's wonderful depictions, focusing mainly on emotions, life and family. These reviewers believe that the novel is moving because of the way in which it is written, and that the novel is a good example of the way in which it is written. These commentators believe that the novel is touching because it is able to express through words a profound reflection on the resilience of the human will. Synthesising these emotional experiences, it can be seen that the English translation of *To Live* has produced far-reaching and diverse emotional resonance among readers, ranging from joy at the wisdom of *To Live* and learning about the historical background through reading the novel, to the emotional experience of deep sadness.

Topic 3 - Life and strong emotions. The comments under this theme mainly dealt with readers' concerns about life and death, strong emotions and heartbreaking plots. Some of the comments expressed that the book triggered readers to think deeply about the subject of life and death, and that the novel's powerful expressive effect made people rethink the meaning of life. Some comments also highlighted the readers' deep experience of the emotional expression of the novel and their recognition of the profound expressive effect of the words during the reading process. The commentators felt that the story's expression was simple yet powerful, and the deep emotional description moved them. In addition, some reviews also focused on the novel's heartbreaking plot development, including the irony embedded in the novel's name and how death can change a person. The reviewers' in-depth analyses of the novel's plot show that they have a keen interest in the story. Taken together, these reviews show that the reviewers' experience of the English translation of *To Live* focuses mainly on the reflection on life and death, the profound experience of emotions, and the attention to the heartbreaking plot.

Topic 4 - Author and theme. The comments under this theme mainly focus on readers' concerns about the author as well as the novel's main idea, which can be divided into the following three categories: 1. Recognition of the author. This category of comments mainly expresses recognition of Yu Hua, including his engaging writing style and excellent character portrayal ability. 2. Views on the main idea. This part of the comments highlights the readers' understanding of the novel's main idea. They believe that this is a historical novel with tragic effects, realism, and unique characteristics. At the same time, the comments mention that the novel incorporates humorous elements, which sometimes appear light-hearted, showing the diversity of readers' understanding of the novel's theme. 3. Impact of the literary work on the experience. This section of the comments emphasised the positive impact that readers had on their lives through reading the novel. Some readers mentioned that reading Yu Hua's works while living in China made them interested in Chinese culture, and exclaimed that reading the novels brought him a deep and wonderful reading journey, which expresses the recognition of the positive emotions brought by the novels. Summarizing these comments, it can be seen that readers have positive views on the author's professional level, understanding of the main theme, and the impact of literary works on life, emphasizing the richness of the novel in both literary and emotional aspects.

Taken together, reader comments on the English translation of *To Live* are complex and varied. Despite criticisms and concerns, positive views of the cultural elements included in the translation, the emotional experience it brings, and the views of the book's author are shown, demonstrating the profound impact on the literary and emotional levels that the English translation of the novel has had on the overseas readership.

5. Conclusions

This study investigates the overseas dissemination effect of the English translation of *To Live* based on reader reviews on Amazon.com and Goodreads. Through an in-depth study of the overseas reader reviews of the English translation of *To Live*, the author found that the number of reviews of the translation has been on an upward trend since its publication. More than 70% of the readers expressed positive views on the English translation of *To Live*. The review themes focus on a number of levels, including the storyline, culture, emotional experience, translation and other aspects. First of all, the cultural background became one of the hotspots of comments. Readers were greatly interested in the presentation of the specific cultural background shown in the novel, reflecting the success of the translation work in conveying the cultural elements. Secondly, the writing style and translation quality are as another concern, showing the readers' high sensitivity to the textual expression and the translator's merit. The wisdom and deep emotional experience conveyed by the novel are also the focus of the comments, show-

ing that readers are not only attracted by the story itself, but also deeply inspired by its connotations in the process of reading. At the same time, the reflections on the novel's main theme and the discussion of the author in the reviews also reflect the readers' deep understanding of the work and their respect for the author. These reviews present a multi-layered and multi-dimensional reading experience, revealing the fascinating aspects of the work on different levels.

The successful dissemination of the English translation of *To Live* overseas is not only the success of a literary work in translation, but also a brilliant display of Chinese literature on the international stage. This study provides multi-level insights for understanding the power of cross-cultural translation and the deeper meaning of literary works. The author expects that this study on the dissemination and acceptance of Chinese culture will further promote the internationalisation of Chinese literature, bring more wonderful works to the world, and promote mutual understanding and exchange of cultures.

6. Data Availability

The author does not have permission to share data.

References

- [1] Shi Chunjian, Zhao Huifang. A study of the relationship between English translation strategies and translation effects in *To Live*[J]. Language Education, 2020, 8 (04): 63-67.
- [2] Hu Anjiang. Study on the Assessment of the Effect of Chinese Literature's Overseas Dissemination--Taking the English Translation of Chinese Literature by American Sinologist Watson as an Example[J]. Shanghai Translation, 2023, (02): 73-78+95.
- [3] Zhu Zhenwu, Yang Leipeng. Bai Ruiwen's Translation Aesthetics and Cultural Commitment--Taking the English Translation of *To Live* as an Example[J]. Foreign Languages, 2016, 32 (03): 89-94.
- [4] Li Hanji. From Subtext to Text: Bai Ruiwen's View of Translation and the Interpretation of English Translation of *To Live*[J]. Journal of Guangdong University of Foreign Studies, 2017, 28 (01): 63-69.
- [5] Fu Yue, Wu Yun. A study on the translation effect of Yu Hua's novels in the United States with the aid of corpus[J]. Journal of Anhui University (Philosophy and Social Science Edition), 2021, 45 (02): 34-45.
- [6] Li Jie, Wei Jiahai. Overseas Acceptance Study of English Translation of Life and Death Fatigue Based on Online Book Reviews[J]. Journal of PLA College of Foreign Languages, 2022, 45 (03): 144-151.
- [7] Cheng Wenting, Wu Jiahui. A study on the theme mining and sentiment analysis of users' opinions on digital reading for preschool children[J]. Library Construction, 2022, (03): 104-112.
- [8] Zhao Yanyan, Qin Bing, Liu Ting. Text Sentiment Analysis[J]. Journal of Software, 2010, 21 (08): 1834-1848.
- [9] MIKOLOV T, SUTSKEVER I, CHEN K, et al. Distributed representations of words and phrases and their compositionality. in: Advances in Neural Information Processing Systems. New York: Curran Associates, Inc. 2013: 3111-3119.
- [10] Zhang Lu. Overseas readers' acceptance and evaluation of Chinese translated literature from Python sentiment analysis: the English translation of Three Bodies as an example[J]. Foreign Language Studies, 2019, 36 (04): 80-86.
- [11] Shi Chunjian, Deng Lin. An examination of the translation and dissemination effects of regional culture-loaded words in the English translation of Jia Pingwa's Happiness based on emotional analysis[J]. Research on Foreign Languages, 2020, 6 (06): 58-66.
- [12] Li Jie, Wei Jiahai. Overseas Acceptance Study of English Translation of Life and Death Fatigue Based on Online Book Reviews[J]. Journal of PLA Foreign Language College, 2022, 45 (03): 144-151.
- [13] Dai Jianhua, Xu Minyun, Zeng Daojian. Research on the foreign communication of Chinese science fiction works in the perspective of digital humanities: Taking Wandering Earth as an example[J]. Foreign Languages and Cultures, 2023, 7 (01): 91-101.
- [14] Liu Jianghua. A text retrieval method based on kmeans clustering algorithm and LDA topic model and validation of effectiveness[J]. Intelligence Science, 2017, 35 (02): 16-21+26.
- [15] Zhang Dongxin, Zhang Min. A review of research progress in the application of LDA topic models in the field of graphical intelligence[J]. Book Intelligence Knowledge, 2022, 39 (06): 143-157.
- [16] BLEI Dongxin, NG A Y, JORDAN M I. Latent dirichlet allocation[J]. Journal of machine learning research, 2003, 3(1): 993-1022.
- [17] Chen Xiaomei, Gao Cheng, Guan Xinhui. An LDA topic modelling approach for online opinion viewpoint extraction[J]. Library and Intelligence Work, 2015, 59 (21): 21-26.