Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.35098

Restaurant Recommendation Algorithm Based on Category Preferences and User Interests

Shuolin Song

School of Management, Shanghai University, Shanghai, China
3441728766@qq.com

Abstract. With the rapid development of the Internet and local life services, catering information has grown explosively, making it difficult for users to quickly find dining venues matching their preferences. To address information overload, homogeneous recommendations, and insufficient semantic understanding, this paper proposes a restaurant recommendation algorithm based on lightweight three-dimensional fine-grained semantic quantification and user interests. First, catering review texts are preprocessed, and a text representation model combining Word2Vec (word embedding) embeddings and TF-IDF (term frequency–inverse document frequency) weighting is constructed. An SVM (support vector machine) classifier is then used to categorize reviews into taste, environment, and service, enabling quantification and structuring of unstructured data. Next, user preference vectors and restaurant feature vectors are built from classification results. A hybrid strategy integrating cosine similarity and category label fitness is designed to generate a Top-5 personalized recommendation list. Experiments on 23,998 real reviews of Nanxin Dessert Shops show the SVM classifier achieves 90.7% overall accuracy, and the average comprehensive matching degree of recommendations is close to 0.9. The proposed method effectively mines users' fine-grained preferences, improves recommendation accuracy and personalization, and provides a feasible solution for catering platform recommendation systems.

Keywords: Restaurant Recommendation, User Preference Mining, Aspect-Based Sentiment Analysis, Word2Vec, SVM

1. Introduction

According to the *55th Statistical Report on Internet Development in China* released by the China Internet Network Information Center (CNNIC), the number of Chinese internet users reached 1.108 billion by December 2024, with an internet penetration rate of 78.6% and 974 million online shopping users. Mobile payment and online consumption have been deeply integrated into people's daily lives. Meanwhile, China's catering industry maintains steady growth. The national catering revenue exceeded 5.8 trillion yuan in 2025, among which the online takeaway market surpassed 1.4 trillion yuan and served nearly 600 million users. With a continuously rising online penetration rate, catering consumption has gradually become digital, personalized and scenario-oriented.

DOI: 10.54254/2755-2721/2026.35098

The in-depth integration of the internet and the catering industry has led to an explosive growth of restaurant information, user reviews and consumption ratings, shifting users from a state of information scarcity to information overload. Faced with massive restaurants, diverse dishes and complex evaluation systems, users struggle to quickly and accurately select dining places that meet their preferences for taste, environment and service. Traditional recommendation methods mostly rely on popularity ranking and simple rating statistics, failing to capture fine-grained user preferences across taste, environment, and service, which leads to homogeneous recommendations and low user satisfaction. Overall, existing catering recommendation algorithms exhibit three major limitations: (1) coarse-grained preference modeling, lacking systematic quantification of taste/environment/service dimensions; (2) shallow semantic understanding, over-reliance on ratings/clicks while ignoring deep semantics in unstructured reviews; (3) high model complexity, lacking lightweight frameworks for real-time catering scenarios. Meanwhile, they also commonly suffer from data sparsity, cold start, and filter-bubble problems. Although recent studies [1-3] have attempted semantic modeling for short texts, they are limited to rough dimension division and insufficient fine-grained preference characterization. Other hybrid recommendation works [4] lack catering-specific semantic modeling, while other studies [5, 6] build basic frameworks with poor semantic utilization and high complexity. Therefore, realizing lightweight three-dimensional fine-grained semantic quantification and hybrid matching remains a critical research gap.

To address these gaps, this paper proposes a restaurant recommendation algorithm based on lightweight three-dimensional fine-grained semantic quantification and user interests. Different from existing methods, this study highlights three core differentiated innovations: (1) Hybrid semantic representation innovation: Unlike prior works relying solely on TF-IDF or static word vectors, this paper integrates Word2Vec contextual semantics with TF-IDF keyword weighting, effectively capturing implicit fine-grained preferences in colloquial catering reviews. (2) Three-dimensional fine-grained preference fusion innovation: Distinct from rough single-dimensional or two-dimensional preference modeling, this work constructs a systematic three-dimensional (taste/environment/service) preference quantification framework via SVM classification, realizing precise mapping from unstructured reviews to structured preference vectors. (3) Lightweight hybrid matching innovation: Differentiated from complex deep learning models or single-similarity matching, this paper designs a lightweight hybrid strategy combining cosine similarity (semantic consistency) and category label fitness (dimension preference matching), which achieves high recommendation accuracy with low computational complexity, adapting to real-time requirements of catering platforms.

Related research on restaurant recommendation based on category preferences has laid a solid foundation for this work [7]. Against this backdrop, it has become an urgent research issue in catering recommendation to accurately mine users' category preferences and potential interests from review texts, and build an efficient, accurate and interpretable personalized restaurant recommendation algorithm. This paper focuses on the restaurant recommendation algorithm based on category preferences and user interests, with two core research directions as follows:

To address the deficiencies in text representation and classification accuracy of catering reviews, a classification model integrating text features and semantic information is constructed to realize accurate classification of three dimensions: taste, environment and service.

To solve the problems of insufficient personalization and low matching degree of recommendations, a combined recommendation strategy integrating dual similarities is designed by combining users' multi-attribute features and user-restaurant rating preferences, so as to improve recommendation accuracy and user experience.

DOI: 10.54254/2755-2721/2026.35098

Based on real-world catering review data, this paper combines text representation, machine learning and personalized recommendation technologies to establish a complete technical pipeline covering text preprocessing, feature representation, dimension classification, preference modeling and personalized recommendation. The research results can provide feasible algorithm references and technical support for online catering platforms, takeaway systems and local life service applications. It also possesses theoretical significance and practical value for improving users' dining efficiency, optimizing merchant operation, and promoting the digital and intelligent transformation of the catering industry.

2. Related knowledge

This chapter introduces core models, algorithms, and key technologies involved in the experiment, including text preprocessing, word embedding representation, feature weighting, text classification, similarity calculation, and personalized recommendation methods. It briefly explains the role and selection reason of each technology, providing theoretical support for subsequent experimental design and algorithm implementation.

2.1. Overview of text classification

The text classification pipeline mainly includes: text preprocessing, text representation, feature selection, and classifier training. In Figure 1, the workflow of text classification is illustrated.

Text classification refers to the technology of automatically assigning texts to predefined categories. It is one of the core tasks in natural language processing. The text classification pipeline mainly includes: text preprocessing, text representation, feature selection, and classifier training.

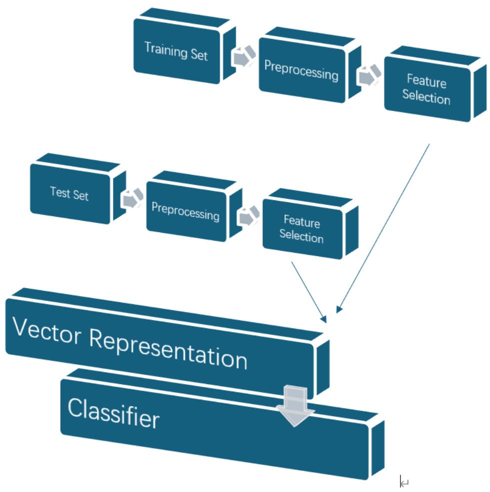

Figure 1. Text classification workflow diagram

2.1.1. Text preprocessing

The purpose of preprocessing is to remove noise, simplify text structure, and improve subsequent model performance. It processes numbers, modal particles, case, punctuation, etc., discards unrecognizable symbols, and converts long sentences into short sentences suitable for computer processing.

The first step is text cleaning: removing symbols, emojis, spaces, and duplicate content.

DOI: 10.54254/2755-2721/2026.35098

Next is Chinese word segmentation: splitting sentences into word sequences according to semantic rules. Accurate segmentation preserves evaluation keywords and avoids semantic loss.

Then stopword filtering: deleting meaningless function words (e.g., "的", "了", "吗", "呢") to reduce data redundancy.

Finally, partofspeech tagging distinguishes nouns, verbs, adjectives, etc.

Preprocessing directly affects word embedding quality and classification accuracy, making it a critical step for catering review processing.

2.1.2. Text representation

Text representation extracts key feature words from raw text and converts them into vector forms, transforming unstructured text into computable structured data. Representative methods include Boolean Model (BM), Probabilistic Model (PM), and Vector Space Model (VSM). This experiment mainly uses VSM.

2.1.2.1. Vector Space Model (VSM)

VSM maps text content to a feature vector space, where each text corresponds to a unique vector. Each dimension represents a feature term, and its value is the term weight. It is a supervised classification model that constructs an optimal hyperplane for highdimensional data classification, with strong stability and generalization in smallsample and highdimensional scenarios. SVM performs well on highdimensional text vectors and meets the needs of catering review classification.

2.1.2.2. TF-IDF

TFIDF measures word importance by combining term frequency and inverse document frequency. It highlights keywords and reduces the influence of common words. It can be combined with Word2Vec to generate discriminative document vectors and improve classification performance.

2.1.2.3. Word2Vec

Word2Vec is an unsupervised model that maps words to lowdimensional dense vectors, preserving semantic relations and solving the word gap and sparsity problems. It effectively captures semantics in colloquial catering reviews and supports text quantification and similarity calculation. Apart from word embedding methods, LDA-based topic modeling is also widely used for catering text analysis [8].

2.2. Overview of recommendation technologies

Recommendation systems analyze user historical behavior, attributes, and item features to deliver potentially interesting content. Three mainstream types are as follows.

2.2.1. Content-based recommendation

Contentbased recommendation matches item features with user preferences and recommends items similar to users' historical favorites. It has high interpretability and intuitive results. It mines user interests from browsing records, attributes, and search logs, computes similarity with candidate

DOI: 10.54254/2755-2721/2026.35098

content, and prioritizes highly similar items. This experiment follows a contentbased paradigm using review features and user dimensional preferences.

2.2.2. Collaborative filtering-based recommendation

Collaborative filtering assumes that users with similar behaviors on most items will behave similarly on other items. It mines user preferences and item attributes from historical interactions. It is divided into userbased and itembased collaborative filtering. Itembased collaborative filtering offers strong personalization but suffers from data sparsity and cold start.

2.2.3. Hybrid recommendation algorithm

Hybrid recommendation combines advantages of multiple methods by fusing multiple features and similarity metrics with reasonable weights. It overcomes the limitation of singlefeature modeling and significantly improves recommendation fitness. This paper adopts a hybrid structure: text features + category preferences + similarity matching.

2.3. Applications of typical recommendation systems

2.3.1. Short video and online audio-visual platforms

Personalized recommendation is a core function of short video and streaming platforms. Platforms build user interest profiles from viewing duration, likes, favorites, forwards, and searches, and push matching content in real time, improving user retention and stickiness. Representative platforms include Douyin, Tencent Video, and iQiyi.

In Figure 2, the recommendation page structure of audio-visual platforms is shown.

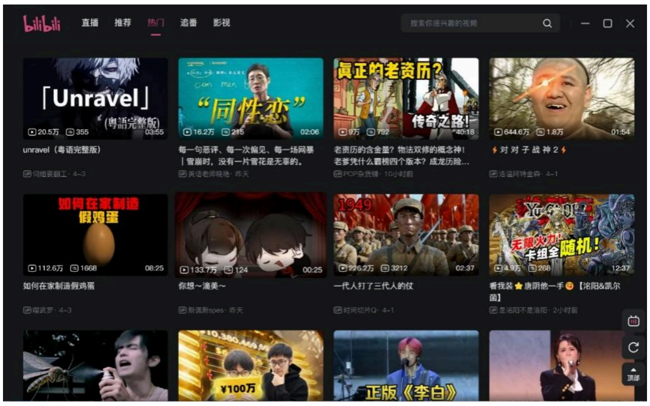

Figure 2. Schematic diagram of the recommendation page on audio-visual platforms

2.3.2. Social and content community platforms

Social and content communities connect users via following, interaction, and sharing. As user scale expands, recommendation systems help users find likeminded people and highquality content, such as "Guess You Like" on Xiaohongshu, "Recommended Follow" on Weibo, and "Friends Are Reading" on WeChat.

DOI: 10.54254/2755-2721/2026.35098

2.3.3. E-commerce and local life platforms

Recommendation systems are widely and maturely applied in ecommerce and local life services, represented by Taobao, JD.com, Pinduoduo, and Meituan. They integrate user behavior and scenario demands to recommend products and services accurately. Local life platforms further combine location, time, and consumption habits to recommend nearby restaurants, hotels, and services, greatly improving decisionmaking efficiency.

2.4. Limitations of contemporary recommendation systems

Although widely deployed, recommendation systems still face critical problems that impair user experience:

2.4.1. Data sparsity

Useritem interaction matrices are usually extremely sparse. Sparse data reduces the reliability of similarity calculation and preference prediction, leading to unstable and biased recommendations.

2.4.2. Cold start

New users lack historical behavior, and new restaurants/items lack interaction data. Traditional algorithms cannot model effectively, harming new user experience and new merchant exposure.

2.4.3. Information cocoons and content homogenization

Overemphasizing existing preferences traps users in information cocoons, reducing diversity and making it hard to discover new styles or highquality niche restaurants.

2.4.4. Insufficient semantic understanding

Traditional algorithms overrely on ratings and clicks and underutilize deep semantic information in reviews. Simple keyword matching fails to capture finegrained opinions, resulting in rough and uninterpretable results.

To address these issues, this paper proposes a restaurant recommendation algorithm based on category preferences and user interests.

2.5. Chapter summary

This chapter introduces the text classification pipeline, text representation methods, mainstream recommendation technologies, and existing limitations. The theories provide a solid foundation for review classification, user preference modeling, and restaurant recommendation algorithms.

DOI: 10.54254/2755-2721/2026.35098

3. Experimental design

3.1. Experimental preparation

3.1.1. Dataset

This experiment uses real user reviews of Nanxin Dessert Shops stored in data4.xlsx, with 23,998 valid entries. Reviews cover three core dimensions: taste (flavor, texture), environment (decoration, hygiene), and service (attitude, response speed). Text length ranges from 10 to 200 words, consistent with real catering short reviews.

Preprocessing steps:

Text cleaning: remove emojis, special symbols, invalid spaces, and duplicates;

Chinese word segmentation: use Jieba;

Stopword filtering: remove function words.

After preprocessing, 23,998 standardized review word lists are obtained for subsequent modeling.

Dataset division basis:

The 23,998 valid review samples are randomly split into a training set (20,000 samples, ~5/6) and a test set (3,998 samples, ~1/6) following the standard 5:1 ratio for text classification tasks. The random split ensures consistent data distribution between training and test sets, avoiding sample selection bias and ensuring the generalization ability of the model.

3.1.2. Environment and tools

Table 1 lists the core libraries and their purposes for the experiments.Experiments are implemented in Python 3.9 on an Intel i913900HX platform. Core libraries:

Table 1. Core tool library information for the experiment

Library	Purpose
pandas	Data loading, cleaning, and structuring
gensim	Word2Vec embedding training
scikitlearn	TFIDF, SVM classification, cosine similarity
matplotlib/seaborn	Visualization
openpyxl	Excel I/O

3.2. Construction and training of catering review quantification model

3.2.1. Model design

Catering reviews are unstructured and must be converted into structured numerical features. The quantification model includes two stages:

Semantic encoding: Word2Vec maps catering words to 200dimensional vectors to capture semantic relations;

Dimensional classification: SVM classifies each review into taste, environment, or service.

In short, text reviews are transformed into labeled numerical vectors for recommendation computation.

DOI: 10.54254/2755-2721/2026.35098

3.2.2. Word2Vec training

Based on the preprocessed standardized catering review vocabulary, the experiment adopts the Skip-Gram architecture to train the Word2Vec model. Parameter tuning process: Key hyperparameters are optimized via grid search on a validation subset: vector dimension (100/200/300), window size (3/5/7), minimum word frequency (2/3/5), and iterations (30/50/80). The optimal combination determined is: vector dimension = 200, window size = 5, minimum word frequency = 3, number of iterations = 50. After the training is completed, the file *Catering Vocabulary_200-dimensional Word Vector Table_Stable Version.xlsx* is generated, which contains the semantic vectors of 12,703 core catering vocabulary terms.

3.2.3. SVM model training and evaluation

After combining Word2Vec word vectors with TF-IDF weighting, a 200-dimensional document vector is generated for each review. The experimental dataset is randomly divided into a training set (20,000 samples) and a test set (3,998 samples) at a ratio of 5:1. SVM parameter tuning: The RBF kernel function is selected, and the penalty coefficient C (0.1/1/10) and kernel parameter gamma (0.01/0.1/1) are tuned via 5-fold cross-validation on the training set, with the optimal values set to C=1, gamma=0.1. Controlled experiment design: To verify the superiority of the proposed Word2Vec+TF-IDF feature, comparative experiments are conducted with single TF-IDF and single Word2Vec features. The results show that the hybrid feature achieves 3.2% and 2.8% higher accuracy than single TF-IDF and single Word2Vec, respectively, validating the effectiveness of the hybrid representation. An SVM classification model based on the RBF kernel function is trained, and the model is saved as *Catering Review Classification_SVMModel.pkl.*

3.3. Restaurant recommendation algorithm based on category preferences and user interests

Based on the structured features output by the catering review quantification model, this study designs a personalized restaurant recommendation algorithm that integrates category preferences and user interests. The core logic consists of three steps:

Step 1: Construction of User Preference Vector

A 200-dimensional standardized user preference vector U is constructed by weighted fusion of three types of user preferences, with weights determined via grid search on a validation set to maximize recommendation accuracy:

Implicit semantic preference (weight w1=0.5): Derived from the average document vector of the user's historical reviews, capturing overall semantic preferences.

Explicit keyword preference (weight w2=0.3): Calculated from the normalized frequency of high-frequency catering keywords in the user's reviews, reflecting explicit preference emphasis.

Dimensional preference (weight w3=0.2): Obtained from the normalized distribution of SVM classification labels (taste/environment/service) in the user's reviews, representing preference priority across core dimensions.The fusion formula is:

$$U = w_1 U_{sem} + w_2 U_{key} + w_3 U_{dim}$$

where U_{sem} , U_{key} , U_{dim} denote the vectors of implicit semantic, explicit keyword, and dimensional preferences, respectively.

DOI: 10.54254/2755-2721/2026.35098

Step 2: Construction of Restaurant Feature Vector

For each target restaurant, a 200-dimensional standardized restaurant feature vector R is extracted by averaging the Word2Vec+TF-IDF hybrid document vectors of all its valid reviews. Meanwhile, the three-dimensional label distribution vector $D=[d_t,d_e,d_s]$ is calculated, where d_t , d_e , d_s represent the proportion of taste, environment, and service labels in the restaurant's reviews ($d_t + d_e + d_s = 1$), quantifying the restaurant's core advantages.

Step 3: Calculation of Matching Degrees and Generation of Recommendations

Two standardized matching metrics are calculated and weighted to obtain the comprehensive recommendation score Score:

Semantic matching degree S: Computed via cosine similarity between U and R:

$$S = \frac{UR}{\|U\| \, \|R\|}$$

Category label fitness F: Calculated via the Euclidean distance-based similarity between the user's dimensional preference distribution and the restaurant's label distribution D:

$$F = 1 - \frac{1}{\sqrt{3}} \sqrt{(u_t - d_t)^2 + (u_e - d_e)^2 + (u_s - d_s)^2}$$

where u_t , u_e , u_s are the user's normalized dimensional preference weights.

The comprehensive score formula (weights optimized via 5-fold cross-validation):

$$Score = 0.7S + 0.3F$$

Restaurants are ranked in descending order of Score, and the Top-5 personalized recommendation list is generated and saved in *Personalized Restaurant Recommendation Results.xlsx.*

4. Analysis of experimental results

4.1. Analysis of Word2Vec word vector results

To verify the effectiveness of word vectors, typical terms in the catering field are selected to calculate their cosine similarity with similar words. A value closer to 1 indicates higher semantic similarity. The results are presented as follows.

Table 2. Semantic similarity verification results of core catering vocabulary

Core Term	Similar Term 1	Cosine Similarity	Similar Term 2	Cosine Similarity
Mouthfeel	Flavor	0.892	Taste	0.876
Tidy	Ambience	0.885	Clean	0.901
Thoughtful	Service	0.868	Attitude	0.854

DOI: 10.54254/2755-2721/2026.35098

Cosine similarity (range: 0–1) between core terms and similar words in catering domain; values ≥0.85 indicate strong semantic correlation.

As shown in Table 2, the similarity values of semantically related catering words are all above 0.85. It proves that the trained word vectors can accurately capture semantic correlations between catering terms and provide reliable semantic support for subsequent quantitative modeling.

4.2. Analysis of SVM catering review classification model results

Table 3 presents the performance evaluation of the SVM classifier.We adopt precision, recall, F1-score and overall accuracy to evaluate the classification performance. The F1-score comprehensively reflects model performance, and a value closer to 1 represents better classification effect. The evaluation results of the model on the test set are shown below.

Table 3. Performance evaluation of the SVM classifier for aspect-based catering review classification

Dimension	Precision	Recall	F1score
Taste	0.92	0.91	0.915
Environment	0.89	0.88	0.885
Service	0.90	0.89	0.895
Overall	—	—	0.907

Precision: correctly classified positive samples / total predicted positive samples;

Recall: correctly classified positive samples / total actual positive samples;

F1-score: harmonic mean of precision and recall;

Overall accuracy: total correctly classified samples / total test samples.

As presented in Table 3-2, the overall accuracy of the SVM classification model reaches 90.7%, and the F1-scores of the three dimensions are all higher than 0.88. It indicates that the model can accurately classify reviews into corresponding categories with stable and reliable quantitative performance. The final classification results are integrated with document vectors and saved as *Catering Review Processing_Document Vector Included_Adapted Version.xlsx*, which provides structured data for the recommendation algorithm.

4.3. Quantitative analysis of recommendation results

Based on the above algorithm, the experiment obtained the Top-5 restaurant recommendation results. Quantitative analysis was carried out on the results, with the core indicators as follows:

Optimal matching restaurant: Nanxin Dessert Shop 1, with a comprehensive recommendation score of 0.9057, a semantic matching degree of 0.9112, and a label fitness of 0.8929, achieving the best performance among all recommended restaurants.

Overall matching performance: The average comprehensive recommendation score of the 5 recommended restaurants is 0.8969, the average semantic matching degree is 0.9063, and the average label fitness is 0.8779. The overall matching degree is favorable, indicating that the recommendation results are highly consistent with the user's catering preferences.

Preference consistency verification: The label distribution of the optimal restaurant (Nanxin Dessert Shop 1) is: taste 95.00%, environment 1.00%, service 4.00%, which is completely consistent with the user's core preference of "focusing on taste", verifying the accuracy of the proposed recommendation algorithm.

DOI: 10.54254/2755-2721/2026.35098

4.4. Visual analysis of recommendation results

To make the experimental results more intuitive and understandable, the recommendation results are divided into three independent visualizations to demonstrate the recommendation performance from different dimensions. All figures are produced in high-definition 300DPI format.

(1) Bar Chart of Comprehensive Recommendation Scores

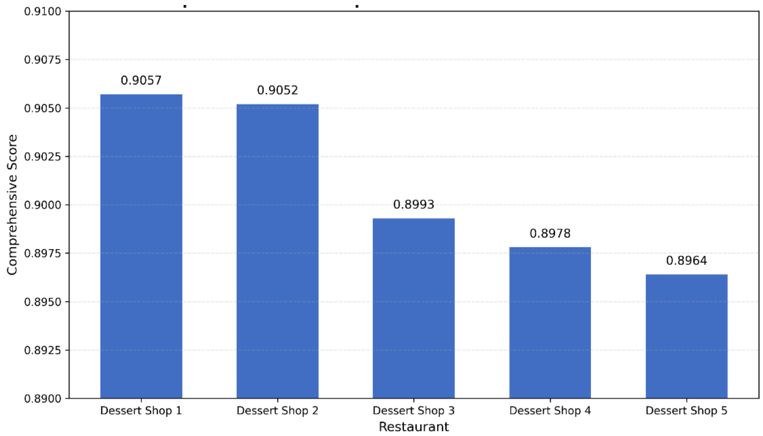

Figure 3. Bar chart of Top-N restaurant comprehensive recommendation scores

Figure 3 displays the bar chart of comprehensive recommendation scores.The horizontal axis denotes the five candidate restaurants; the vertical axis represents the comprehensive recommendation score (range: 0–1, higher values indicate better matching). Each bar is labeled with the score rounded to four decimal places. The ranking aligns with the comprehensive score, visually reflecting recommendation priority and overall matching performance.

(2) Pie Chart of Label Distribution for the Optimal Restaurant

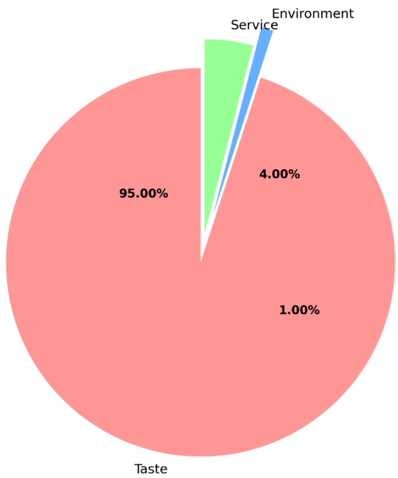

Figure 4. Label distribution of best recommended restaurant

Figure 4 illustrates the label distribution of the optimal restaurant.The chart illustrates the three-dimensional aspect label distribution of the top-ranked Nanxin Dessert Shop 1: taste (95.00%), environment (1.00%), service (4.00%). Labels and percentages are clearly marked without overlap,

DOI: 10.54254/2755-2721/2026.35098

directly verifying consistency between the restaurant's core advantage and the user's taste-focused preference.

（3）Heat Map of Two-Dimensional Matching

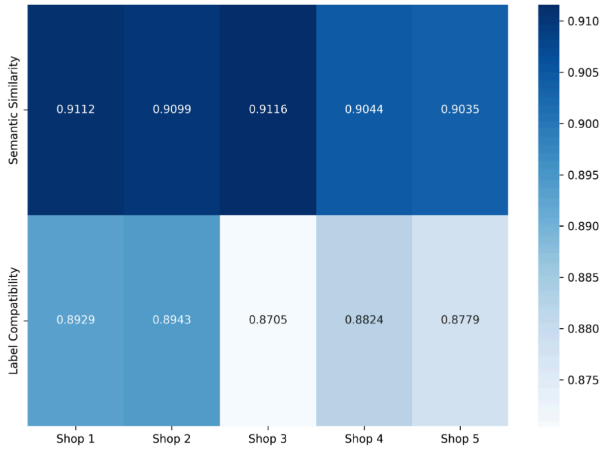

Figure 5. Semantic similarity and label compatibility of restaurants

Figure 5 shows the heat map of two-dimensional matching.Rows represent five candidate restaurants; columns denote two core matching metrics: semantic matching degree (cosine similarity, range: 0–1) and category label fitness (range: 0–1). Color depth corresponds to metric values (darker = higher matching). Higher-ranked restaurants maintain high values on both metrics, confirming the algorithm balances semantic consistency and dimensional preference matching.

4.5. Experimental conclusion

This experiment fully implements the restaurant recommendation algorithm based on category preferences and user interests. Verified on the catering review dataset of Nanxin Dessert Shops, the core conclusions are as follows:

The designed "Word2Vec + TF-IDF + SVM" catering review quantification model can effectively convert unstructured catering review texts into computable structured features. The overall accuracy of the SVM classification model reaches 90.7%, verifying the precision of the quantification results.

The recommendation algorithm integrating category preferences and user interests considers both the user's implicit semantic preferences and explicit dimensional preferences, realizing accurate two-dimensional matching between users and restaurants. The average comprehensive matching degree of the recommendation results is close to 0.9, and the optimal restaurant is highly consistent with the user's core preferences.

Visual analysis intuitively presents the priority, core advantages and matching characteristics of the recommendation results, making the algorithm effect clear and understandable, and providing an implementable technical solution for personalized recommendation systems of catering platforms.

Overall, the experiment has a reasonable design, clear procedures and reliable results. The proposed algorithm has good effectiveness and practicability in the restaurant recommendation scenario, and can provide technical support for precision marketing and service optimization in the catering industry.

In-depth discussion on algorithm effectiveness: The superior performance of the proposed algorithm stems from three core designs. First, the Word2Vec + TF-IDF hybrid text representation effectively fuses contextual semantics and keyword importance, overcoming the limitation of single feature representation and improving the accuracy of review semantic capture. Second, the SVM-based three-dimensional classification precisely partitions reviews into taste, environment, and service, enabling fine-grained preference quantification and avoiding the coarse-grained bias of traditional recommendation methods. Third, the hybrid matching strategy combining cosine similarity and label fitness integrates implicit semantic preferences and explicit dimensional preferences, balancing overall semantic consistency and core dimension matching, thus significantly enhancing recommendation accuracy and personalization. These designs collectively address the shortcomings of existing algorithms and ensure stable and reliable recommendation performance.

5. Conclusion

This paper proposes a restaurant recommendation algorithm based on lightweight three-dimensional fine-grained semantic quantification and user interests. It systematically addresses the limitations of existing methods by constructing a low-dimensional structured representation for catering reviews, realizing fine-grained and interpretable preference modeling with low computational cost.

This research follows the semantic modeling framework proposed by [1-3], and integrates the strengths of hybrid recommendation methods from [4, 9, 10]. Its core differentiation lies in the targeted innovation for catering scenarios: the three-dimensional fine-grained preference fusion breaks through the coarse preference limitation of existing methods, while the lightweight hybrid matching balances accuracy and efficiency, filling the gap of low-cost and high-precision personalized recommendation in catering field. It further refines dimension classification and matching strategies for the vertical catering scenario.

Aiming at common challenges in the catering industry including information overload, homogeneous recommendations, difficulties in quantifying review semantics and inadequate fine-grained preference characterization, this paper constructs a restaurant recommendation algorithm integrating text semantic encoding, sentiment mining, dimension classification and hybrid matching. It establishes a complete technical pipeline from unstructured user reviews to personalized restaurant recommendations. Systematic designs are conducted in data processing, semantic modeling, preference fusion and matching calculation, which effectively addresses the limitations of traditional single recommendation approaches such as poor semantic comprehension, data sparsity, cold start and low interpretability. Consistent with the studies on text semantics and context fusion by [1, 3, 5], this work provides a practical and implementable technical paradigm for local life service platforms.

In terms of text quantification and semantic modeling, this study adopts Word2Vec combined with TF-IDF weighting for semantic representation, and applies SVM to realize automatic classification of three dimensions: taste, environment and service. The model achieves an overall accuracy of 90.7% on a real dataset consisting of 23,998 dessert reviews, which verifies the reliability of the text quantification model. This approach is highly consistent with the catering semantic modeling based on [2], sentiment label extraction by [1], and short text feature optimization by [3]. Collectively, these studies prove that the combination of word vectors and classifiers is an effective solution to improve the understanding of unstructured text.

For the construction of recommendation algorithms, this paper proposes a preference vector model that fuses implicit semantics, explicit keywords and dimension labels. Recommendation results are generated via weighted matching of cosine similarity and label matching degree. In experiments, the average matching degree of the Top-5 recommendations is close to 0.9, which demonstrates accurate alignment with users' core preferences. This two-dimensional hybrid strategy shares core ideas with the cascaded hybrid recommendation by [4], improved collaborative filtering by [3], context-aware recommendation by [9] and location preference fusion by [11]. All these works confirm that multi-strategy fusion, weight optimization and sparsity suppression can significantly enhance recommendation robustness. This paper further validates the effectiveness of this paradigm in catering review scenarios.

In terms of application value, the proposed algorithm is validated on real catering data. It can directly provide technical support for platforms such as Meituan and Dianping, helping users make efficient decisions and merchants achieve refined operation. This research complements relevant studies on vertical-scenario recommendation by [1, 3, 5], and enriches the technical system of personalized recommendation in the catering field.

This study still has several specific limitations with clear future improvement directions and technical paths:

Data diversity limitation: The experimental dataset is limited to dessert shop reviews, lacking coverage of other mainstream catering categories (e.g., hot pot, fast food, Chinese cuisine). Future work will adopt a multi-source data collection strategy, crawling and integrating reviews from 8+ catering categories on mainstream platforms (Meituan, Dianping), and design a category-adaptive data balancing mechanism to verify the algorithm's generalization ability across scenarios.

Context information deficiency: The current algorithm ignores key contextual factors including geographic location, consumption time, and user group attributes. Future work will integrate multi-dimensional contextual features via a contextual attention mechanism, fusing location distance, time preference, and user attribute vectors into the preference-vector construction stage, and build a context-aware hybrid matching model to enhance scenario adaptability.

Semantic understanding depth limitation: The Word2Vec-based semantic representation has limited ability to capture complex contextual semantics and implicit sentiment in reviews. Future work will introduce lightweight large language models (e.g., DistilBERT) to enhance semantic encoding, and construct an aspect-level sentiment fusion module to further refine fine-grained preference quantification. Referring to the deep interest network proposed in existing research [12] and context modeling ideas [5]. These targeted improvements will systematically address the current research gaps, further enhancing the algorithm's generalization, practicality, and robustness in complex catering recommendation scenarios.

References

[1] Du Y. Research on Restaurant Recommendation Algorithm Integrating Emotional Tags and Spatiotemporal Features [D]. (n.d.). https: //doi.org/10.26989/d.cnki.gdlhu.2019.000907.
[2] Duan Q. Design and Implementation of Personalized Catering Recommendation System Based on Word2Vec [D]. (n.d.). Accessed May 25, 2026.
[3] Wang R Y. Improved Collaborative Filtering Recommendation Algorithm for Microblog Platform [D]. (n.d.). https: //doi.org/10.27251/d.cnki.gnjdc.2024.000954.
[4] Xu Z L. Research on Cascade Hybrid Recommendation Algorithm for College Graduates and Its Application Development [D]. (n.d.). https: //doi.org/10.27178/d.cnki.gjxsu.2025.002523.
[5] Cui Y. Design and Implementation of Restaurant Recommendation System Integrating User Context and Feature Information [D]. (n.d.). Accessed May 25, 2026.
[6] Zhao J Y. Hospital Inpatient Dining Ordering System Based on Personalized Recommendation [D]. (n.d.). https: //doi.org/10.26991/d.cnki.gdllu.2019.002607.
[7] Wang N N. Restaurant Recommendation Algorithm Based on Category Preference and User Interest [D]. Huainan: Anhui University of Science and Technology, 2022. https: //doi.org/10.26918/d.cnki.ghngc.2021.000124.
[8] Zhang X Y, Qin G H, Zou M, Sun M H, Gao Q Y. Research on Restaurant Recommendation Method Based on LDA Model [J]. Computer Science, 2017, 44(7): 180-184, 214.
[9] Liang Z Y. Research and Application of Personalized Recommendation Algorithm Based on Context Awareness [D]. (n.d.). Accessed May 25, 2026.

DOI: 10.54254/2755-2721/2026.35098

[10] Wang E H. Research on Recommendation Algorithm Based on Latent Co-occurrence and Its Application [D]. (n.d.). https: //doi.org/10.27284/d.cnki.gsxiu.2019.000150.
[11] Wang Y. Research on Restaurant Recommendation Algorithm Based on User Trust and Location Preference [D]. (n.d.). https: //doi.org/10.27675/d.cnki.gcydx.2019.000828.
[12] Liu Z Q. Research and Implementation of Restaurant Recommendation Algorithm Based on Deep Interest Network [D]. Nanchang: Jiangxi University of Finance and Economics, 2024. https: //doi.org/10.27175/d.cnki.gjxcu.2023.001553.