

Research on Exploring the Potential of Diversified Traditional Chinese Medicine Culture in Xinjiang based on Binary Logit Regression Model, K-prototypes Clustering Algorithm, and Structural Equation Model

Weitao Qin*

Department of School of Politics and Public Administration (National Security School),
Xinjiang University, Urumqi, China

*Corresponding author: 3210295167@qq.com

Abstract

This article takes Urumqi as the research object, comprehensively using methods such as questionnaire surveys, semi-structured interviews, and text mining, combined with binary Logit regression models, clustering algorithms, and structural equation models, to deeply explore the development potential of Xinjiang's diverse traditional Chinese medicine culture. Research has found that although Urumqi has achieved certain results in the development of traditional Chinese medicine culture, such as improving public awareness and making breakthroughs in international expansion, there are also problems such as policy implementation and demand disconnect, supply-demand mismatch, and cultural dissemination gap. To this end, it is necessary to establish a precise publicity mechanism and government enterprise cooperation platform, strengthen design compliance innovation and doctor awareness, break through the European and American markets and carry out cultural differentiation output, strengthen the construction of new media matrix and the integration of ethnic medicine, optimize market supply to reach the youth group.

Keywords

Diversified Traditional Chinese Medicine Culture in Xinjiang, Binary Logit Regression Model, K-prototypes Clustering Algorithm, Structural Equation Model.

1. Introduction

Xinjiang's diversified traditional Chinese medicine culture is an important part of the treasure house of Chinese medicine. It is not only an important link to strengthen the awareness of the Chinese national community, but also a key carrier to deepen the "the Belt and Road" health cooperation. Urumqi, as a core node city of the Silk Road Economic Belt, relies on characteristic medicinal resources such as Tianshan snow lotus and *Cistanche deserticola*, and integrates the wisdom of ethnic medicine such as Uyghur medicine and Kazakh medicine to form a full industry chain ecology of "medicinal planting cultural inheritance health care services international dissemination". However, existing research mostly focuses on the macro level, with insufficient attention paid to the regional practice and micro mechanisms of traditional Chinese medicine culture in Xinjiang.

The role of traditional Chinese medicine culture in regional development is increasingly prominent. Li Candong (2011) [1] proposed that the localization of traditional Chinese medicine health literacy should be combined with the concept of "harmony between heaven and man", while Qian et al. (2023) [2] emphasized the need to balance traditional and modern needs in cross-cultural communication. In the study of Xinjiang's diversified traditional Chinese

medicine culture, Zhang Yong and Zhang Yanxia (2023) [3] pointed out through SWOT analysis that the development of traditional Chinese medicine in Xinjiang faces the challenges of talent gap and western medicine dominance, but the "the Belt and Road" policy and geographical advantages provide strategic opportunities. Zhao Panzhen et al. (2019) [4] noted the need to address terminology translation barriers through digital technology. Li Zhongkai and Li Jinye (2019) [5] emphasized that Xinjiang should rely on the "Healthy Silk Road" strategy and leverage the geographical advantages of traditional Chinese medicine. Yu Jinrong et al. (2022) [6] also pointed out that Chinese consumers have a high sense of identity and belonging to Chinese culture. For example, Zhang Xiaoqin and Liao Zhaoyu (2016) [7] believe that Xinjiang has the potential to develop tourism related to traditional Chinese medicine and culture; Hailiang (2020) [8] emphasizes that the integration of traditional Chinese medicine and tourism can promote cultural heritage; Wang Liqiong et al. (2018) [9] suggested using new media to expand the dissemination of Mongolian medicine, while Gong and Gao (2020) [10] constructed a performance evaluation system for Mongolian medicine cultural tourism projects, providing reference for their scientific development. Xinjiang ethnic medicine has both cultural value and industrial potential, but faces challenges such as inheritance gaps and brand deficiencies. We need to build a sustainable development path through the integration of culture and tourism, policy support, and new media communication.

However, existing research lacks empirical analysis of Urumqi, especially in terms of policy evaluation, market supply optimization, and international expansion paths, which lack systematic exploration. This study takes Urumqi as a sample and focuses on the "Regulations on Traditional Chinese Medicine in Xinjiang Uygur Autonomous Region". Through a mixed research method, it fills the gap in micro perspectives and provides a solution that combines theory and practice for the development of traditional Chinese medicine culture in Xinjiang.

2. Data Source

The study adopted a multi-source data fusion method: 1. Questionnaire survey: covering 6 administrative districts of Urumqi City, 1026 valid questionnaires were obtained through stratified sampling and three-stage PPS sampling. The sample covered 19 ethnic groups, mainly aged 26-45 (52%), and the occupation was mainly enterprise employees and public servants (61%); 2. Semi structured interviews: Conduct in-depth interviews with the heads of 10 units including the National Health Commission, pharmaceutical companies, and medical institutions, focusing on policy implementation pain points and market supply and demand contradictions; 3. Text mining: Using Python crawlers to collect 43888 traditional Chinese medicine culture related blog posts on Weibo platform, 89630 comment data were obtained. After cleaning, 42759 valid blog posts (cleaning retention rate of 97.43%) and 76131 valid comments (cleaning retention rate of 84.94%) were retained. Combined with sentiment analysis and word frequency statistics, public cognitive characteristics were mined.

3. Specific Model

3.1. Binary Logit Regression Model

This model is suitable for situations where the dependent variable is a binary variable, and is used to explore the influence of the independent variable on the probability of event occurrence. (1) Data preparation and descriptive statistics: Collect 1014 valid samples, with the dependent variable being "purchase intention" (0=none, 1=present), distributed as 64.2% vs. 35.8%. Excluding 268 missing data, the proportion of valid samples is 79.1%. (2) Model construction and testing: Include independent variables - health products, cultural and creative accessories, medicinal herbs and drugs, specialty foods, and specialty drinks. The overall effectiveness of

the model was evaluated through likelihood ratio test (chi square=99.00, $p<0.05$), confirming the significance of the model. Calculate AIC (1309.674) and BIC (1399.904) as the benchmark for model comparison.(3) Regression analysis and hypothesis testing: Use Wald test to determine the significance of each variable ($\alpha=0.05$), calculate the OR value and its 95% confidence interval. Collinearity diagnosis ($VIF<1.02$) eliminates multicollinearity interference. Calculate the marginal effect (dydx) to evaluate the impact of variable unit changes on purchase probability.(4) Model validation and optimization: Evaluate prediction accuracy (overall 65.98%) and classification error rate. Use Hosmer Lemeshow test (specific values not provided) to determine the goodness of fit of the model.

Table 1. Significance and Influence Direction of Variables

Independent variable category	regression coefficient	p-value	OR value	95% CI	Marginal effect (dydx)
Health Products	0.612	0.000	1.84	[1.41, 2.40]	0.137
Medicinal herbs and drugs	0.289	0.032	1.33	[1.03, 1.75]	0.064
Speciality Foods	0.283	0.036	1.33	[1.02, 1.75]	0.063
Cultural and Creative Accessories	0.164	0.224	1.18	[0.91, 1.55]	0.036
Specialty drinks	0.007	0.963	1.01	[0.76, 1.33]	0.002

Note: $p<0.05$, $p<0.01$; The marginal effect value is the effect at the mean.

The significance and direction of influence of variables are shown in Table 1; Overall performance of the model: Likelihood ratio test: $\chi^2=99.00$, $p<0.001$, significantly better than the intercept only model; Explanatory power: McFadden $R^2=0.024$, Cox&Snell $R^2=0.031$, indicating limited explanatory power of the model; Prediction accuracy: Overall 65.98%, with an accuracy rate of 92.93% when the true value is 0 and an error rate of 82.37% when the true value is 1; Collinearity diagnosis: All independent variables have VIF values <1.02 , tolerance >0.98 , and no multicollinearity issues.

3.2. Improved K-prototypes Clustering Algorithm

Clustering algorithm selection: K-prototypes algorithm: suitable for mixed data types (numerical+categorical), by combining the advantages of K-means (processing numerical variables) and K-modes (processing categorical variables), a comprehensive distance measure is defined:

$$D(X, Y) = \alpha \sum_{j=1}^p (x_j - y_j)^2 + \beta \sum_{j=p+1}^m \delta(x_j, y_j) \quad (1)$$

Among them, α and β are weight coefficients, and δ is the difference function of categorical variables. Improvement point: Introducing contour coefficient and Calinski Harabasz index to dynamically optimize the number of clusters, solving the problem of traditional methods relying on empirical settings.

Data preprocessing: Standardization: Z-score standardization is used for numerical variables such as age and income to eliminate the influence of dimensionality; Classification variable encoding: Convert classification variables such as "appearance design preference" and "brand

awareness" into exclusive hot codes; Missing value handling: Multiple imputation method is used to fill in missing values to ensure data integrity.

Table 2. Summary of Basic Information of Cluster Categories

Cluster category	frequent and continuous	percentage
Young middle-income high-frequency buyers	314	36.09%
Middle aged, high-income, low-frequency buyers	270	31.03%
Middle aged and high-income high-frequency buyers	286	32.87%
amount to	870	100%

Definition of high-value group: Based on three indicators: comprehensive purchase frequency (≥ 3 times/month), unit price per customer ($\geq$ top 30% in the industry), and repurchase rate ($\geq 50\%$), a high-value scoring model is constructed: $\text{Score} = 0.4 \times \text{frequency} + 0.4 \times \text{unit price per customer} + 0.2 \times \text{repurchase rate}$ $\text{Score} = 0.4 \times \text{frequency} + 0.4 \times \text{unit price per customer} + 0.2 \times \text{repurchase rate}$.

Data loading and cleaning: Import 870 user data, remove 12 invalid samples, and retain 858 valid data; Variable filtering: Retain 8 key features such as "age", "monthly income", "purchase frequency", "appearance design preference", and "brand attention".

Cluster parameter optimization: Determine the optimal number of clusters k by using the elbow rule (SSE descent inflection point) and maximizing the contour coefficient. Set the weight coefficients $\alpha = 0.7$ (numerical variable) and $\beta = 0.3$ (categorical variable) to balance the contributions of the two types of data.

Model training and validation: The improved K-prototypes algorithm was iterated 50 times until convergence, with a contour coefficient of 0.62 and a Calinski Harabasz index of 285.7, which is superior to traditional K-means (0.51, 198.3).

The significance of inter cluster differences was verified through ANOVA and chi square test ($p < 0.01$).

Table 3. Characteristics of Cluster Centers

variable	Young middle-income high-frequency buyers	Middle aged, high-income, low-frequency buyers	Middle aged and high-income high-frequency buyers
age	1.476 (Youthfulness)	0.131 (middle-aged)	-2.134 (older)
monthly income	-1.319 (middle-income)	1.588 (high-income)	1.099 (high income)
purchase frequency	0.614 (high frequency)	-1.217 (low frequency)	0.324 (medium high frequency)
Cultural connotation	0.000 (low attention)	1.000 (high attention)	0.000 (low attention)
Cultural cognition	1.000 (high cognition)	0.000 (low cognition)	1.000 (high cognition)

3.3. Structural Equation Modeling (SEM)

(1) Theoretical basis and research hypotheses: Based on consumer behavior theory and cultural psychology theory, nine hypotheses are proposed, including the significant positive impact of cultural cognition on emotional identification (H1) and the significant positive impact of cultural cognition on product integration (H2). (2) Model setting: divided into two parts: measurement model and structural model. The measurement model sets observation variables

to measure variables such as cultural cognition, emotional identification, and product integration; The structural model describes the relationship between latent variables, such as the influence path of cultural cognition on emotional identification, product integration, etc.(3) Model estimation method: Maximum likelihood estimation (MLE) is used, and professional statistical software is used to iteratively find the optimal parameter estimation values, obtaining non standardized regression coefficients, standard errors, z-values (CR values), p-values, and standardized regression coefficients.(4) Model fitting indicators: Evaluate the degree of model fitting through various indicators such as chi square test, chi square degree of freedom ratio, goodness of fit index (GFI), and root mean square approximation error (RMSEA).

4. Results

4.1. Binary Logit Regression Model Results

(1) Variable significance and direction of influence: Health products, medicinal herbs, and specialty foods have a significant positive impact on purchase intention ($p < 0.05$), while cultural and creative accessories and specialty drinks have no significant impact on purchase intention ($p > 0.05$). (2) Overall performance of the model: Likelihood ratio test shows that the model is significantly better than the intercept only model; Limited explanatory power (McFadden $R^2 = 0.024$, Cox&Snell $R^2 = 0.031$); The overall prediction accuracy is 65.98%, with an accuracy of 92.93% when the true value is 0 and an error rate of 82.37% when the true value is 1; The collinearity diagnosis shows no multicollinearity issues.

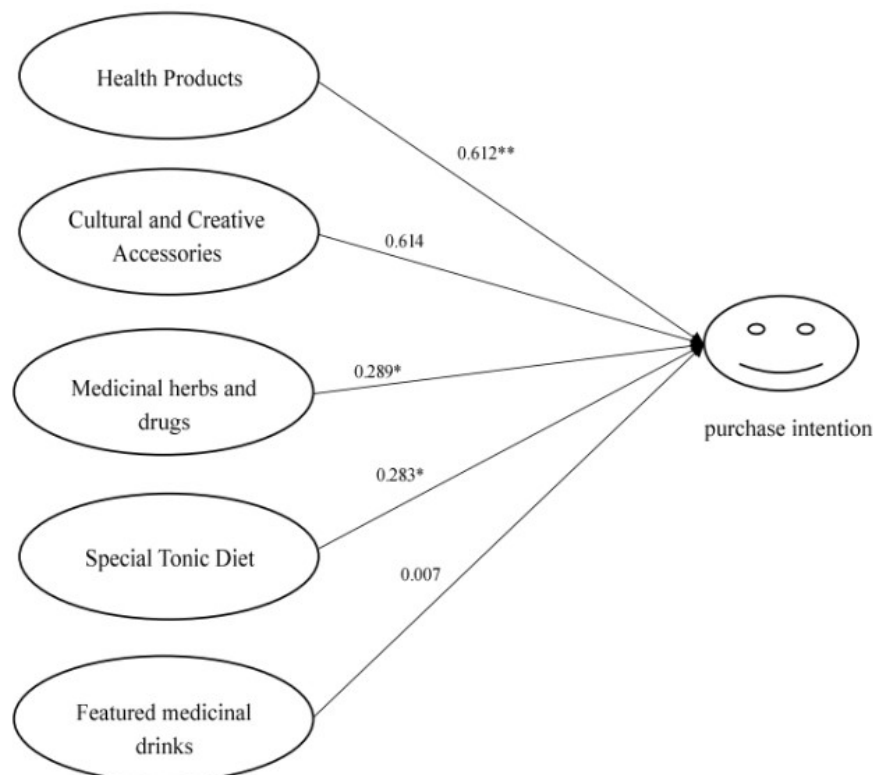


Figure 1. Purchase intention of various types of cultural and characteristic products of diverse traditional Chinese medicine in Xinjiang

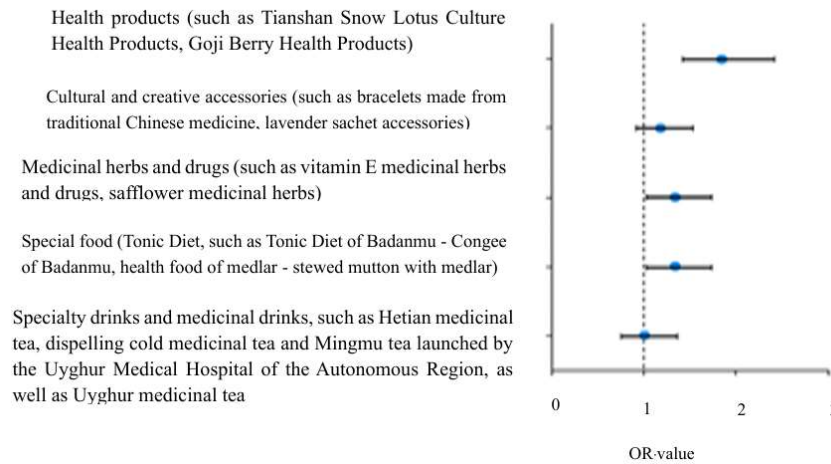


Figure 2. Model Results Chart

4.2. Improved K-prototypes Clustering Algorithm Results

(1) Basic information of clustering categories: Consumers are divided into three categories: young middle-income high-frequency buyers, middle-aged high-income low-frequency buyers, and middle-aged high-income high-frequency buyers, with frequencies of 314, 270, and 286, accounting for 36.09%, 31.03%, and 32.87%, respectively. (2) Cluster center characteristics: Different clusters have differences in age, monthly income, purchase frequency, cultural connotation attention, cultural cognition, and other aspects. For example, young and middle-income high-frequency buyers with younger ages and lower attention to cultural connotations; Middle aged, high-income, low-frequency buyers have high income and high attention to cultural connotations. (3) Model training and validation: The improved K-prototypes algorithm was iterated 50 times until convergence, with a contour coefficient of 0.62 and a Calinski Harabasz index of 285.7, which is superior to traditional K-means (0.51, 198.3). The significant differences between clusters were verified through ANOVA and chi square test ($p < 0.01$).

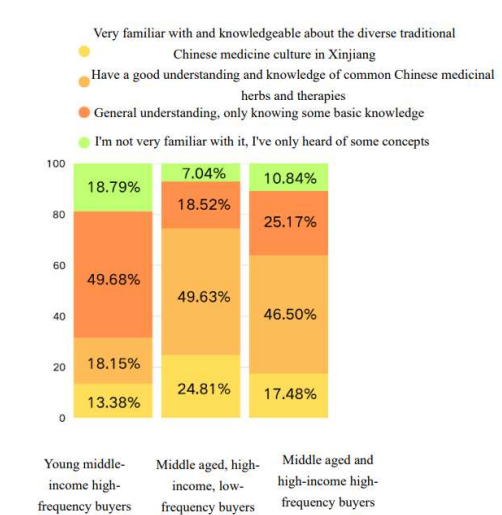


Figure 3. Analysis of the Cognitive Situation of Diversified Traditional Chinese Medicine Culture in Xinjiang Based on Consumption Characteristics

4.3. Structural Equation Modeling Results

(1) Model regression coefficient: Cultural cognition has a significant positive impact on multiple variables such as emotional identification and product integration; Emotional identification and product integration have a significant positive impact on purchase intention; Purchase

intention has a significant positive impact on purchase frequency.(2) Model fitting indicators: Some indicators show that the overall fitting effect of the model is good, such as GFI (0.992), CFI (0.986), etc., which are greater than 0.9. However, the chi square degree of freedom ratio (χ^2/df) is 4.323, slightly higher than the recommended value of 3. Some indicators, such as PGFI (0.236), do not meet the ideal standard.(3) Model fit R^2 : Some variables such as emotional identification and purchase intention have relatively high R^2 values, indicating that they can be well explained by the independent variables of the model; Q24 tourism experience and Q9 cultural complex have low R^2 values and limited explanatory power.(4) Covariance and Residual Estimation Values: Covariance displays the correlation between variables; The residual term standard estimation coefficients are mostly small, but some variables such as emotional identification residuals are relatively large.

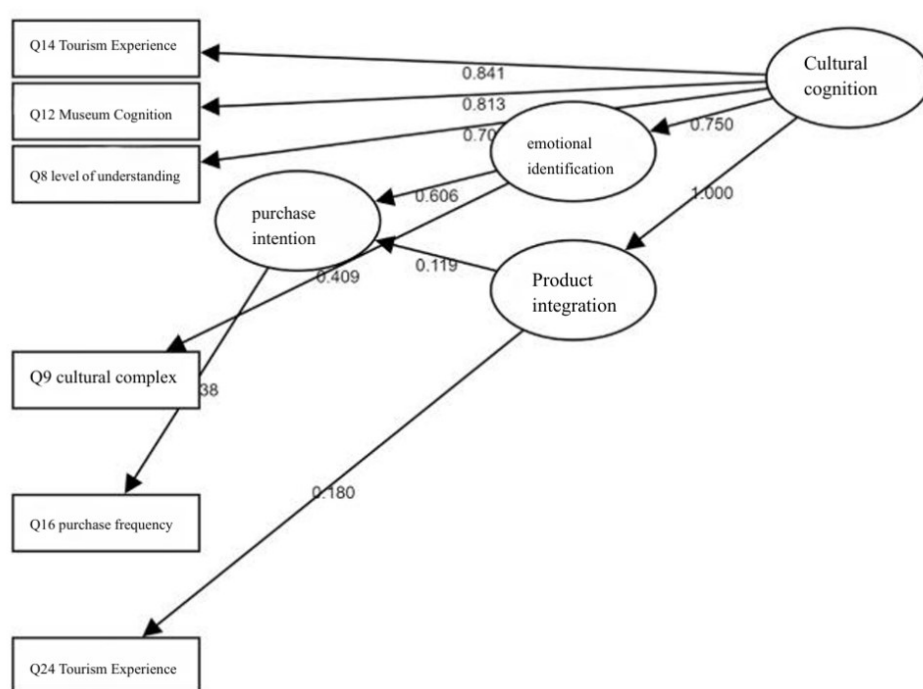


Figure 4. Multivariate Relationship Analysis Based on Structural Equation Modeling

(1) Conclusion of the binary Logit regression model: Health products, medicinal herbs, and specialty foods are the core driving factors that affect consumers' purchase intention, while cultural and creative accessories and specialty drinks have no significant impact on purchase intention. The explanatory power of the model is low, and the error rate in predicting true purchase intention (1) is high. It is necessary to optimize variable selection or introduce interaction terms. At the same time, a high sample loss rate may introduce selection bias.

(2) Conclusion of the improved K-prototypes clustering algorithm: Through this algorithm, fine segmentation of consumer groups has been achieved, and different groups have different consumption characteristics. Young people pay attention to product appearance design, while middle-aged high-income low-frequency buyers are sensitive to cultural connotations and brand awareness. Middle aged high-income high-frequency buyers have higher requirements for price rationality and cultural value.

(3) Conclusion of structural equation modeling: Cultural cognition has a significant positive impact on multiple variables such as emotional identification and product integration; Emotional identification and product integration have a significant positive impact on purchase intention, with emotional identification playing a more prominent role; Purchase intention has a significant positive impact on purchase frequency. The overall fitting effect of the model is

good, but there is room for improvement in some indicators, and the explanatory power of some variables needs to be strengthened.

(4) Comprehensive conclusion: The development of traditional Chinese medicine culture in Xinjiang has achieved certain results, such as improved public awareness and breakthroughs in international expansion. However, there are also problems such as policy implementation being disconnected from demand, supply-demand mismatch, and cultural dissemination gaps. Enterprises should focus on strengthening consumer cultural awareness and optimizing product design to incorporate cultural elements; The government needs to improve precise publicity mechanisms, strengthen government enterprise cooperation, and assist in market expansion to promote the sustainable development of Xinjiang's traditional Chinese medicine cultural industry.

5. Conclusion

Significant achievements have been made in the development of traditional Chinese medicine culture in Xinjiang. In terms of policy implementation, public awareness has been improved, breakthroughs have been made in international expansion, and the effect of cultural integration is good; In terms of market supply, product innovation is accelerating and potential demand is being released. However, there are also structural contradictions, such as a disconnect between policy implementation and demand, a mismatch between supply and demand with technological barriers, and insufficient cultural dissemination and ethnic integration; The mechanism of cultural enrichment in Xinjiang is not closed loop, cultural identity is decoupled from consumer behavior, and international communication is trapped in a "regionalization trap". We should implement targeted publicity, strengthen government enterprise cooperation, and enhance public policy awareness. Innovative drug appearance design, strengthening doctors' understanding of traditional Chinese medicine, and launching cultural integration products. Support international certification for enterprises, adopt cultural differentiation strategies, and build an integrated system to promote the internationalization of medicinal materials. Build a new media matrix, integrate ethnic medicine, and create an immersive experience platform. Optimize market supply, match the needs of young people, enhance product diversity and purchase conversion rates.

References

- [1] Li Candong, Ji Lijin, Lu Yuhui, etc On the Logical Starting Point of Traditional Chinese Medicine Health Cognition Theory [J]. Chinese Journal of Traditional Chinese Medicine, 2011, 26 (01).
- [2] Qian Z, Wang G Y, Henning M, Chen Y. Understanding health literacy from a traditional Chinese medicine perspective[J] . Journal of Integrative Medicine, 2023, 21(3).
- [3] Zhang Yong, Zhang Yanxia Strategic Exploration of Traditional Chinese Medicine's Participation in the Cultural Enrichment Project Based on SWOT Analysis [J]. Xinjiang Traditional Chinese Medicine, 2023, 41 (02).
- [4] Zhao Panzhen, Shi Long The Challenges and Strategies of Translating and Spreading Ethnic Minority Medical Classics from the Perspective of the "Healthy Silk Road" [J]. Traditional Chinese Medicine Culture, 2019, 14 (04).
- [5] Li Zhongkai, Li Jinye Exploration of the Path for Xinjiang to Build a Medical Service Center in the Core Area of the Silk Road Economic Belt [J]. Finance and Economics, 2019, (16).
- [6] Yu Jinrong, Tu Mengjun, Aliyeguli Abulimiti A Study on the Relationship between Cultural Cognition, Differential Patterns, and Brand Attitudes: Mediated by the Consensus of Social Mainstream Goodness [J]. Business Economics Research, 2022, (19).
- [7] Zhang Xiaoqin, Liao Zhaoyu Research on the Development Strategy of Uyghur Medical Culture Tourism Industry [J]. Science and Technology Entrepreneurship Monthly, 2016, 29 (22).

- [8] Hailiang A Study on the Influence of Motivation and Perceived Value of Inner Mongolia Mongolian Medicine Culture Tourism on Tourism Behavior Intention [D]. Xinjiang University, 2020.
- [9] Wang Liqiong, Su Yalatu Opportunities and challenges brought by new media to the dissemination of Mongolian medicine culture in the context of the "the Belt and Road" [J]. Chinese Journal of Ethnic Medicine, 2018, 24 (10).
- [10] Gong X L, Gao W. Research on performance evaluation system of Mongolian medicine culture tourism construction project [J]. Science & Technology for Development, 2020, 16(1).