EXPLORING FOOD CULTURE AND REGIONAL CHARACTERISTICS IN ANCIENT LITERATURE USING TEXT CLUSTERING MODELS

Xu Zhang*

- Jiangsu College of Tourism, Yangzhou, Jiangsu, 225000, China
- Education Management, Krirk University, Bangkok, 10220, Thailand
- xuzhang202388@163.com

Reception: 15 February 2024 | **Acceptance**: 10 April 2024 | **Publication**: 23 May 2024

Suggested citation:

Zhang, X. (2024). **Exploring Food Culture and Regional Characteristics in Ancient Literature Using Text Clustering Models**. *3C Empresa. Investigación y pensamiento crítico, 13(1)*, 214-230. https://doi.org/10.17993/3cemp.2024.130153.214-230

ABSTRACT

In order to explore the connection between food culture and regional characteristics in ancient literature, a text clustering model is used to bring together literary works with similar food culture descriptions. First, selection criteria such as era, region, and author's social background are set. Then, the works are iteratively assigned to the closest cluster centers by the K-Means algorithm, and these cluster centers are continuously updated to find the best clustering results. In the text preprocessing stage, keywords related to food culture were extracted from each work to form a basic feature set. Finally, the K-Means algorithm is used to identify food culture themes with different regional characteristics. The entropy values of the text clustering model are 92.85 and 72.6, which reveal the common dietary elements of the literary works in each cluster and reflect the close relationship between regional characteristics and dietary habits.

KEYWORDS

Text clustering; K-Means algorithm; cluster center; text preprocessing; food culture

https://doi.org/10.17993/3cemp.2024.130153.214-230

INDEX

https://doi.org/10.17993/3cemp.2024.130153.214-230

1. INTRODUCTION

The traditional Chinese culture is so broad and profound that the traditional food culture is the one with the most enduring vitality, deepest influence and most conscious dissemination [1]. In the long period of feudal empire, no matter how the kingdoms and mountains changed, no matter how the dynasties changed, it must be the food culture that has always been firmly passed on to the world [2]. Food culture not only reflects the geographical location, living habits, social customs and national character of a region or tribe, but also reveals the deeper historical and cultural qualities such as religious culture, international exchanges, and the rise and fall of national power [3]. Taking food culture as a breakthrough to study the historical change process of a tribe, nation and country is undoubtedly a more documentary, persuasive and creative research angle [4]. However, in addition to the intergenerational transmission by word of mouth, there is no better way of transmission of food culture, as a slang culture which is parallel to the elegant art, there are not many records about food culture in the education system or in the proper historical documents, and the reason why the Chinese food culture has been able to flourish for thousands of years is mainly realized by the folk education of the word of mouth [5-6].

Adams, S. A explores the ways in which ancient authors used mnemonic language to express apparent inter- and intra-textual allusions, using specific language and expressions to quote, echo, or reinterpret allusions in other texts [7]. Peterson, M. R et al. examined the characterizations discussed by ancient literary theorists in their works, comparing these characterizations with the differences and similarities between the main structural relationships identified by David R. Bauer, and Robert A. Traynor to identify the similarities and differences between the major structural relationships [8]. Baumard, N et al. employed a combination of qualitative and quantitative methods to build a database of ancient literary fiction, using probabilistic generative models to reconstruct the potential evolution of love and to assess the respective roles of cultural diffusion and economic development on the growth of love [9]. Kellner, A explored the possible dialog between ancient Greek and Mesopotamian chronicles, proposing the idea that fragmentary preservation of Greek texts and cuneiform tablets may have influenced the development of the Greek script, and that reviewing textual evidence from ancient Greek and Mesopotamian chronicles can lead to a better understanding of the similarities and differences, and for this purpose the history of Greek and Akkadian temples can be used as a test case [10]. Chiu, B et al. propose a graph-based representation to reconstruct two sets of features and use GAE to aggregate this paper. , by clustering the learned representation and using vector dimensions to classify the document. The results show that the features learned by this method such as word frequency inverse document frequency and average embedding outperform other existing features in terms of document clustering performance [11]. Yu, P et al. proposed a model based on potential space energy with good interpretability in text modeling and attempted to solve some of the problems of data space EBMs. By introducing a novel relationship between diffusion models and potential space EBMs, and a regularization method based on geometric

https://doi.org/10.17993/3cemp.2024.130153.214-230

clustering, this model performs well in interpretable text modeling [12]. Ghosal, A et al. briefly describe the types of clustering methods available, and then survey the areas in which clustering analysis has been effectively applied in pattern recognition and knowledge discovery [13]. Su, H et al. introduce a new method called INSTRUCTOR to compute the text embedding of a given task instruction. Unlike previous methods, INSTRUCTOR is a single embedder that generates text embeddings applicable to different downstream tasks and domains without additional training [14].

Since existing studies have not targeted the connection between diet and geography, this paper uses text clustering models, especially the K-Means clustering algorithm. First, the selection criteria of literary works are established to ensure the representativeness and diversity of the samples, including the era span, author background, and genre categories. Then, after selecting the works, keywords and expressions related to food culture and regional characteristics are extracted from the text, such as food names, cooking styles, and food sources. The basic process of clustering starts with randomly selecting K centroids, classifying the text features according to their similarity to these centroids, and optimizing the centroid positions in each round of iteration until the optimum is reached or the stopping condition is satisfied. Finally, the clustering results reveal the geographical distribution patterns of food elements in ancient literature, and observe the differences and connections of food culture among different regions.

2. CRITERIA AND SOURCES FOR THE SELECTION OF LITERARY WORKS

It is crucial to select appropriate literary works that not only need to cover a wide range of different historical periods and geographic regions, but also be able to reflect the diversity and richness of food culture [15]. The criteria for the selection of literary works and the sources of the works are as follows:

1. The selected literary works should cover different historical periods in ancient China, such as the pre-Qin, Han, Tang, Song, Ming, and Qing dynasties, in order to analyze the changes in food culture in different eras.

2. Considering the vast area of China, different regions have their own unique food cultures, so it is necessary to select literary works that reflect different regional characteristics.

3. The diversity of genres includes works in different genres such as poetry, prose, novels, opera, etc., in order to comprehensively capture the many manifestations of food culture.

4. Preference will be given to literary works that involve more descriptions of food and drink, especially those that depict food, eating habits, and eating scenes in detail.

https://doi.org/10.17993/3cemp.2024.130153.214-230

5. Select classic works that have an important position and wide influence in the history of Chinese literature.

6. Many ancient literary works have been organized and published, and these publications are one of the main sources for acquiring texts. For example, the Chinese Philosophical Books Electronic Program and the Ancient Books Library of the National Library of China, these databases provide a large number of digitized ancient literary texts. Referring to scholars' research in the field of ancient literature and food culture, the frequently cited literary works in them are selected.

A key step in the text clustering process is to calculate the similarity between texts [16]. The commonly used method is cosine similarity, which is calculated as follows:

$$(A, B) = \frac{A \cdot B}{\| A \| B \|} \quad (1)$$

where A, B is the two text vectors, A, B is the dot product of the vectors, and $\| A \|$ and $\| B \|$ are the modes of the vectors.

3. TEXT CLUSTERING MODEL

3.1. K-MEANS CLUSTERING ALGORITHM

3.1.1. FUNDAMENTALS

k-means is an iterative relocation method that minimizes the sum of squared errors as an objective function, and each iteration consists of two steps [17]. In textual data, each literary work can be considered as a data point, and the words and expressions about food and drink contained therein constitute a multidimensional feature space. Using the K-means algorithm, the works can be assigned to different clusters based on their descriptions of dietary culture, each representing a specific geo-cultural feature. Given the center of the cluster, each data point is assigned to the cluster where the closest clustering center is located according to the nearest neighbor principle. The clustering centers are adjusted so that the data points in the cluster where they are located have the smallest SSE to the cluster center, i.e., so that the similarity between the cluster center and the other data points in the cluster is maximized.

$$sse = \sum_{i=1}^{k} \sum_{x \in c_i} \mathrm{dist}\left(x, o_i\right)^2 \quad (2)$$

Where k is the number of clusters, o_i is the clustering center of the i rd cluster c_i, and $\mathrm{dist}\left(x, o_i\right)$ refers to the dissimilarity between data points o and o_i. Different

https://doi.org/10.17993/3cemp.2024.130153.214-230

dissimilarity calculations often lead to different clustering results, and the Euclidean distance metric is usually used.

3.1.2. BASIC CLUSTERING PROCESS

In order to reduce the effect of different magnitudes on clustering, the data needs to be standardized, using a method of departure normalization so that the data falls on the interval [0,1]. The deviation normalization is as follows:

$$x_p^* = \frac{x_p - \min_p}{\max_p - \min_p} \tag{3}$$

Where $\max_p$ and $\min_p$ are the maximum and minimum values on the p attribute respectively. The basic process is shown in Fig. 1, using the definition to normalize the data in the dataset, the number of iterations $t = 0$, calculate the mutual distance between the data points in the dataset, if $t = k$ then the algorithm terminates. Otherwise redefine, if $\| M \| = 0$, it means that there is not yet a data point in the set of clustering centers, take each data point in the dataset as a clustering center, calculate its corresponding sse, select the data point corresponding to the smallest sse as the first initial clustering center Z_1 and add it to the set M. Record the distance from data point x_j to Z_1. md_j^1, otherwise, according to the definition, select the data point that can minimize the sse as the clustering center Z_t for the tth iteration and added to the set M. Update the shortest distance md_j^1 from all data points x_j in the dataset to the set M. The number of iterations t is incremented by 1 and go to $t = k$.

https://doi.org/10.17993/3cemp.2024.130153.214-230

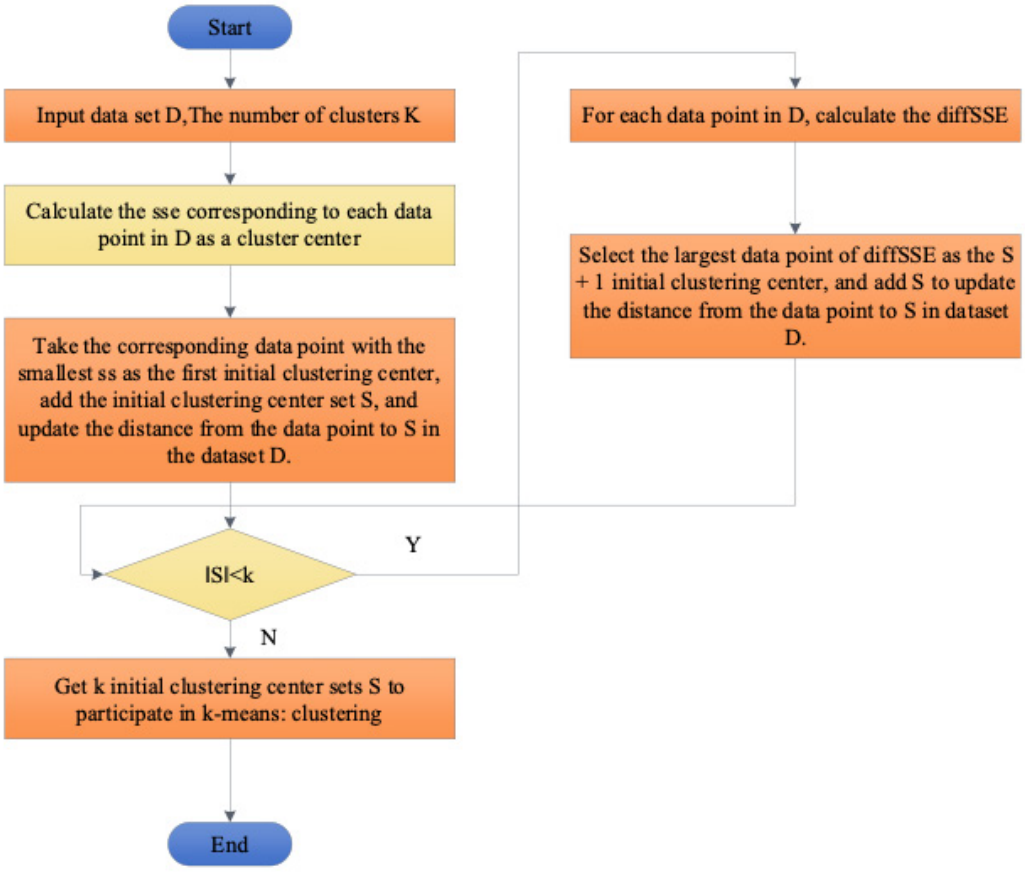

Figure 1. Basic clustering process

4. EXTRACTION OF CHARACTERISTICS OF FOOD CULTURE AND REGIONAL FEATURES

4.1. TEXT FEATURE EXTRACTION

Feature selection has a basic assumption that each feature is independent of each other, but this assumption does not hold true for text data, because the phenomenon of proxemics and polysemy, which is common in text data, makes the text of each feature between the existence of extremely complex semantic links [18-19]. The problem of near-synonyms and polysemous words cannot be solved by feature selection and can only rely on feature extraction.

4.1.1. DOCUMENT FREQUENCY

The document frequency of a lexeme is the number of documents in which the lexeme occurs in the training corpus.4 DF is used as the feature extraction, and it is basically assumed that the lexeme with DF value below a certain threshold is a low-frequency lexeme, which does not contain, or contains less information about the category [20]. Removing such words from the original feature space not only reduces

https://doi.org/10.17993/3cemp.2024.130153.214-230

the dimensionality of the feature space, but also has the potential to improve the classification accuracy.

Document frequency is the simplest feature extraction technique and can be easily used for large-scale corpus statistics due to its linear computational complexity relative to the size of the training corpus. However, in information extraction studies it is commonly recognized that words with low DF values have more information relative to words with high DF values and should not be removed completely.

4.1.2. CHI STATISTICS

The CHI statistic measures the degree of correlation between lexical entry t and document category C and assumes that the x-distribution with first order degrees of freedom is met between t and C. The higher the value of the X2 statistic of a lexical item for a category, the higher the correlation with that category and the more category information it carries. Let N denote the total number of documents in the training corpus, C is a particular category, t denotes a particular lexical item, A denotes the frequency of documents belonging to category C and containing t, B denotes the frequency of documents not belonging to category C but containing t, C denotes the frequency of documents belonging to category C but not containing t, and D is the frequency of documents belonging to neither C nor t. The CHI value of t for C is calculated by the following formula:

$$\chi^2(t,c) = \frac{N \times (AD - CB)^2}{(A+C)(B+D)(A+B) + (C+D)} \tag{4}$$

For multi-category problems, the CHI value of t for each category is calculated separately. The CHI value of lexeme 1 for the whole corpus is then calculated using the following formula and tested separately:

$$\chi^2_{\max}(t) = \max_{t=1}^{m} \chi^2\left(t, c_i\right) \tag{5}$$

where m is the number of categories. The lexical entries below a specific threshold are removed from the original feature space, and the lexical entries above that threshold are retained as the features of the document representation [21]. By this method, the most relevant words and phrases related to food culture and regional features in ancient literature can be identified. Removing words and phrases below a specific threshold from the original feature space while retaining words and phrases above that threshold as features for document representation optimizes the performance of the clustering model and ensures that the model focuses on the most informative features, thus revealing more accurately the deep connection between food culture and regional features in literature.

https://doi.org/10.17993/3cemp.2024.130153.214-230

4.2. MUTUAL INFORMATION

Mutual Information MI is widely used in statistical language modeling. If we use A to denote the frequency of documents containing lexeme 1 and belonging to category C , B is the frequency of documents containing but not belonging to C , C denotes the frequency of documents belonging to C but not containing t, and N denotes the total number of documents in the corpus, the mutual information of t and C . The mutual information of 5 and 6 can be calculated by the following formula:

$$MI_{\max}(t) = \max_{i=1}^{m} I(t, c_i) \tag{6}$$

Where m is the number of categories. The words below a specific threshold are removed from the original feature space, the dimension of the feature space is reduced, and the words above the threshold are retained.

In conclusion, although CHI and MI perform well in the English text classification problem, their performance is far less than that of DF. After careful analysis, it is found that the reasons for this difference come from the fact that the feature extraction methods using category information rely on low-frequency words and the fact that Chinese has a higher dimension of the feature space compared with English.

4.3. REPEATED STRING FEATURE EXTRACTION

Most clustering algorithms regard a document as a collection of words only, completely ignoring the order and co-occurrence relationship between words, which may provide important information for document clustering. Therefore, we extract the key repetitive strings from the whole document collection as text features. In order to quickly extract most of the repetitive strings, the introduction of maximized repetitive strings is necessary and sufficient. It can capture all the meaningful repetitive structures of the strings in a very concise way, and also avoids generating a lot of unnecessary output. Non-maximized repeated strings do not need to be reported, because the text must be contained in some maximized repeated strings. However, not all maximized repetitive strings are useful, many of them are only part of phrases, which are semantically incomplete and meaningless, so further filtering is needed to filter out the meaningful and interesting parts of them.

The algorithm first scans each document in the corpus and removes deactivated words and non-word symbols such as numeric punctuation. A document is treated as a string, and all documents are concatenated into a pseudo-document. Each word is converted to a 2-byte integer so that each English word or Chinese character can be treated as a unit. At the same time, each subscript in the record string corresponds to the document number to which the character belongs, and the documents are separated by a specific boundary symbol, which does not appear in any of the original documents. Obviously, across the document boundary of the substring, is meaningless, we limit the algorithm to find the duplicate string in a document. More

https://doi.org/10.17993/3cemp.2024.130153.214-230

strictly, since sentence boundaries often imply a change of topic, repeated strings can also be limited to a sentence. This also reduces the cost of the duplicate string discovery algorithm. The output of the preprocessing is a string containing all the documents in the corpus and the corresponding document number records.

Through all substrings, which are clustered into a relatively small number of classes, the statistical information of all substrings can be obtained by calculating the frequency of at least 2N-1 substrings. The data structure used is the array of suffixes and the corresponding array of maximal common prefixes created from the input text strings. This part of the algorithm outputs all the maximized repeated strings and their frequency statistics. A complete repeated string should be both left-maximized and right-maximized. After obtaining the complete repeated strings and their frequencies in document T, it is easy to calculate the stability and independence of each repeated string. The quality of each clustering algorithm is significantly improved when using repetitive strings as features.

4.4. EXTRACTION PROCESS

In exploring the food culture and regional characteristics in ancient literature, the extraction process using the text clustering model is shown in Figure 2. Firstly, ancient literary texts need to be collected, including various literary works, historical records and cultural documents. Then, these texts are preprocessed, including text cleaning, word splitting, removal of deactivated words, and other operations to prepare the data. Next, the features of the texts are extracted, which can be achieved through TF-IDF, Word Embeddings, or topic modeling. Subsequently, an appropriate text clustering model, such as K-Means or hierarchical clustering, is selected to cluster the text features. Once the model training is completed, the text content in each clustered cluster will be analyzed to identify the food culture and geographical features in it. Finally, the results are presented through visualization tools with in-depth analysis and conclusions [22]. This process can help to gain a deeper understanding of cultural characteristics and regional differences in ancient literature.

https://doi.org/10.17993/3cemp.2024.130153.214-230

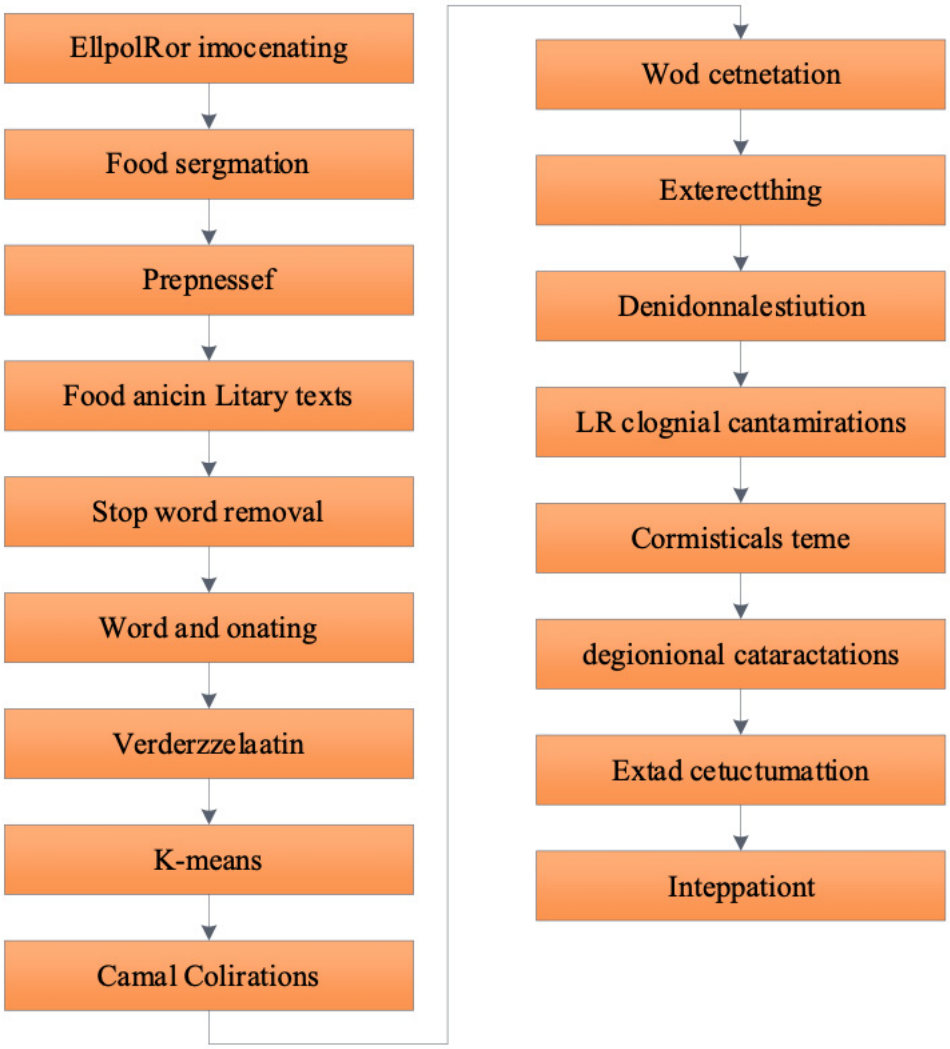

Figure 2. Feature extraction process

5. THE CONNECTION BETWEEN FOOD CULTURE AND REGIONAL CHARACTERISTICS IN ANCIENT LITERATURE

As an important cultural element, food culture is not only a subject matter in ancient literature, but also a carrier that conveys the cultural characteristics of a specific region through literary works.

First of all, food culture in ancient literature reflects the resources and climatic conditions of the region. The ingredients and cooking styles of different regions are often influenced by the local natural environment, which is vividly expressed in literary works. For example, literature located in seaside regions may emphasize the importance of seafood and seafood products, while mountainous regions may highlight mountain delicacies and wild foods. This way of reflecting regional characteristics not only enriches the depiction of literary works, but also enables readers to better understand the geographical distribution of the society at that time. Secondly, the relationship between food culture and regional characteristics is also

https://doi.org/10.17993/3cemp.2024.130153.214-230

reflected in the cultural exchange and integration in literary works. Ancient literature often describes cultural exchange and borrowing between different regions, of which food culture is an important aspect. For example, the trade on the Silk Road made the food culture of the East and the West influence each other, which was often reflected in literary works. The exchange and fusion of regional characteristics make the literary works more diversified and also convey the importance of cultural exchange. Finally, the relationship between food culture and regional characteristics in literature also reflects social and cultural changes. With the evolution of history, the food culture of different regions has changed, which can be traced back to the depiction of different historical periods in literary works. Through literature, one can understand the evolution of food supply, eating habits and cooking skills in ancient societies, thus gaining a deeper understanding of the historical process and cultural development.

To sum up, the relationship between food culture and regional characteristics in ancient literature is multi-layered and multi-dimensional, intertwined, and together they constitute a rich and colorful literary work.

6. ANALYSIS OF CLUSTERING RESULTS

6.1. THEME IDENTIFICATION

A series of ancient literary works were collected, which cover the cultural backgrounds of different regions and periods. These textual data were entered into a textual clustering model to group them into clusters with similar themes. A suitable number of clusters was chosen in order to categorize food culture and regional characteristics into two main themes. The clustering results of different texts were analyzed and calculated by the model to determine the themes of each cluster. The clustering results of the first-level and second-level texts are shown in Table 1, in which the frequency and relevance of keywords and the semantic information of the text content were paid attention to ensure the accuracy and consistency of the theme identification.

Table 1. Results of first - and second-level text clustering

Class variety	Number of samples included	Conceptual structure of text	Theme
C1	54	Ingredients + dishes + techniques	Menu
C2	53	Dishes + Restaurants + Ingredients	Food guide
C3	50	Celebrity + Dish + ingredients	Cultural story
C4	28	Ingredients + a few other concepts	Ingredients introduction
C11	26	Other cuisines + various ingredients	Recipes Other than Sichuan Cuisine Recipes of Sichuan cuisine
C12	25	Sichuan cuisine + various ingredients	Recommend the famous dishes around the food guide

https://doi.org/10.17993/3cemp.2024.130153.214-230

C21	24	Cuisine + various ingredients	Recommend the famous restaurants around the food guide
C22	28	Restaurants + various cuisines	Recipes Other than Sichuan Cuisine Recipes of Sichuan cuisine
C31	30	Celebrities + various cuisines	Cultural stories related to food in different places
C32	23	Non-vegetable ingredients	Introduction to non-vegetable ingredients
C41	24	Vegetable based ingredients	Vegetable ingredients introduction

The comparison of entropy value of the first layer clustering results is shown in Table 2, in the text clustering model, the entropy value is 92.85, indicating that the clustering effect of this model is relatively good. In the feature word clustering, the entropy value is 221.25, which is relatively high, indicating that the clustering effect of this model is poor. The text clustering model proposed in this paper is more effective, similar to the original knowledge base grouping, and can effectively improve the clustering accuracy of text.

Table 2. Clustering results of the first layer

Evaluation index	Entropy value	The number of samples contained in a class cluster			
		C1	C2	C3	C4
Evaluation index	92.85	54	53	50	28
Feature word clustering	221.25	125	88	6	4

The results of the second layer clustering are shown in Table 3, in the text clustering model, the evaluation index is 72.6, which is lower than the 144.8 of the feature word clustering, the evaluation index of the text clustering model is lower, indicating that its clustering effect is relatively good, and it is more suitable to be used in the text clustering task. The higher evaluation index of feature word clustering indicates that its clustering effect is relatively poor and may be less suitable for text clustering tasks. Therefore, the text clustering model may be more suitable for clustering tasks in ancient literature.

Table 3. Results of second-layer clustering

Evaluation index	Entropy value	The number of samples contained in a class cluster						
Evaluation index	72.6	C11	C12	C21	C22	C31	C32	C41
		26	25	24	28	30	23	24
Feature word clustering	144.8	C1	C2	C3	C4	C5	C6	C7
		100	47	26	15	12	8	2

https://doi.org/10.17993/3cemp.2024.130153.214-230

6.2. MODEL PERFORMANCE ANALYSIS

In this paper, K-mean clustering is chosen as the method of text clustering. The text data of literary works are input into the model and the optimal number of clusters is found by adjusting different numbers of clusters. After performing K-mean clustering, we obtain a set of clusters of literary works. Each cluster contains literary works with similar textual features. The food traditions of different regions are evident in the literary works, and the text clustering model shows a good performance in analyzing these relationships.

The clustering results are shown in Fig. 3, and it can be seen that the points within each cluster are similar to each other, and the eigenvalues vary in the range of 2-4, while there is a certain distance from the points in other clusters, which indicates that the clustering effect is good.

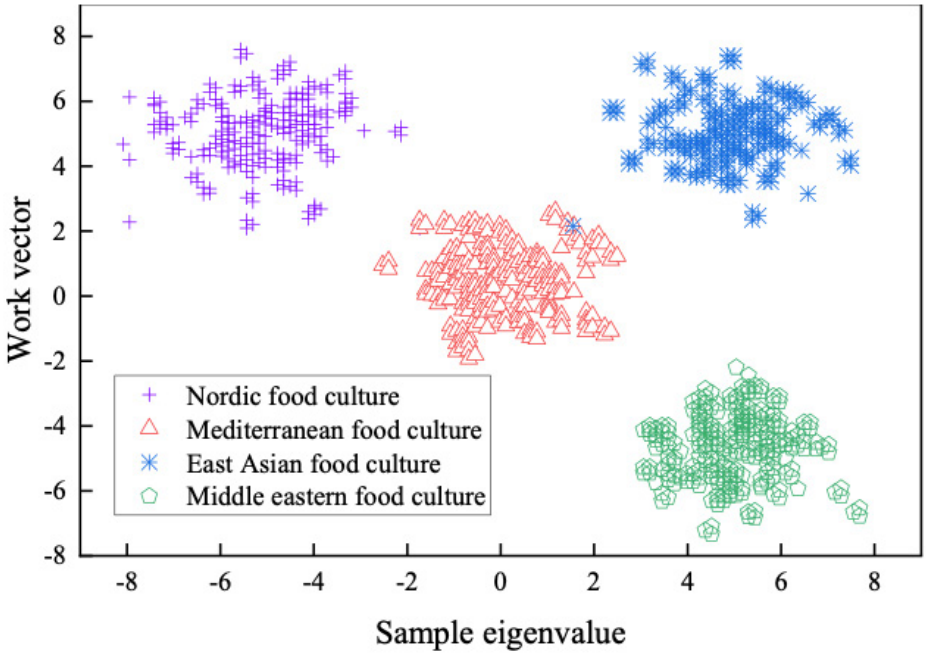

Figure 3. Clustering results

7. DISCUSSION

In future research, the scope of the study can be further expanded to cover more ancient literary works and historical documents in order to gain a more comprehensive understanding of food culture in different regions and periods. Reveal more regional characteristics and cultural differences to provide more in-depth insights for cultural and historical research. Or explore more advanced text clustering models and natural language processing techniques to improve clustering accuracy and efficiency of text feature extraction. The application of methods such as deep learning and neural networks can help to better mine the hidden information in the text for finer analysis and theme identification. Furthermore, the application of the research results to the fields of cultural heritage protection and cultural tourism can promote the inheritance

https://doi.org/10.17993/3cemp.2024.130153.214-230

and promotion of ancient literature, as well as enrich cultural exchange and understanding.

8. CONCLUSION

In this paper, when using the text clustering model to study the connection between food culture and regional characteristics in ancient literature, it is verified that the text clustering model shows high effectiveness. Compared with feature word clustering, the text clustering model showed a significantly lower evaluation index in the second level of clustering, 72.6 vs. 144.8, indicating that it can more accurately identify and group similar food culture descriptions. In addition, the lower entropy value exhibited by the model in the first level of clustering, 92.85 vs. 221.25, confirms its high consistency and accuracy in clustering ancient literature. The visualization of the clustering results further reveals the high degree of intra-cluster similarity and clear inter-cluster boundaries. Overall, the text clustering model is not only suitable for this type of research task, but also provides highly accurate clustering results that strongly support a deeper understanding of the connection between dietary elements and regional characteristics in literary works.

ABOUT THE AUTHOR

Xu Zhang was born in Yangzhou, Jiangsu, P.R. China, in 1983. She received the Master degree from YangZhou University, P.R. China. Now, she works in Jiangsu College of Tourism. She is studying for Ph.D. at krirk university. Her research interests include education management, diet culture, ancient literature.

FUNDING

This work was supported by B/2022/02/94 “Practical Research on Modern Apprenticeship System Promoting the Effect of Moral and Technical Integration of Higher Vocational Students” 2022 Provincial Education Science Planning Project; and 150 ZCZ150 “An Empirical Study on Promoting the Development of Professional Core Literacy through Modern Apprenticeship System: Taking Cooking as an Example” the fifth issue of vocational education and teaching reform in Jiangsu Province.

REFERENCES

(1) Suntikul W, Pratt S, Chong YWJ. Factors that influence Chinese outbound tourists’ intention to consume local food. J China Tourism Res. 2020;16(2):230-47.

(2) Apak ÖC, Gürbüz A. The effect of local food consumption of domestic tourists on sustainable tourism. J Retail Consum Serv. 2023;71:103192.

(3) Savoie-Roskos MR, Hood LB, Hagedorn-Hatfield RL, Landry MJ, Patton-López MM, Richards R, et al. Creating a culture that supports food security and health equity at higher education institutions. Public Health Nutr. 2022;1-7.

https://doi.org/10.17993/3cemp.2024.130153.214-230

(4) Leer J. Designing sustainable food experiences: Rethinking sustainable food tourism. Int J Food Des. 2020;5(1-2):65-82.

(5) Lee HY. Linguistic politeness in the Chinese language and culture. Theory Pract Lang Stud. 2020;10(1):1-9.

(6) Cao C, Zhu C, Meng Q. Chinese international students' coping strategies, social support resources in response to academic stressors: Does heritage culture or host context matter? Curr Psychol. 2021;40:242-52.

(7) Adams SA. Memory as overt allusion trigger in ancient literature. J Study Pseudepigrapha. 2022;32(2):110-26.

(8) Peterson MR, Smith DA. Ancient Literary Criticism and Major Structural Relationships: A Comparative Analysis. J Inductive Biblical Stud. 2020;7(2):4.

(9) Baumard N, Huillery E, Hyafil A, Safra L. The cultural evolution of love in literary history. Nat Hum Behav. 2022;6(4):506-22.

(10) Kellner A. Time Is Running. Ancient Greek Chronography and the Ancient Near East. J Anc Hist. 2021;9(1):19-52.

(11) Chiu B, Sahu SK, Thomas D, Sengupta N, Mahdy M. Autoencoding keyword correlation graph for document clustering. In: Proceedings of the 58th annual meeting of the association for computational linguistics. 2020 Jul;3974-81.

(12) Yu P, Xie S, Ma X, Jia B, Pang B, Gao R, et al. Latent diffusion energy-based model for interpretable text modeling. arXiv preprint arXiv:2206.05895. 2022.

(13) Ghosal A, Nandy A, Das AK, Goswami S, Panday M. A short review on different clustering techniques and their applications. In: Emerging Technology in Modelling and Graphics: Proceedings of IEM Graph 2018. 2020;69-83.

(14) Su H, Shi W, Kasai J, Wang Y, Hu Y, Ostendorf M, et al. One embedder, any task: Instruction-finetuned text embeddings. arXiv preprint arXiv:2212.09741. 2022.

(15) Mattana A. Antiquitas non fingo: Newton, the moderns and the science of ancient history. J Eighteenth-Century Stud. 2020;43(4):447-61.

(16) Wang Y. Similarity detection of English text and teaching evaluation based on improved TCUSS clustering algorithm. J Intell Fuzzy Syst. 2021;40(4):7555-65.

(17) Rashid J, Shah SMA, Irtaza A. An efficient topic modeling approach for text mining and information retrieval through K-means clustering. Mehran Univ Res J Eng Technol. 2020;39(1):213-22.

(18) He Y, Chen C, Zhang J, Liu J, He F, Wang C, et al. Visual semantics allow for textual reasoning better in scene text recognition.

(19) Veyseh APB, Dernoncourt F, Dou D, Nguyen TH. A joint model for definition extraction with syntactic connection and semantic consistency.

(20) Singh R, Singh S. Text similarity measures in news articles by vector space model using NLP. J Inst Eng India Ser B. 2021;102:329-38.

(21) Alpizar D, Adesope OO, Wong RM. A meta-analysis of signaling principle in multimedia learning environments. Educ Technol Res Dev. 2020;68:2095-2119.

(22) Chu CY, Park K, Kremer GE. A global supply chain risk management framework: An application of text-mining to identify region-specific supply chain risks. Adv Eng Inform. 2020;45:101053.

https://doi.org/10.17993/3cemp.2024.130153.214-230