GEORGE BROWN PRESS
JOURNALS, BOOKS & PROCEEDINGS

Journal of Computer, Signal, and System Research

Article

A Personalized Recommendation System for Housing Information Based on Deep Learning

Su Yu [1,*]

[1] Beijing Panwen Information Technology Co., Ltd., Beijing, 102300, China
* Correspondence: Su Yu, Beijing Panwen Information Technology Co., Ltd., Beijing, 102300, China

Abstract: In view of the problems of complex housing information in the real estate market and the difficulty for users in finding houses, the personalized recommendation system of housing information based on deep learning is studied. The system adopts a hierarchical architecture. The data layer collects multi-source data from the housing information platform and stores it in MySQL and MongoDB databases respectively according to the structural characteristics. The preprocessing layer improves the data quality by removing noise, processing stopped words and eliminating duplicate data. The recommendation layer uses the convolutional neural network to generate the recommendation results. The service layer realizes invalid housing filtering, grouping and sorting optimization, and recommendation reason adding functions. The experiment shows that the recommendation reliability of the system is high, and significantly improves the efficiency of user house hunting, greatly shortens the house search and transaction time, and has a good effect of house information recommendation.

Keywords: deep learning; housing information; personalized recommendation

1. Introduction

With the wave of digital transformation sweeping across industries, the real estate market is also facing challenges and opportunities brought by the information explosion. With the rapid development of Internet technology, the number of housing resources on housing information platforms has increased exponentially, leading to an overwhelming amount of complex housing data. For home buyers or renters, it is difficult to accurately locate houses that meet their needs amidst this vast sea of information [1]. The traditional housing search method, which relies on manual filtering by users, not only consumes significant time and energy but is also highly inefficient. Furthermore, because users often struggle to comprehensively consider all factors, finding an ideal house that fully meets both their preferences and practical needs becomes challenging. At the same time, deep learning technology has made significant progress in recent years. Its ability to automatically learn complex patterns from massive datasets provides new approaches to addressing the challenges of housing recommendation [2]. Given these challenges, this paper proposes a personalized housing recommendation system based on deep learning. This system aims to improve user efficiency, optimize market resource allocation, and inject new vitality into the digital transformation of the real estate industry [3,4].

2. Overall Architecture Design of the Housing Information Personalized Recommendation System

In the housing information recommendation scenario, housing data contain multiple types of features, which deep learning algorithms can simultaneously process without requiring complex manual feature engineering. To address this, this paper investigates a personalized housing recommendation system based on deep learning and designs its overall architecture, as shown in Figure 1.

Received: 16 February 2025
Revised: 28 February 2025
Accepted: 16 March 2025
Published: 20 March 2025

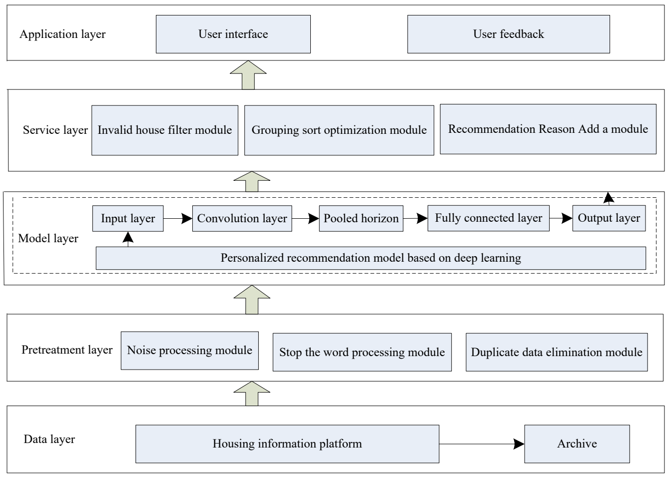

Figure 1. The overall architecture of the personalized recommendation system for housing information.

The system facilitates data interaction between the client and the database via the server. The server is built using Web Service and the MyBatis framework and communicates with the client via HTTP [5]. The architecture is divided into five layers: data layer, preprocessing layer, model layer, service layer, and application layer. The data layer collects housing and user data from the housing information platform and stores them in the database. The preprocessing layer performs noise removal and stop word processing to enhance data quality. The model layer employs deep learning algorithms, such as convolutional neural networks, to construct a personalized recommendation model. The service layer refines the recommendation results. The application layer provides users with a visual interface to display recommended housing information and collects feedback for model optimization. The hierarchical architecture fully leverages the advantages of deep learning algorithms to achieve efficient and accurate personalized housing recommendations.

3. Specific Module Design of the Housing Information Personalized Recommendation System

3.1. Functional Design of the Data Layer

(1) Data source: The system mainly collects data from the housing information platform. This data includes basic house information, such as area, house type, floor, orientation, and decoration status, as well as geographical location details, including community, transportation, and amenities. Additionally, user behavior data on the platform, such as browsing records, collection preferences, search keywords, and consultation history, is gathered for analysis.

(2) Data storage: For structured data, the MySQL database is chosen to facilitate efficient querying and management. For unstructured data, such as user reviews, the MongoDB database is chosen to better accommodate diverse data storage requirements.

3.2. Functional Design of the Pretreatment Layer

The preprocessing layer in this paper mainly processes the collected text data (such as housing descriptions, user evaluations, and search keywords) to improve data quality and provide a reliable basis for subsequent analysis and recommendations. It involves three key steps: noise removal, stop word processing, and data deduplication.

(1) Remove noise: the collected housing information is often mixed with special symbols and punctuation marks, which affects the information understanding. So this noise

T needs to be removed $N = \{n_1, n_2, \ldots, n_k\}$ to make the data T_{clean1} purer. Let the original information $T_{clean1} = T - N$ data be, the noise character T set N be, and the data after noise removal be. The noise removal operation can be expressed as a text operation that removes all characters or strings, which can be implemented using regular expressions.

(2) Processing stop words: stop words are words that appear frequently but do not contribute much to the semantic meaning in the housing information data, such as "yes", "yes", "in $S = \{s_1, s_2, \ldots, s_l\}$", etc. Removing the stopped T_{clean2} words can reduce the data dimension and improve $T_{clean2} = T_{clean1} - S$ the processing S T_{clean1} efficiency. Let the set of stop words be, and the text after removing the stop words be as. The stop word removal operation can be expressed as a text operation that eliminates all stop words from the text.

(3) Deduplication: there may be duplication in the data, which will affect the analysis results. In this paper, we identify and eliminate duplicate data $C = \{T_1, T_2, \ldots, T_n\}$ by $T_j T_i CS(T_i, T_j)$ calculating $CS(T_i, T_j) > \theta$ the θ data cosine T_i T_j similarity. Let the data set be, and the cosine similarity of the data sum be. If (for similarity threshold), it is considered and repeated and one can be retained.

Through the above steps, the collected text data is processed, and the high-quality, non-noise and non-repetitive housing information data is obtained, providing high-quality data for the subsequent personalized recommendation.

3.3. Functional Design of the Recommended Layer

In this housing information personalized recommendation system, the recommendation layer selects the convolutional neural network in the deep learning algorithm to realize the personalized recommendation of housing information. The algorithm mainly consists of the following five parts, which gradually extracts valuable features from the housing and user information obtained by the preprocessing layer to generate personalized housing recommendations for users.

(1) Input layer design: The input layer is the entrance of the entire convolutional neural network, which is responsible for transforming the housing and user information obtained by the pre-processing layer into a vector form suitable for model processing. n In this paper, the text vectorization method is selected as the processing means of the input layer. Its core idea is to map the input housing and user-related information to the dimension vector space with the help of word embedding, so that the text information can be transformed into numerical form to facilitate the calculation and learning of the model.

Suppose the data input from the preprocessing layer is in $\left(y_i, b_{i0} \oplus b_i \oplus \ldots \oplus b_{in}, \frac{1}{n}\sum_{j=1}^{n} h_{ij}\right)$ a format $\oplus$ where the house information y_i is $i$connected b_{i0} by b_{in} convolution characters. Here represents the first user, including various $\frac{1}{n}\sum_{j=1}^{n} h_{ij}$ detailed information of the house, such as house area, house type, decoration, etc., is the sum of information $\left(y_i, p_i, \bar{h}_i\right)$ related p_i to the user's preference for all aspects of the house. Through $\bar{h}_i$ the integration and processing of the input data, the user information $u_i$of the house can be obtained. All the information representing the house is a comprehensive expression of the aforementioned detailed information of the house, the relevant information indicating the user's preference, and comprehensively reflects the user's preference of the preference for the house. Represents the combined housing information data through the vector:

$$u_i = Doc2VecC(p_i) \tag{1}$$

Through the above procedure $Doc2VecC$, n the function is returned to the dimension vector. At this point, the user sample data is converted into the following vector:

$$L = \left(y_i, u_i, \bar{h}_i\right) \tag{2}$$

Among them L, this vector combination includes user identification, housing information vector, and user preference information, which are passed to the subsequent convolution layers as input.

(2) Convolutional layer design: the main function of convolution layer is to convolve the information after text vectorization processing $b^j \in R^1$, so as to extract the characteristics of housing related information. Using the features after convolution, the convolution process can be expressed as:

$$b_j^i = f\left(M_j^b \oplus p\big(:, i: (i + M_S - 1)\big)\right) + r_j^b \quad (3)$$

Where, representing $\oplus$ the M_j^b convolution j operation, is the first convolution kernel, which is like a feature extractor $p\big(:, i: (i + M_S - 1)\big)$, able to extract p specific characteristic i, $i + M_S - 1$ patterns from the input M_S content. The fragment of information taken from the first $r_j^b \in R$ position to each position from the input data matrix is the width of the f convolution kernel and determines the range $f(x) = max(0, x)$ of information $b_j \in R^{i-Ms+1}$ covered M_j^b by each convolution $b_j = \left[b_j^1, b_j^2, \ldots, b_j^i, \ldots, b_j^{i-Ws+1}\right]$ operation. Represents the bias vector, adding a constant offset to the convolution result, increasing the flexibility of the model. The ReLU activation function of the convolutional layer, expressed using eigenvectors, is applied through the convolution process of the convolution kernel. Through the above process, using multiple sets of different convolution kernels, a variety of different features can be extracted from the input content, which will help the model to more fully understand the relationship between housing information and user preferences.

(3) Pooling layer design: After the processing of the convolution layer, this paper obtains a large amount of feature information, but these feature information has a high dimension, which may contain some redundant information, but also easy to lead to model overfitting. To solve these problems, this paper uses the pooling layer to reduce the features acquired by the convolution layer.

Adopted $D_t = \{d_1, d_2, \ldots, d_{z-s+1}\}$ and G_t set t as the characteristics of the first convolution D_t layer. Select the maximum value as the output of the layer, and the pooling layer processing process can be expressed as:

$$G_t = max(D_t) = max\{d_1, d_2, \ldots, d_{z-s+1}\} \quad (4)$$

This maximum value pooling method can retain relatively important features in convolutional neural networks, because the maximum value tends to represent the most significant feature information in the range. By reducing the dimension, the overfitting can be avoided.

(4) Design of the full connection layer: the function of this layer is to extract the output of the previous layer, so as to obtain the implied features G_t of the house information. Enter the output m of the previous layer into the fully connected layer, assuming that the number of neurons is, the specific operation process of the fully connected layer can be expressed as:

$$\varphi_i = Re\,l\,u(w_t G_t + r_t) \quad (5)$$

Where, is the $\varphi_i \in R^m$ fully w_t connected r_t layer weight, indicating the w_t bias r_t coefficient. The input information is linearly combined by weight and bias, and then nonlinearly transformed φ_i through the Relu activation function to obtain the implied features of the house information. These implied features contain comprehensive information of housing information and user preferences and are an important basis for personalized recommendation of the model.

(5) Output layer design: The output layer is the last layer of the entire convolutional neural network, which is responsible for transforming the high-dimensional matrix with sparse property information obtained by the fully connected layer into the final personalized recommendation result of housing information. In this paper, we apply the matrix factorization method to the output layer of convolutional neural networks. After applying this method, the output of the output layer of the neural network model can be expressed as:

$$V_j = CNN(Z, A_j) \tag{6}$$

It represents V_j the housing recommendation information vector, which A_j contains the housing Z recommendation score and related feature information for the user information input; it is the weight set. Through matrix decomposition, the high-dimensional information matrix is decomposed into low-dimensional V_j feature matrix. These low-dimensional feature matrix can represent the key information of houses more simply, and also better reflect the personalized needs of users. The final housing recommendation information vector is the personalized recommendation result of housing information.

3.4. Functional Design of the Service Layer

This layer is primarily responsible for processing the personalized recommendation results for houses. It includes the following functions:

(1) Invalid housing filter module: Filters out sold-out or rented houses. If the house status is marked as 'sold out' or 'leased', it will be removed from the recommended list.

(2) Group and sorting optimization module: Groups the recommended list based on certain characteristics of the houses, such as area, house type, etc. For example, the houses are grouped by region and then sorted within each group according to their recommendation score, enhancing the readability and relevance of the results.

(3) Recommended reasons module: Provides reasons for each house recommendation. The recommendation reasons are generated by analyzing the user's historical behavior and house characteristics. For example, if a user frequently browses houses with two rooms and one living room, and the recommended house also fits this criterion, the recommendation reason might be "Based on your browsing history, this two-bedroom, one-living room house may meet your needs."

4. Application Effect Analysis

In order to comprehensively evaluate the performance of the personalized recommendation system based on deep learning, this paper verifies its recommendation effect through experiments. The effectiveness of the system is investigated from different angles to provide strong support for the practical application of the system.

(1) Recommended reliability test: the matching degree of the house recommended by the test system and the actual needs of users. Select a data set that contains a large amount of user history and behavior data (browsing, collecting, consulting, etc.) and the corresponding housing information. The data set was proportional between the training set and the test set. The convolutional neural network model was trained using the training set. In the test phase, the user information in the test set is input into the trained model to obtain the list of houses recommended by the system. The cumulative gain of normalized loss (NDCG) was used as the evaluation index. NDCG considers the correlation score of each house in the recommended list and the location of the house in the recommended list, which can be calculated by formula (7):

$$NDGG_i = \frac{DCG_i}{IDCG_i} \tag{7}$$

The loss of DCG (Discounted Cumulative Gain) is the cumulative gain of the actual recommendation list, obtained by the weighted sum of the correlation scores of each house in the list, with the weighting factor depending on the position of the house in the list. The ideal cumulative gain is the cumulative gain when the recommendation list is arranged in descending order based on the correlation with user demand. The housing information was recommended 10 times, and the reliability of the recommended results was analyzed based on the above evaluation indicators. The analysis results are shown in Table 1.

Table 1. Reliability analysis of the recommended results.

Recommended times / times	The cumulative gain of the normalized loss
1	0.83
2	0.87
3	0.82
4	0.84
5	0.88
6	0.86
7	0.85
8	0.82
9	0.86
10	0.88

According to the 10 recommendations in Table 1, the NDCG value is basically kept between 0.82 and 0.88. Overall, these values are at a high level, indicating that the houses recommended by the system match well with the actual needs of users. This means that the system can more effectively recommend users to meet the needs of the houses according to the user's historical behavior, which has a high practical value in the field of housing recommendation.

(2) This section analyzes whether the efficiency of user search can be effectively improved after the application of this recommendation system. Twenty users were selected from the housing information platform and randomly divided into two groups: the experimental group and the control group. The number of users in the two groups was as balanced as possible, with a random assignment approach used for grouping. Users in the experimental group used the personalized recommendation system designed in this paper, while users in the control group used the original traditional house screening and display function of the platform. Both groups were tested for one month, analyzing the average time from the start of the search to the collection of the first house of interest, as well as the average time from the start of the search to the completion of the transaction. These two time indicators are used to measure the improvement in users' house-hunting efficiency. The specific analysis results are shown in Table 2.

Table 2. Analysis of user room search efficiency.

user group	Average time/day to start finding a house and collect the first house of interest	Average time/day to start the house search and to close the transaction
experimental group	3.5	12
control group	6	20

It can be seen from the data in Table 2 that, after using the personalized recommendation system, the average time for users in the experimental group to find and save interested houses is greatly reduced, 2.5 days less compared to the control group. Additionally, the average time to complete transactions is also decreased, with a reduction of 8 days compared to the control group. This fully demonstrates that the personalized recommendation system has significantly improved the efficiency of users' house hunting.

5. Conclusion

This paper designs and constructs a personalized recommendation system for housing information based on deep learning to realize efficient and accurate housing recommendation through hierarchical architecture design. The application effect analysis shows that the NDCG value in the recommended reliability test is stable at 0.82~0.88, indicating that the recommended house matches the user demand. The user house-hunting efficiency shows that the experimental group using the system has significantly reduced transaction time. Therefore, the system holds high practical value, effectively improving the service quality of the real estate platform and providing users with an efficient house-hunting experience.

References

1. S. Sharma, V. Rana, and V. Kumar, "Deep learning based semantic personalized recommendation system," *Int. J. Inf. Manag. Data Insights*, vol. 1, no. 2, p. 100028, 2021, doi: 10.1016/j.jjimei.2021.100028.
2. F. Liu and W. W. Guo, "Research on house recommendation model based on cosine similarity in deep learning mode in grid environment," in *2019 Int. Conf. Virtual Reality Intell. Syst. (ICVRIS)*, IEEE, 2019, doi: 10.1109/ICVRIS.2019.00039.
3. S. Zhang, et al., "Deep learning based recommender system: A survey and new perspectives," *ACM Comput. Surv. (CSUR)*, vol. 52, no. 1, pp. 1-38, 2019, doi: 10.1145/3285029.
4. X. Yuan, et al., "Toward a user-oriented recommendation system for real estate websites," *Inf. Syst.*, vol. 38, no. 2, pp. 231-243, 2013, doi: 10.1016/j.is.2012.08.004.
5. E. G. Muñoz, et al., "Housing fuzzy recommender system: A systematic literature review," *Heliyon*, vol. 10, no. 5, p. e26444, 2024, doi: 10.1016/j.heliyon.2024.e26444.