Big Data Analysis of the Formative Factors of Text Meaning in Ancient Chinese Literature of Higher Education

Jie Hang(✉)

Xian FanYi University, Xian 710105, Shaanxi, China

Abstract. The research on the meaning of ancient literary texts has always been a hot topic in literary criticism. Traditional criticism holds that the meaning of the text should be traced back to the original meaning of the author when he created the text, and the author's meaning is the meaning of the text. Under the background of big data, it is of great significance to analyze the generative factors of the text meaning of ancient literature in higher education.

Keywords: Ancient literature · Text meaning · Particularity · Generative factors

1 Introduction

What kind of thoughts and ideas can be realized only with the help of a certain text form, so this form has become an important means and carrier of human ideological and cultural communication after word of mouth. The formation of this form depends on language and characters. Since with the help of character symbols, both western alphabetic characters and Chinese characters have direct and obscure language features. Any text is not entirely meaningful, which leaves room for different readers to understand it differently. In this way, since the birth of the literary text, it has been a long and continuous process theory of literary text accepted, understood and interpreted by people. Its own existence is not the key, but the important thing is to explore the language intention hidden behind the form of the text. In western literary theory, the research on this aspect belongs to the special field of hermeneutics [1]. Of course, this paper does not intend to follow the footsteps of many western theoretical schools, but only hopes to use some theoretical achievements of Western hermeneutics to provide some help for exploring the meaning generation of ancient Chinese literary texts.

2 Common Data Mining Algorithms

The so-called frequent pattern mining refers to the commodity sets found in the commodity transaction database records. The frequency of these commodity sets is higher than a threshold. These frequent commodity sets are called frequent patterns. The idea

M. A. Jan and F. Khan (Eds.): BigIoT-EDU 2021, LNICST 392, pp. 165–174, 2021.
https://doi.org/10.1007/978-3-030-87903-7_23

of frequent patterns is very simple. First of all, the number of occurrences of each single commodity is counted, which constitutes a dimension table. Then, according to the one-dimensional table, a two-dimensional table is generated by pairwise combination of commodities.

The so-called relationship mining, is worth mining out the causal relationship between the various projects [2]. The basis of relation mining is frequent pattern mining. It's easy to get relations through complex pattern mining. For example, we get a frequent set. The typical data mining algorithm is shown in Fig. 1.

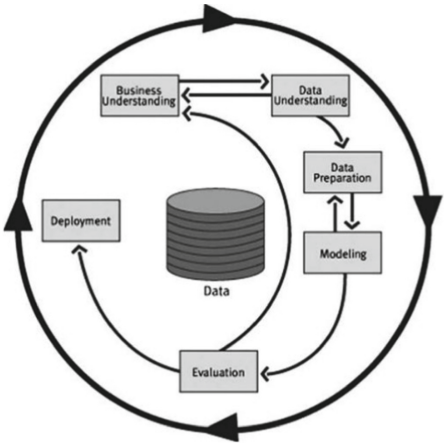

Fig. 1. The typical data mining algorithm

3 The Particularity of Chinese Ancient Literature Text

If we regard the generation of text meaning as a process of understanding, then for us, people's understanding process starts from the text symbol system, and then gradually integrates its associative factors on this basis. As far as ancient Chinese literary texts are concerned, the function of the Chinese character expression system to readers' understanding mainly lies in two aspects.

3.1 The Imagery of the Structure of Chinese Characters

Character is the most basic factor of the existence of text, which is "the writing symbol of language, the conventional visual signal system of information exchange between people". These symbols should be able to flexibly write the language composed of sound, so that the letter can be sent to distant places and to future generations. "The universal characters can be divided into phonetic characters and ideographic characters, both of which can record the language in the form of symbols, so that the information can be preserved and the limitation of time and space can be broken [3–5]. Compared with the western alphabetic characters, Chinese characters have very special ideographic

characteristics, especially the vast majority of them are the combination of sound, form and meaning. The ancients paid attention to the description of the physical characteristics at the beginning of the word making, so that people can know the meaning of the hieroglyphs such as "water".

From the perspective of information acquisition, human beings get more information from the outside world through the visual symbol system. Compared with the western alphabetic characters, the Chinese characters with rich physical characteristics have a large amount of written Chinese information. Therefore, they are more suitable for the implicit and rich style of font shape, and even can associate with the principles and objectives of life, as sung in the first popular song.

$$l = \{I1, I2, I5\} \tag{1}$$

Then the subset set can be obtained by permutation and combination:

$$\{I1, I2\}, \{I1, I5\}, \{I2, I5\}, \{I1\}, \{I2\}, and\{I2\} \tag{2}$$

In relation mining, a very useful relation pattern mining: the relation patterns of miassoc:

$$A_{quan1} \wedge A_{quan2} \Rightarrow A_{cat} \tag{3}$$

3.2 The Duality of Written Language

Chinese people pay attention to "harmony between man and nature", which is the highest realm they pursue all their lives. Even if they can't climb this height, they also strive for natural harmony. They have the philosophy of unity of opposites and the aesthetic principle of "unity of opposites" and attach importance to the "symmetry and harmony" of things in pairs. Some scholars attribute it to thinking. This characteristic has a very important impact on the formation of the whole Chinese traditional culture. As for the literary creation we are going to talk about, just as the Chinese people pay attention to symmetry and harmony in art design, they are also striving for a kind of balance and harmony in the complicated strokes and radicals of Chinese characters [6–8]. For example, the knots of many Chinese characters in "Tanaka" are axisymmetric, There are also some slightly different but generally symmetrical left and right sides - the structure of these Chinese characters is symmetrical.

With the Chinese people's symmetrical paintings and pianbian, they are striving for balance and harmony. For example, most of them are axisymmetric, and some of them are slightly different on the left and right sides, We can feel a special way of thinking and aesthetic of Chinese people from it. This symmetry and balance is fully reflected in the creation of characters.

4 Research Results

4.1 Statistics of Topics

Through the vectorization of the original document and the analysis of the structural topic model, 15 research topics of AI education are obtained. Each topic is made up of

high ratio keywords of AI education, which are arranged from big to small, and they represent the main meaning of this topic. Each root will also be distributed in different topics with different ratios, which represents the degree of their thematic meaning in different topics. The topic of AI education is expressed by the matrix of topic word "and" word topic". Through the synthesis of the meaning of words with a high proportion of topics and the review of relevant literature, two peer experts are invited to assign a topic to each topic after many consultations.

4.2 Topic Model Estimation

Taking the literature on "artificial intelligence + education" as the analysis data source, the country and time of the literature were set as the influencing factors of the change of topic ratio. The regression equation constructed by 15 topics was estimated by using the estimated effect function of structural topic model, and the uncertainty parameter in the model was set as "Goba". Table 2 shows the p value of each topic under the influence of country and time ($P < 0.001$ is very significant, $P < 0.05$ is significant, $P < 0.1$ is not significant). From the perspective of topic preference, there are significant differences between China and the United States except topic 7, topic 8 and topic 15. From the time development of the topic, there is no significant change in the intensity of all topics in the process of time development.

4.3 Topic Preference Estimation

Estimating the relationship between metadata and topic is the core advantage of structural topic model. The mapping function of estimating effect can deal with the results of estimating effect of artificial intelligence education topic model. For the differences between China and the United States, researchers can use the difference option to map the changes of topics from one specific value to another. Using the structural topic model package (STM), taking the country as the covariate, the two values of the covariance parameters are set to China and the United States respectively, and the specific names of 15 topics are specified. Finally, the topic preference diagram of "artificial intelligence + education" research in China and the United States is drawn, as shown in Fig. 1. Chinese researchers mainly focus on four topics: Games and intelligent agents, intelligent teaching system, intelligent medical and nursing, and educational intelligent software [9, 10]. American researchers mainly focus on knowledge management system, educational robot, intelligent research field, educational intelligent technology, and machine learning, The number of "Ai + education" topics preferred by American researchers is significantly more than that in China.

4.4 Topic Content Comparison

The analysis of preference differences between Chinese and American researchers for a certain topic content can reveal the specific content of research preference in detail. The content preference difference analysis of the same topic can be realized by using the content (CO η ten) parameter covariates of the structural topic model. First, we estimate

the structural topic model with country covariates, and then use the plot function to draw a comparative diagram of each topic between China and the United States. Researchers can analyze 15 topics one by one and conclude the preference differences between China and the United States in the same topic.

5 The Theory of Text Meaning in Ancient Chinese Literary Theory

5.1 "The Distinction Between Words and Meanings"

In the philosophy of language in this historical period, "speech" refers to the meaning of expressing speech, while "meaning" refers to the origin or fundamental law of all things in the universe (equivalent to the purpose, destination and realm of showing life in Laozi and Zhuangzi's Philosophy), More often, people are used to analyze it from the perspective of philosophy; but as for the ancient Chinese literary texts, it has become a deep analysis of the relationship between words and meanings in literary discourse [11–13]. The proposition of "Confucius is not complete, not full of meaning" in Zhouyi · Xici is the first of its kind, and a series of discussions on "words" and "meaning" are launched.

5.2 "Against the Will with the Will"

The understanding of "Zhi" has always been controversial. Some people think that "Zhi" is a verb, which means "record", that is, the record of historical facts. However, the author thinks that although these three views are different, they are not as incoherent as the understanding of "meaning". They can be integrated together, and they do not have essential differences. No matter how it is interpreted, "Zhi" can be attributed to the original intention of the author or the work. Of course, it is not enough to have "intention" and "ambition", but also to be able to "reverse". The literal understanding of "inverse" is to trace and explore what "Zhi" is; in fact, it is also equivalent to the process of "fusion of Horizons" in western theories. Readers need to eliminate the distance between themselves and the text and the author in time and space, and organically integrate their own "realistic horizon" and the author's "initial horizon", so as to grasp the exploration of "Zhi".

5.3 The Interaction Effect of Topic

It is estimated that the topic intensity of AI + education research is closely related not only to country or time, but also to the interaction effect of country and time. The conclusion of the topic model provides an explanation for the influence of country on topic intensity. In the process of structural topic model fitting, the product of country and time is taken as the independent variable of topic intensity, and then the estimated effect function is used to estimate and draw the fitting trend chart of each topic intensity one by one [14]. As shown in Fig. 2. Taking topic 13 "intelligent medical and nursing" as an example, this paper uses the effect estimation function to describe the adjustment effect of countries on the temporal change intensity of topic.

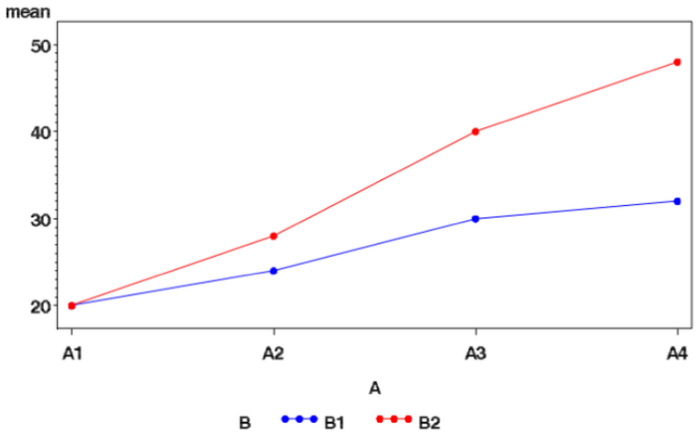

Fig. 2. Interactive effect of topic

6 Analysis and Discussion

By analyzing the structural theme model of "artificial intelligence + education" literature in China and the United States, this paper reveals that researchers in China and the United States mainly focus on 15 topics, including educational robots, educational intelligence software and intelligent teaching system. According to the model of "artificial intelligence + education" and the topic preference model of Chinese and American researchers, the differences between Chinese and American topics are mainly reflected in the number and content of topics. The differences of topic contents are mainly reflected in the comprehensiveness of the research on "artificial intelligence + education" and the differences of their interests, involving educational artificial intelligence software, artificial intelligence teaching and learning, intelligent learning evaluation, educational robot and other topics.

6.1 Topics of Common Concern

The development and application of artificial intelligence education technology is a topic of common concern of researchers in China and the United States. There is no significant difference in the three topics of knowledge management system, intelligent processing system and machine learning. The topic of knowledge management system mainly focuses on the knowledge representation algorithm and model construction of artificial intelligence system, The purpose of this paper is to explore how agents can solve complex problems, improve reasoning ability or help students understand complex behaviors through knowledge management, knowledge representation algorithm and complex system understanding, such as knowledge representation and extensible computing object network design of knowledge-based system, structural behavior function model for complex system understanding, intelligent complex problem solving based on metaphor Algorithm, knowledge understanding based on golden ratio and genetic algorithm, interactive multimedia representation development system, cognitive

activities and reasoning based on purposive topological reasoning theory. The topic of intelligent processing system mainly focuses on manufacturing, semantic understanding, sharing decision and effective learning. Intelligent manufacturing is the key to improve the quality of processing process [15, 16]. Intelligent manufacturing is realized through automatic video monitoring, knowledge reasoning to process safety, human-machine physical system and optimization, It can also be used as a framework or ruler to understand the impact of semantic technology on individuals and society, design educational courseware, design the process and content attributes of decision-making system, and measure the effect of human use of cognitive resources. Machine learning topics mainly focus on machine learning process, data mining, student learning performance and learning evaluation supported by artificial intelligence. As far as machine learning is concerned, intelligent machine can read and understand unstructured text through problem feature oriented attention mechanism.

6.2 Estimation of Country Moderating Effect

6.2.1 Overall Differences in Topic Content

First of all, there are significant differences in the number of topics and research scope of AI + education between China and the United States. The research scope of the topic "artificial intelligence education" in the United States is more comprehensive, more specific and more profound than that in China, covering 11 topics such as educational artificial intelligence technology, machine learning, educational robot, knowledge management, machine learning and learning evaluation, However, Chinese researchers only focus on educational intelligence software, educational games, intelligent medical and nursing, intelligent teaching system and other topics. There are also great differences in the research topics of "artificial intelligence + education" between China and the United States [17]. The United States is inclined to the innovation of artificial intelligence education technology, especially the research of machine learning, educational robot and intelligent knowledge management system; in contrast, Chinese researchers pay more attention to the teaching and learning application of artificial intelligence education technology, such as educational games, teaching agents, intelligent medical treatment and intelligent teaching system.

6.2.2 Content Comparison of Single Topic

The content covariates in the structural topic model can realize the difference preference comparison between China and the United States for each topic content, and the high proportion words of China and the United States for each topic are drawn by using the plot function. As follows, we will analyze the difference words of China and the United States in each topic one by one, The review of educational AI robots is a topic of common concern of Chinese and American researchers, covering travel, medical care, intelligent education and super AI. The United States pays more attention to the development of artificial intelligence and risk prevention strategies, while it is difficult for China to identify the high proportion of research preferred words. From the perspective of Topic 2 "educational AI", American researchers pay more attention to the topic content than Chinese researchers. They mainly focus on the purpose and value of AI education from a

macro perspective. It is difficult to distinguish the topic content that Chinese researchers pay attention to. From the content of Topic 3 "learning and education", researchers in China and the United States pay almost the same attention to topic content. Chinese researchers tend to pay attention to the influence of AI on learners, online learning, resource recommendation and learning performance, while American researchers tend to focus on teaching design, learning tasks, interactive conversation and learning experience supported by AI. The comparison of high-frequency words in topic 4 learning evaluation shows that Chinese researchers mainly focus on neural network, intelligent algorithm and evaluation accuracy, while American researchers mainly focus on intelligent evaluation, event prediction and measurement model. There are significant differences between China and the United States in the topic of educational intelligence technology [18–20]. The United States pays particular attention to learner learning supported by artificial intelligence technology, while Chinese researchers focus on specific research fields such as intelligent education, development and innovation, and intelligent platform.

7 The Generative Factors of the Text Meaning of Ancient Chinese Literature and Simulation Analysis

In the process of generating the meaning of ancient literature text, some factors play a fixed role in the whole process of readers' reading, which mainly refers to the text itself. The text has its own inherent stipulation. "Once it gets rid of the speaker's immediacy, the text can go beyond the historical, psychological and sociological limitations of the speaker listener context." The text is a kind of permanent existence, which is independent of the perception of the acceptor. Its existence does not depend on the aesthetic experience of the acceptor, and its structure will not change because of people's events [21–24]. "In other words, once the text is formed, it will be out of the author's" control", all the text symbols that make up it have been determined, and the meaning of the text on the level of these text symbols will also change At the same time, it is fixed. Regardless of the author's original intention and the reader's participation in understanding activities, the structure, rhetorical devices, stylistic style and writing skills of the text itself are fixed.

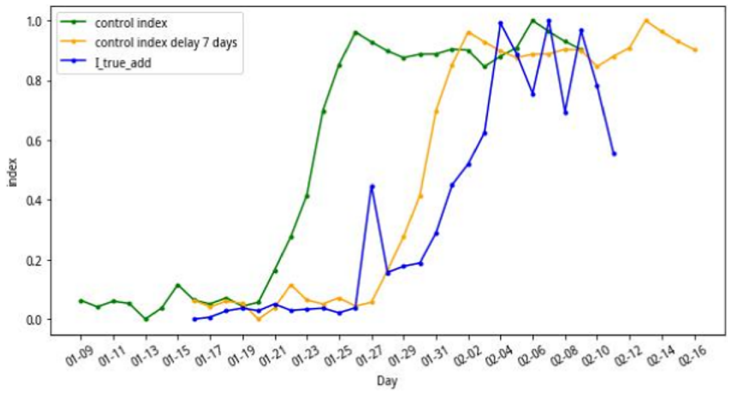

Fig. 3. Text generation with bigdata

We do the simulation as shown in Fig. 3. From Fig. 3, we can see that with the gradual enhancement of the algorithm, that is, with the passage of time, the generated words and words also show an upward trend, which shows that the algorithm proposed in this paper has played a certain effect.

8 Conclusions

In this paper, the author always emphasizes that the generation of text meaning is a dynamic and complex "generation" process, and tries to analyze the influence of various factors on the readers' understanding of the text meaning item by item through the systematic framework of "decoding the text from the readers" and combining the comprehensive factors such as reading situation, etc., The meaning of ancient Chinese literature text is relatively ignored in the generation. The system they set up is generally aimed at all texts. There is no special research on my unique text. We can't take the ancient Chinese text by its number in many of their theoretical frameworks. There are many unique text nouns and symbols between the lines. Therefore, the significance of this paper is to try to re-establish a system to analyze the generation process of text meaning, which can be used to study the meaning generation of ancient Chinese literary texts in the current environment on the premise of ensuring that the generalized texts are established.

Acknowledgements. Supported by Xi'an Translation Research Team (XFU18KYTDC01).

References

1. [song] Zhu, X.: Notes on the Chapters and Sentences of the Four Books Notes on the Analects of Confucius. Beijing: Zhonghua Book Company (1995)
2. Xiuyan, F.: Textology: a Systematic Study of Textualism [M], p. 12. Peking University Press, Beijing (2004)
3. Gadamer, H.: Shanghai: Shanghai Far East publishing house (2003)
4. Pan, D.: Character, interpretation and tradition: the modern transformation of Chinese interpretation tradition. Shanghai: Shanghai Translation Publishing House, October 2003
5. Brief introduction to the war narrative of Murakami's works in the war memory crisis of contemporary Japanese literature. J. Ningbo Univ. (Hum. Ed.) **33**(05), 133 (2020)
6. Chen, Y.: Practice of digging the theme meaning of text from the perspective of deep learning. Basic Foreign Lang. Educ. **22**(04), 76–82+108–109 (2020)
7. The intertextuality construction and significance expression of Xi Li Pumi funeral ceremony "giving sheep Zi" and the text -- Taking nansongyuan, Lanping County, Yunnan Province as an example. J. Kunming Univ. **42**(04), 102–107 (2020)
8. Guikui, L.: From the perspective of intertextuality the jidacheng of the dream of Red Mansions. Chin. Cult. Res. (03), 84–93 (2020)
9. Ling, Y.: Story, ability and Metaphor: a brief discussion on the mutual text structure and meaning interpretation of Lu Xun's the moon running. J. Taiyuan Univ. (Soc. Sci. Ed.) **21**(04), 60–65 (2020)
10. Dong, Y., Guo, S.: Exploring visual text meaning in English teaching in the internet environment. J. Higher Educ. **25**, 48–51 (2020)

11. Liyazhi, G.: Re mention of feelings: experience aesthetics and text interpretation. J. Xi'an Univ. Archi. Technol. (Soc. Sci. Ed.) **39**(04), 88–93 (2020)
12. Lu, J., Chen, K.: Context parameters, textual interpretation and meaning confirmation – on the constraints of context on interpretation. Philos. Res. **08**, 90–97 (2020)
13. Pang, H.: Boundary of interpretation and hermeneutics of literature and art of hesch. Foreign Lit. (03), 1–10+156 (2020)
14. Qiao, Z.: The teaching model of college English reading in higher vocational colleges based on the theory of personal cognition. Overseas Engl. **15**, 157–158 (2020)
15. Xu, Y.: The double construction of compound borrowing, time form and text internal and external meaning -- narrative research of the mystery of the beloved Van Gogh and the stars. New film works, (04), 72–75 (2020)
16. Hongju, Z., Shi, J.: The analysis of the cultural communication path of Hongsao in the perspective of coding decoding theory. J. Party Sch. Jinan Municipal Committee Communist Party China **04**, 96–99 (2020)
17. Yuan, Z.: Innovative design of primary school Chinese literacy teaching with Chinese. Sch. Educ. **20**, 38 (2020)
18. Zhang, W.: Exploring the significance of multi text reading in the teaching of Chinese reading in senior high school. New Curriculum **33**, 67 (2020)
19. Zhao, L.: Barry brumet's communication thought research. Shandong University (2020)
20. Shiqun, S.: Acceptance of classical poetry and literature: an hermeneutics observation of literature text teaching in normal university. J. Guangxi Norm. Univ. Sci. Technol. **35**(04), 94–97 (2020)
21. Li, Y.: The penetration of Chinese traditional culture in the lower grade Chinese texts and its educational significance . Prose Hundred (New Chin. loose Page) **08**, 190 (2020)
22. Fu, F.: Analysis of the guiding strategies for the deep reading teaching reform in junior high school Chinese. Read. Writ. Calculation **23**, 115 (2020)
23. Wang, J., Yuchunxia, H., Wufengyun, Chen, L., Suli, H.: The significance of the same book heterogeneity of teachers is significant -- the text characteristics and Class Innovation Discussion in the whole book teaching perspective (8). Ref. Chin. Teach. Mid. Sch. (23), 4–8+20 (2020)
24. Baiwu, J.: Textual interpretation and contemporary significance of historical records in Jiangnan under KangQian. Knowl. Libr. (16), 9+193 (2020)