Analysis of College Students' Consumption Behavior Based on LDA Thematic Modeling

Shaoning Sun [1,a*], Jiayi Guo[2,b], Yunxiang Bai[3,c], Jijia Yang[4,d], Wenyi Dai[5,e], Runyu Jia[6,f]

{sunsn@cau.edu.cn[a*], 2021505460102@cau.edu.cn[b], sddzbyx@cau.edu.cn[c], 2022308310229@cau.edu.cn[d], 2021505460130@cau.edu.cn[e], 2018505460109@cau.edu.cn[f]}

Yantai Institute of China Agricultural University, Yantai City, Shandong Province 264670, China [1*]
China Agricultural University, Beijing 100083, China [2,3,4,5,6]
Jiayi Guo and Yunxiang Bai contributed equally to this work and should be considered co-first authors.

Abstract: In the present consumer market, there is an increasing focus on college students as the primary consumers and influencers of consumption trends. Utilizing the LDA model, this article examines the text and theme clustering of data gathered on college students' consumption behavior. Logistic regression is employed to affirm that the disparity in students' consumption is not primarily triggered by gender or grade. Additionally, solutions are proposed to address the current irrational consumption patterns among college students, with an emphasis on promoting the concept of responsible consumption.

Keywords: LDA model; rational consumption; logistic regression; sampling; Consumption guidance

1 Introduction

The LDA model [1] is a popular method for text analysis that has been widely used in recent years to uncover hidden topic structures in text data. This technique assumes that each document is composed of a blend of multiple topics. By examining the frequency of word occurrences in the text, the model can infer the distribution of topics within each document, as well as identify common words associated with each topic. The LDA model can effectively address various text-related issues, such as text classification[1], topic extraction[2], and modeling user behavior[3][4][5]. It has significantly contributed to reducing information redundancy in text processing, enhancing text readability and comprehensibility, and playing a vital role in numerous fields.

With the ongoing progress of the Internet sector, the amount of information accessed by university students has experienced a tremendous increase, consequently leading to a transformation in their consumption patterns. Currently, research and analysis on the spending habits of college students mainly revolve around the percentage of expenditures on food[6][7][8], alterations in consumption behavior during and after the COVID-19 pandemic[9][10], and the establishment of a sustainable consumption environment[11][12][13].

Shaoning Sun: Corresponding author. Associate Professor (1975-), mainly engaged in research in the direction of agricultural informatization

There is a lack of research on comprehending the psychological traits, behavior, and structure of contemporary college students' consumption. This research aims to guide and assist college students in developing sensible and accurate consumption concepts, overcoming misconceptions about consumption, and establishing sustainable consumption habits. Ultimately, the goal is to promote healthy consumption habits. Given that college students are a key demographic and play a significant role in shaping consumption trends, studying their consumption behavior holds immense value. To delve deeper into this subject, we conducted searches using keywords such as "college students' consumption" and "college students' preferences" on platforms like "Zhihu" and "Baidu online shopping mall". From the 383 relevant results obtained using the "college students' preferences" keywords, we analyzed and categorized the most popular terms. Our findings revealed that 43% of college students are primarily concerned with the sense of psychological gap and adaptability changes, while 36% focus on changes in their consumption behavior. This includes alterations in consumption psychology, consumption products, and consumption demands. The LDA model, which can automatically identify themes from extensive text data, offers a flexible and adaptable approach to analyze complex and diverse consumption behaviors of college students. It provides an unsupervised learning method based on data, aiding in understanding consumption behavior patterns and serving as a reference for related research and decision-making. Therefore, we have chosen to analyze college students' consumption behavior using the LDA model.

2 Building LDA models with topic clustering

The theme of "College Students' Consumption Psychology" encompasses various aspects of college students' attitudes, motives, and behaviors during the consumption process. By selecting this theme, we aim to gather insights and experiences from diverse groups of college students, enabling us to create a more comprehensive and diverse research text. To begin, we collated data from the Zhihu platform and converted it into written form. Subsequently, the text underwent cleaning, with punctuation and special characters removed, and deactivated words excluded. Additionally, existing dictionaries were utilized, and a thesaurus was constructed to perform jieba segmentation on the text content. After eliminating low-frequency and high-frequency words, 100 feature words were extracted to construct a document-vocabulary matrix. This matrix was then inputted into the LDA model for modelling, and the optimal number of topics was determined through topic confusion analysis[14]. In our study, we utilized the confusion calculation method to evaluate the effectiveness of the chosen number of topics. We applied Eq.(1) on the Zhihu "college students' consumption" fixed text set to acquire the necessary data and conducted the calculation. Among the various number of topics tested, we observed that the topic perplexity performed optimally when the number of topics was 3. In this formula, M represents the size of the test corpus (i.e., the number of documents), N denotes the size of the text (i.e., the number of words or tokens), and w denotes the relevant documents.

$$P(\overline{W}|M) = \prod_{m=1}^{M} p(\overline{\overline{w}}_{\overline{m}}|M)^{-\frac{1}{N}} = \exp - \frac{\sum_{m=1}^{M} \log p(\overline{\overline{w}}_{\overline{m}}|M)}{\sum_{m=1}^{M} N_m} \tag{1}$$

Next, the article "psychology of consumption among university students" on Zhihu platform was analyzed and grouped into different categories, visually represented by a word cloud (Figure 1). The LDA keywords were then sorted into three main themes: sources of income,

reasons for consumption, and factors influencing consumption (Table 1), leading to the creation of a consumer profile of university students. By examining these three themes, it becomes evident that the discussions among university students on Zhihu mainly revolve around their living expenses, the proportion of their spending, and the factors that influence their consumption choices, with a particular emphasis on the proportion of spending. The word cloud effectively illustrates the most frequently mentioned keywords, providing a clear picture of the consumption patterns among university students. The central words, such as "living expenses","university","saving money," "roommate," and "game equipment," align well with the identified themes derived from the text. In terms of word frequency, the top-ranking keywords account for , of the total, demonstrating their significant relevance and offering an accurate portrayal of the consumption habits specific to university students.

Fig. 1. Word cloud of the text "Consumption behavior of college students"

Table 1. College Student Consumer Theme Sheet

Thematic	Byword
Economic sources	Scholarships, Parents' salaries, Labor, Part-time jobs,Loans, Family, Living expenses
Consumption purpose	Exams, Gaming equipment, Social spending, Clothing, Cosmetics, Laptops, Lodging fees, Tuition, Meals
Consumption influencing factors	Girls, Roommates, Family, Impulse spending, Environment, Grade level, Rights awareness

3 Data analysis of college students' consumption behavior

3.1 Research Plan and Methodology

This paper presents an analysis of the consumption behavior among college students using the LDA model. To provide a more objective and accurate understanding of the consumption patterns among college students, a sampling of consumption behaviors and a survey on consumption psychology were conducted on three grades of students at China Agricultural University. The survey was administered in dormitories, resulting in 534 valid questionnaires. The analysis primarily focuses on three dimensions of college students' consumption behavior: composition, level, and psychology. Quantitative analysis was employed for data processing. Closed-ended questions were summarized and statistics were calculated using Excel 2016 to

determine the proportion of respondents for each option. For open-ended questions, answers were categorized and analyzed accordingly. Cross-analysis and data processing were conducted using different questions, and the resulting data were represented in charts and graphs. The data was then correlated and analyzed using SPSS to draw systematic conclusions and provide specific recommendations[15].

3.2 Research content and data analysis

In terms of consumption, online shopping APP has gradually become an important means of consumption for college students with its convenient, fast and rich products. According to the survey results, college students' consumption is mainly on e-commerce platforms such as Taobao and Pinduoduo, accounting for 67.24% and 19.83% of the total number of students respectively. Moreover, with the help of big data and personalized recommendation system, online e-commerce platform APP can capture students' needs more quickly, save consumers' shopping time, and make targeted suggestions and guidance according to consumers' needs. At the same time, since there is no need to pay for the rental cost of physical stores, the stores are more attractive to college students, who mainly rely on their parents' living expenses as their financial basis, because of their low prices.

Regarding the composition of consumption, the survey data reveals that over 60% of students allocate a higher proportion of their spending towards healthcare, web-based learning, and sports and recreational activities. Additionally, students also devote a significant share of their expenses to supplementary learning resources. In recent times, the pressure to secure employment has gradually intensified, leading most fresh graduates and current college students to invest more funds in exams, public assessments, and professional training courses to enhance their competitiveness. Presently, college students exhibit dissatisfaction with traditional teaching methods, hence they actively seek improved educational resources through diverse avenues such as online video platforms and offline educational counselling classes. Concurrently, the education industry actively responds to students' numerous demands by constantly developing and enhancing its offerings, thereby augmenting and expanding the sector's structure, and effectively fostering increased consumption among students in this domain.

Regarding the degree of consumption, the survey findings indicated that 61.34% of university students spent an average ranging from 1,000 to 1,500 yuan. Through in-depth interviews, it was observed that the majority of college students perceived an enhancement in their quality of life. This can be attributed to two key factors: firstly, the steady growth in China's economic development, along with the successful triumph over poverty, which has resulted in a sustained rise in students' families' income. Consequently, the family stands as the primary economic source for students' expenditures, and the improvement in family earnings further contributes to the elevation of students' consumption levels. Additionally, the impact of inflation, causing price hikes, has also propelled the level of student consumption. However, a small subset of students with extremely low consumption levels was identified during our interviews, a large proportion of whom possessed meager and inconsistent incomes.

In addition to these three aspects, this paper also explores and examines the impact of gender on consumption. Consumption behavior among college students displays distinguishable disparities. Female college students tend to be more emotional and enjoy making comparisons.

They are highly influenced by their surroundings. If they see their classmates buying fashionable clothes, they tend to disregard their own financial situation and impulsively purchase similar items. This behavior indicates that female college students have an emotional and casual approach to consumption. On the other hand, male college students exhibit a more rational approach. They are capable of calmly handling the consumption patterns of their peers and spending according to their own means. Additionally, the influence of age on college students' consumption is also crucial. Freshmen, who are unfamiliar with the university environment, are more susceptible to blindly following the herd compared to their senior counterparts. Based on previous research, this project suggests that there is a certain disparity in consumption behavior based on gender and age. This may be due to the fact that males and females, as well as different grade levels, have varying consumption patterns. Therefore, this paper employs SPSS statistical analysis software to establish a binary Logistic regression model for college students' overall consumption. In this model, a value of 0 represents a monthly consumption below 1,500, while a value of 1 represents a monthly consumption exceeding 1,500. Gender is represented by a value of 0 for females and 1 for males. Grade level is represented by the values 1, 2, 3, and 4, corresponding to the first, second, third, and fourth years of college, respectively. The output results are presented in Table 2 below.

Table 2. Summary Of Results From Binary Logistic Regression Analysis

Item	Regression coefficient	Standard error	Z-value	P-value	OR 95% CI
Sex(s)	-0.826	0.666	-1.241	0.215	0.119 ~ 1.614
Grade(g)	-0.389	0.273	-1.424	0.155	0.397 ~ 1.158
Intercept(i)	1.936	0.738	2.622	0.009	1.631 ~ 29.468

From the above table, it can be seen that the binary Logit regression analysis with gender,, grade as independent variables and monthly consumption as dependent variable from the above table implies that gender, grade explains the reason for 0.08 change in monthly consumption. From the above table (where p represents the probability that the monthly consumption is 1, represents the probability that the monthly consumption is 0[16]). The final analysis shows that the regression coefficient of gender is -0.826, but it does not show significance , which means that gender does not have an impact on monthly consumption. The regression coefficient value of grade is -0.389, but does not show significance , which means that grade does not have an impact on monthly consumption. Analysis shows that gender, and grade all do not have a large impact on the relationship between monthly consumption[17]. This suggests that prior research has failed to effectively elucidate the consumption disparity solely based on gender and age as independent variables. Additionally, this indicates that Logit models for low-density development may encounter a dimensional catastrophe when confronted with high-dimensional data. Excessive features lead to escalated training and computational complexities, thus rendering the model susceptible to overfitting. Moving forward, we will delve deeper into investigating additional factors influencing monthly consumption among college students and employ the stepwise elimination method to derive a precise regression equation.

4 Correct guidance for university students' consumption behavior

As a group in the transitional stage between family and society, college students' lives primarily revolve around practicing and learning, with limited prior experience before entering the workforce. Regarding the consumption psychology of college students, a survey indicates that approximately 27% of them possess a strong desire to consume. These students tend to indulge themselves in various activities and typically belong to a consumption bracket of over 1,500 yuan. Conversely, 73% of students display rationality in their consumption behavior, exercising emotional control and avoiding impulsive expenditure. Furthermore, 70% of college students utilize advanced spending platforms to expand their personal spending capacity. Among these individuals, 34.6% employ Ant Check Later and borrowing, while the remaining proportion resorts to alternative micro-lending platforms such as Jingdong Baitiao. Through extensive research, it was revealed that college students predominantly allocate their advanced spending towards non-essential items such as trendy footwear, high-end cosmetics, and branded smartphones. External factors easily influence students' consumption psychology, often leading them to purchase items beyond their financial means. This phenomenon, commonly referred to as "advance spending," establishes a detrimental cycle, perpetuating further instances of such behavior. Thus, it is imperative to establish a healthy consumption mindset among college students and provide appropriate guidance to foster responsible consumption behaviors.

To foster and guide college students in developing a strong understanding of consumption, the following measures can be implemented: Firstly, it is vital to promote the concept of rational consumption among college students. By cultivating the practice of rational consumption and discouraging mindless pursuit of fashion and wasteful spending, students can shape their consumption behavior appropriately. Next, campuses can arrange financial literacy training programs and offer courses on consumption planning to continually enhance students' knowledge about consumption and their financial management skills. It is essential to strengthen financial management as well. Schools can provide deliberate financial subsidies and distribute them fairly among students, ensuring that each student receives an appropriate amount of financial assistance. Moreover, the institution can introduce features such as consumption records and analysis to enable students to practice independent and standardized financial management. The establishment of a consumption supervision system is necessary. Schools can form a dedicated management team responsible for monitoring students' consumption behavior. This team can educate and caution students against overspending and other undesirable behaviors, while also imposing penalties for any violations, effectively maintaining order in students' consumption habits. Building stronger ties with families is crucial as well. Students' families should actively participate in their financial management, collaborating with schools to promote proper consumption concepts, reinforce family education, and establish a comprehensive approach to managing students' consumption through home-school cooperation. Lastly, efforts should be directed towards enhancing the quality of consumption and service standards. Schools should allocate additional funds and resources to improve the quality of student consumption and the level of service. Implementing preferential policies to encourage student consumption is essential, while ensuring that businesses strictly adhere to relevant service standards and quality requirements to establish an effective and efficient student service system.

5 Conclusion

Through this paper, we examine the consumption behavior of college students and identify various factors that influence it, such as economic level and gender. By analyzing the themes of consumption discussions, we identify different topics related to economic sources, consumption purposes, and factors impacting consumption behavior. These themes are closely tied to the overall atmosphere and needs of consumption behavior among college students. Moreover, we investigate the motivations behind college students' consumption behaviors, including social identity, self-expression, and personal reality needs. Utilizing text mining technology, we conduct an in-depth exploration of the characteristics and patterns observed in college students' consumption behavior. Our findings reveal not only gender-based differences in consumption patterns but also differences in consumption preferences among students from various majors. However, we acknowledge certain limitations and issues in our study. For instance, we recognize that the questionnaire used as a data source introduces a certain degree of subjectivity. Additionally, there are other social consumption factors, such as psychological and cultural aspects, that could potentially influence consumption behavior but were not fully encompassed in our analysis. Therefore, moving forward, we aim to enhance the rigor of our research methodology and data sources to attain more precise and objective conclusions.

References

[1]   Li W, Sun L, Zhang D K. Text classification based on labeled-LDA model[J]. CHINESE JOURNAL OF COMPUTERS-CHINESE EDITION-, 2008, 31(4): 620.

[2]   Nikita M. Select number of topics for LDA model[J]. CRAN R Project, 2016.

[3]   Xu G, Zhang Y, Yi X. Modelling user behaviour for web recommendation using lda model[C]//2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology. IEEE, 2008, 3: 529-532.

[4]   Moghaddam S, Ester M. ILDA: interdependent LDA model for learning latent aspects and their ratings from online product reviews[C]//Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval. 2011: 665-674.

[5]   Wan C, Peng Y, Xiao K, et al. An association-constrained LDA model for joint extraction of product aspects and opinions[J]. Information Sciences, 2020, 519: 243-259.

[6]   Kamenidou I C, Mamalis S A, Pavlidis S, et al. Segmenting the generation Z cohort university students based on sustainable food consumption behavior: A preliminary study[J]. Sustainability, 2019, 11(3): 837.

[7]   De Piero A, Bassett N, Rossi A, et al. Trends in food consumption of university students[J]. Nutricion hospitalaria, 2015, 31(4): 1824-1831.

[8]   Bernardo G L, Rodrigues V M, Bastos B S, et al. Association of personal characteristics and cooking skills with vegetable consumption frequency among university students[J]. Appetite, 2021, 166: 105432.

[9]   Gómez-Galán J, Martínez-López J Á, Lázaro-Pérez C, et al. Social networks consumption and addiction in college students during the COVID-19 pandemic: Educational approach to responsible use[J]. Sustainability, 2020, 12(18): 7737.

[10]　White H R, Stevens A K, Hayes K, et al. Changes in alcohol consumption among college students due to COVID-19: Effects of campus closure and residential change[J]. Journal of studies on alcohol and drugs, 2020, 81(6): 725-730.

[11]　Valor C, Antonetti P, Merino A. The relationship between moral competences and sustainable consumption among higher education students[J]. Journal of Cleaner Production, 2020, 248: 119161.

[12]　Waris I, Hameed I. Promoting environmentally sustainable consumption behavior: an empirical evaluation of purchase intention of energy-efficient appliances[J]. Energy Efficiency, 2020, 13(8): 1653-1664.

[13]　Degli Esposti P, Mortara A, Roberti G. Sharing and Sustainable Consumption in the Era of COVID-19[J]. Sustainability, 2021, 13(4): 1903.

[14]　LIAO Huiping,LIU Jiaxuan. Exploration of Discourse Expression of "Heartwarming Offer" Based on LDA Thematic Model[J]. New Media Research, 2022,8(16):112-115.

[15]　Zhang, Xingming. A survey report on consumption of frozen food during the new crown pneumonia epidemic[J]. Modernization of Shopping Malls,2021(01):8-10.

[16]　HUANG Wenze,GUO Chenglong. A study on the influence of gender on college students' online shopping tendency[J]. E-commerce, 2020(08):95-96.

[17]　Zhang Jing. Regression Analysis of Influencing Factors of Grassroots Employment Intention--Taking the Rural Student Source of Loop College as an Example[J]. China Market, 2022(07):31-35+104.