Research on the Construction of Housing Recommendation and Rights Protection Platform Based on Data Mining

Xiaoshu Peng[1](✉), Siying Li[1], Junjie Peng[1](✉), and Silei Li[2]

[1] Zhuhai College of Science and Technology, Jinwan, Zhuhai, Guangdong, China
1797215019@qq.com, 823750063@qq.com
[2] Nanfang College Guangzhou, Conghua, Guangzhou, Guangdong, China

Abstract. The purpose of building a platform for house purchase recommendation and rights protection is to provide consumers with an intelligent recommendation platform that can reduce the risk of house purchase. This paper first analyzes the risks that home buyers may encounter in the process of buying houses at the present stage. Secondly, this paper introduces the construction scheme of the housing recommendation and rights protection platform. Finally, this paper expounds the research on the recommendation function of the housing recommendation platform. Its innovation lies in the use of new data mining technology to help consumers make more sensible purchase decisions and protect the rights and interests of consumers.

Keywords: Data mining · Recommendation system · Rights protection platform · DDCF

1 Introduction

In recent years, the rapid development of the real estate industry has also made it difficult to form a unified regulatory system in the short term [1]. At present, research teams at home and abroad have made research on this. However, due to the different purchase and after-sale policies at home and abroad and different ways of safeguarding rights, foreign consumers are more inclined to entrust agents to protect their rights, while domestic consumers prefer the difference of group or unwilling to protect their rights.

Based on the above situation, the domestic housing recommendation and rights protection platforms need to make more differentiation and convenience according to the local public conditions while reducing the risk of buying dangerous housing sources. Therefore, this paper puts forward a new type of [3] purchase recommendation and rights protection platform structure combined with dual decoupling collaborative filtering algorithm technology, to provide a new idea for domestic purchase recommendation and rights protection platform.

R. B. B. M. Hussain et al. (Eds.): ICHSSR 2023, ASSEHR 765, pp. 844–850, 2023.
https://doi.org/10.2991/978-2-38476-092-3_103

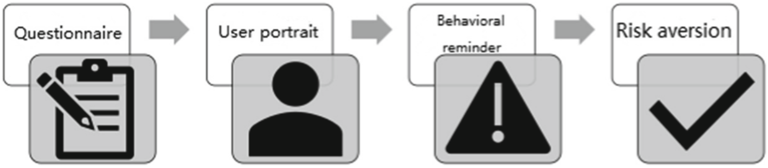

Fig. 1. Purchase risk avoidance process

2 The Construction Plan of the Housing Recommendation and Rights Protection Platform

Platform first in the form of questionnaire survey intelligent analysis user portrait to help users find out the problem, and based on big data technology recommend good housing to users, as far as possible to avoid the hidden danger of dangerous housing, help consumers to make more sensible purchase decisions, at the same time for the rights demand users to provide online rights channels.

2.1 Platform Process

For the users with house purchase demand: to provide risk aversion. For users in the process of buying houses: guide them to conduct user portraits in the form of questionnaire survey, and find out the problems [4] in the purchase process in time. For the users who need to protect their rights: recommend the rights protection methods according to the analysis of the user conditions. So as to help users to enhance the legal awareness, to find a way to protect their rights (Fig. 1).

2.2 Main Functions of the Platform

Housing Source Recommendation
The housing recommendation function of the housing recommendation and rights protection platform can help home buyers make more sensible decisions and reduce the risk of costly mistakes when buying houses.

Intelligent Answer Questions
In the form of questionnaire survey, the answers to the current real estate problems are obtained and big data is applied. By analyzing the situation of related problems and combining with the existing information, the latest answers are obtained to solve the potential problems in the purchase process of users.

Legal Advice
Analyze user characteristics, provide users with professional legal consultation functions by matching lawyers or artificial intelligence, and help users to strive for the maximum rights and interests in the most reasonable and legal way.

3 Double Decoupling Collaborative Filtering Algorithm

Double Disentangled Collaborative Filtering (DDCF) [6] can be used in recommender systems to perform double decoupling during scoring modeling of explicit feedback, resulting in more robust and interpretable hidden vector representations. It divides the traditional shared hidden space into two new hidden spaces, naming them intent recognition networks and preference decomposition networks, respectively.

As shown in Fig. 2, DDCF is designed to perform dual decoupling representation learning to better model user-housing interaction records. DDCF can learn two hidden vector representations of user and housing source, namely intention distribution representation and preference representation.

As shown in Fig. 3, the user behavior may be driven by multiple intentions. The attraction to users may come from many fine-grained concepts, such as the attention to housing, convenient transportation, and price. So different intentions occupy in end-user decisions. Here, a user intention is defined as a probability distribution over all listings, while a representation of a user's intention distribution is a probability distribution over all intentions. Then, based on the obtained intention distribution representation, the user's score record R_i is further decomposed into segmented input $R_{il} \in \mathbb{R}^M$ according to different intentions, and the final score prediction is performed by jointly considering the predicted scores under L intentions.

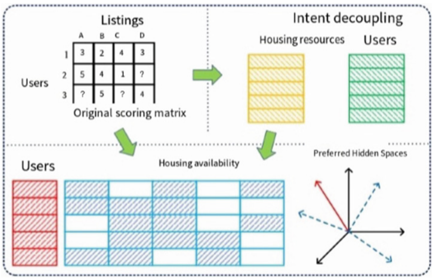

Fig. 2. Scoring modeling process diagram

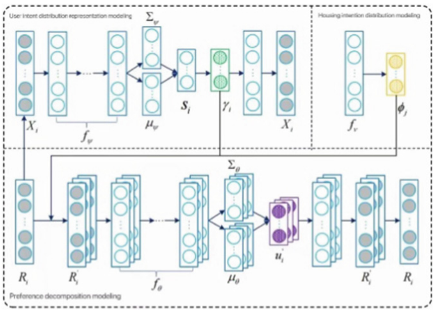

Fig. 3. Diagram of the network structure

3.1 Intention Recognition Modeling

Assuming there are a total of K intents in the data, each user's 0/1 score record X_i. It is generated by a mixture of scores under K intents $\beta = (\beta_1, .., \beta_k) \in \mathbb{R}^{M\times K}$. A higher probability value in β_k means that the listing is more likely to belong to the intent, and the sum of the probability values of all listings in each intent should be equal to the distribution of user intent for the ith user, $\gamma_i \in \mathbb{R}^K$ is defined as a probability distribution on all intents, The distribution follows the Dirichlet distribution: $\gamma_\text{i} \sim$ Dirichlet (α). Here $\alpha \in \mathbb{R}^K$ is the parameter to be learned, The kth of γ_i dimension represents the proportion of the kth intent. For example, $\gamma_\text{i} = (0.2, 0.3, 0.4, 0.1)$ means that the user's intent is related to four intents (20%, 30%, 40%, 10%).Assume that the user's 0/1 record, χ_i. There is, N_i, then X_i follows the following distribution: $X_i \sim$ Multinomial (N_i, β_{γ_i}). Under this assumption, the 0/1 record of the user, X_i. Marginal likelihood function of:

$$p(X_i|\alpha,\beta) = \int\limits_{\gamma_i} p(X_i|\beta,\gamma_i)p(\gamma_i|\alpha)d\gamma_i \tag{1}$$

As for obtaining the intention distribution representation $\phi \in \mathbb{R}^{M\times K}$ of the listings, the user representation γ_i can be used as a prior probability of interactive listings. Then use MLP as a posterior inference network $f_v\left(W_j\right)$:

$$\phi_j \sim Multinormal(1, \sigma(f_v\left(W_j\right))/\tau \tag{2}$$

Here W_j is the jth row of the first layer weight matrix of the network f_ψ so that the initial embedding of the listing is shared between the two inference networks. The KL divergence is used as follows to measure the distance between prior and posterior probabilities during variational inference:

$$\mathcal{L}_1 = \sum\nolimits_i \sum\nolimits_{j\in\{j|X_{ij}=1\}} \sigma(f_v(\frac{W_j}{\tau})^T log \frac{\sigma(\frac{f_v(W_j)}{\tau})}{\gamma_i} \tag{3}$$

3.2 Preference Decomposition Modeling

The above intention recognition modeling network to obtain the intention distribution of users and houses. In this subsection, in order to construct the decoupled preference representations, the user preferences are further decomposed into preferences under the corresponding respective intentions.

The user listing score R_{ij} can be viewed as a comprehensive result of the preferences of the ith user under all different intentions. Therefore, first decompose the user rating R_i into each intention. The resulting modified scoring order may differ substantially from the original observed scoring order. For example, house A score 5, while house B scores 3. However, with the intention of high correlation with B but not A, the modified scores for A and B may be 0.4 and 0.7, respectively.

The following are evidence lower bounds for modeling networks by preference decomposition:

$$-\mathcal{L}_2 = \mathbb{E}_{\boldsymbol{q}}\left[\log p_\theta\left(R^{'}|U^{'}\right)\right] - \eta\mathbb{KL}(q_\theta(U^{'}|R^{'})||p\left(U^{'}\right)) \tag{4}$$

4 Realize the Housing Source Recommendation Function

4.1 Introduction of the Data Set

Based on python's requests, pandas, urllib3 and other libraries, this article takes a considerable number of some real estate information data of a domestic city from the housing websites such as shell house hunting. This data set has a total of 22,462 lines of data, including 10 columns of features, including: the area where the house is located, the title of the introduction, the community, basic information, the release time and the number of followers, the way of house viewing, whether it is close to the subway, whether duty-free, total price (10,000), price per square meter.

4.2 Data Cleaning

Process the characteristics of the housing resources, extract the characteristics of each house, and use the correlation of the same characteristics in different houses to screen out the housing problems such as malicious price raising, inconsistent with the real description and a long release time.

4.3 Data Storage

The data storage component includes user data, historical data and other related data. The system plans to use the relational database and NoSQL database. To provide users with personalized advice to meet their specific needs and preferences by effectively organizing and storing large amounts of data.

4.4 Recommended Methods

First, use word cloud to draw word clouds of the user's favorite apartment features. According to the frequency of feature occurrence, DDCF scores according to the key frequency of the word cloud are introduced and ranked, so that the recommended results will be in line with the overall interest of the user (Fig. 4).

4.5 To Calculate the Overall Average Rating of the Platform

In the trial stage, to investigate the user experience of the platform recommendation function. We collected the scores of some users, and made a result display platform interface, platform services, recommendation function, and user experience rating table for each problem (Table 1).

The score of each question in the table is the score of each question by multiplying the number of points by the frequency of specific score points. In order to find the average, the following formula is introduced [7]:

$$\sum_{i=1}^{n} \frac{\sum_{i=1}^{4} fx_i}{4} \tag{5}$$

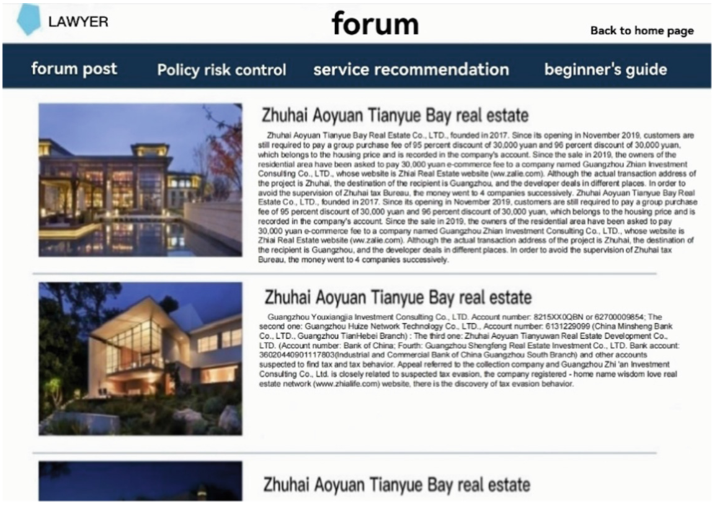

Fig. 4. Expected recommended results

Table 1. User experience rating table

S/N	QUESTION	V. GOOD [4]	GOOD [3]	FAIR [2]	POOR [1]	RATING
1	Are the recommender houses satisfactory?	9	6	4	1	3.15
2	Can the housing supply on the platform meet the demand?	11	6	3	0	3.4
3	Does the platform help you buy a house?	9	4	5	2	3
4	Is the interface design reasonable?	8	9	3	0	3.25
5	Is the service of the platform satisfactory?	13	6	1	0	3.6

To calculate the overall average rating of the platform:

$$\text{Averagesystemrating} = \frac{3.15 + 3.4 + 3 + 3.25 + 3.6}{5} = 3.28 \tag{6}$$

Therefore, we found that by referencing the recommendation obtained by the DDCF algorithm, a better recommendation effect can theoretically be obtained.

5 Conclusions

On the construction of a house purchase recommendation and rights protection platform under data mining, this paper uses new data mining [5] technology to collect and analyze the current problems in the real estate industry, and creates a new house purchase recommendation rights protection platform to reduce the user's purchase risk by recommending better housing listings for users, and also provide rights protection channels for users in need. However, at present, there are still certain limitations in the house purchase recommendation and rights protection platform, and the structure of the platform needs to be further improved. In addition, the research is still in the theoretical research stage, and a large number of experiments are still needed to verify and analyze.

References

1. Ding Xiaojing. Real estate bubble warning Research [D]. Hebei University of Technology,2011. [1] Ding Xiaojing. Research on Real Estate Bubble Warning [D]. Hebei University of Technology,2011.
2. Lin Danyao. Research on construction cost control in the implementation stage of real estate engineering project [J]. Ju She,2021(24):3-4.
3. Pan Meizhu. Tianjin state-owned small and medium real estate enterprise development strategy research [D]. Tianjin university, 2019. https://doi.org/10.27356/dcnki.Gtjdu.2019.000609
4. Wang Xiaoyang. Study on Problems and Countermeasures of Urban Affordable Housing in Liaoning Province [D]. Liaoning University, 2011.
5. Liying Li.Analysis and data mining of intellectual property using GRNN and SVM [D].Institute of Science and Technology, North China Electric Power University, Baoding, 071003, China.
6. Chao Wang. Personalized Recommendation Techniques and Applications Based on User Decision-making Behaviors[D]. Computer Application Technology. University of Science and Technology of China, 2022.
7. Ojokoh, B. A., Olayemi, O. C., Babalola, A. E. & Eyo, E. O. A User-Centric Housing Recommender System[J]. (ISSN 2220–3796). 2018