

LAWYER RECOMMENDATION SYSTEM BASED ON USER PROFILES AND COLLABORATIVE FILTERING

Yu Enbo^{1, a}, Wang Chunxi^{1, b}, Zou Xinlin^{1, c}, You Cancan^{2, d}, He Xiangyu^{2, e}

¹College of Cybersecurity, Sichuan University, Chengdu 610000, Sichuan, China;

²Law School of Sichuan University, Sichuan University, Chengdu 610000, Sichuan, China.

^a scu_enboy@163.com, ^b evernightstar@foxmail.com, ^c 1412871486@qq.com

^d eiheicc@163.com, ^e 1251750044@qq.com

Abstract. This study discusses the difficulties that ordinary citizens face when dealing with legal problems in the process of China's rule of law, including high lawyer fees, lengthy litigation processes, the shortcomings of existing lawyer recommendation methods, and the lack of channels for efficient matching between user needs and lawyers. Currently, when seeking legal rights protection, citizens usually rely on referrals from acquaintances or search engines to find lawyers, but these methods have some issues. Lawyer recommendation platforms are still in the early stages of development in China, with problems such as data gaps, opaque information sources, and quality issues. To address these problems, this study proposes a user profile-based lawyer recommendation system, aiming to provide citizens with more accurate and centralized legal aid services, as well as establish an efficient user-lawyer matching channel.

Keywords: Legal aid services; User profile; Collaborative filtering.

1. Introduction

Rule of law is a significant marker of human civilization's progress. Although the Chinese government's initiative of constructing a "rule of law in China" has greatly propelled the country's process of legalization, many ordinary citizens still find themselves perplexed when dealing with legal issues. They often face expensive legal consultation fees and lengthy litigation processes, which dissuade them from defending their rights. Therefore, effectively addressing the awakening and rational insufficiency of citizens' rights awareness, the shortcomings of existing methods, and the absence of efficient channels for matching user needs with lawyers becomes essential to enhance the efficiency of legal services, improve the legal service environment, and reasonably match the supply and demand of legal services.

Currently, most citizens' initial response when asserting their rights is to seek advice from a lawyer. However, the prevailing method of lawyer recommendations relies on acquaintances, which does not guarantee that the recommended lawyer possesses the necessary expertise or can address the citizens' legal issues. Hence, finding a suitable lawyer is not an easy task, and citizens often incur high costs of trial and error, which is why it is commonly said, "Finding a lawyer is difficult, finding a good lawyer even more so^[1]."

Besides consulting acquaintances for lawyer recommendations, another common approach citizens take when facing legal issues is searching for solutions via search engines. However, the results often include a mix of advertisements and spam, requiring significant time and effort to sift

through. Moreover, while some legal websites (like the "Judgment Document Website") offer search services, their functionality is not user-friendly enough for laypersons and does not meet the needs of citizens.

In China, the lawyer recommendation service industry is still in its early stages of development. Existing lawyer recommendation platforms often suffer from data deficiencies, opaque information sources, and issues with advertisements and false promotions. For example, the intelligent legal consultation feature of the China Legal Service Network produces mechanical and lexical combinations in its search results. The so-called "intelligent legal opinions" it generates fail to address complex practical issues, making it difficult for citizens to trust the content provided by these platforms.

In light of the aforementioned problems in domestic legal services, this paper explores a lawyer recommendation system based on user profiling, aimed at providing citizens with precise, centralized legal aid services and establishing an efficient user-lawyer matching channel.

2. Constructing a Multidimensional Portrait Model for Lawyers

Judgment documents are reliable and have a relatively fixed structure, making it feasible to assess a lawyer's performance during the trial process to obtain data. The main source of lawyer profiling data in this paper comes from the judgment documents publicly available on the "China Judgments Online."

The construction of lawyer profiles consists of three steps: data extraction and preprocessing, profile dimension design, and profile tag generation. The construction process of user profiles is the result of a collaborative effort between data and analysis, as detailed in Figure 1: (Please note that Figure 1 is not displayed here but would typically be included in the academic paper to visually represent the process).

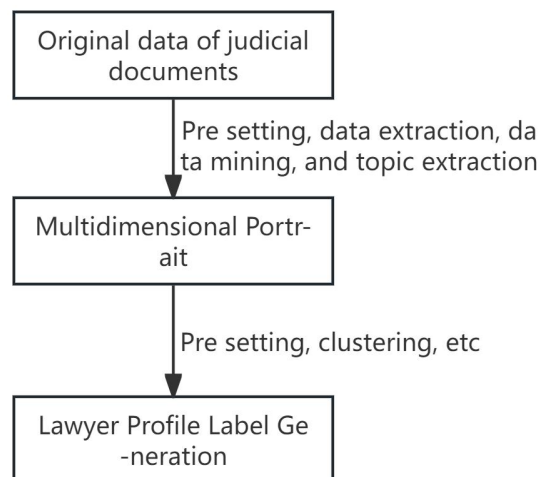


Fig. 1 The process of constructing a lawyer's profile

2.1 Data extraction and preprocessing

As the foundation for constructing user profiles, data is crucial for building lawyer profile models. Therefore, this paper initially requires the collection of comprehensive information on lawyers. To enhance the quality of the generated profiles, text mining technology is employed to

preprocess judicial documents, extracting key information about lawyers and minimizing the interference of irrelevant data in the modeling process.

As previously mentioned, judicial documents have a somewhat fixed format, which includes relatively fixed types of information and their positional features. By analyzing the structure of the documents and the indicative words or positions of effective information within them, an extraction plan targeting this information is designed.

In this paper, the Jieba segmentation tool was utilized to construct a lexicon for segmenting and part-of-speech tagging judicial documents, and the documents were processed into blocks based on their textual structure. On the basis of using regular expressions to identify case role entities, the extracted attribute information was matched with corresponding entity classes.

To better handle information related to grammatical structures, this paper also processed specific attributes based on part-of-speech tagging. For instance, in criminal cases, depending on different contextual keywords, it was determined whether the extracted time referred to the time of the crime, the time of arrest, or other times. The detailed analysis process is illustrated in Figure 2.

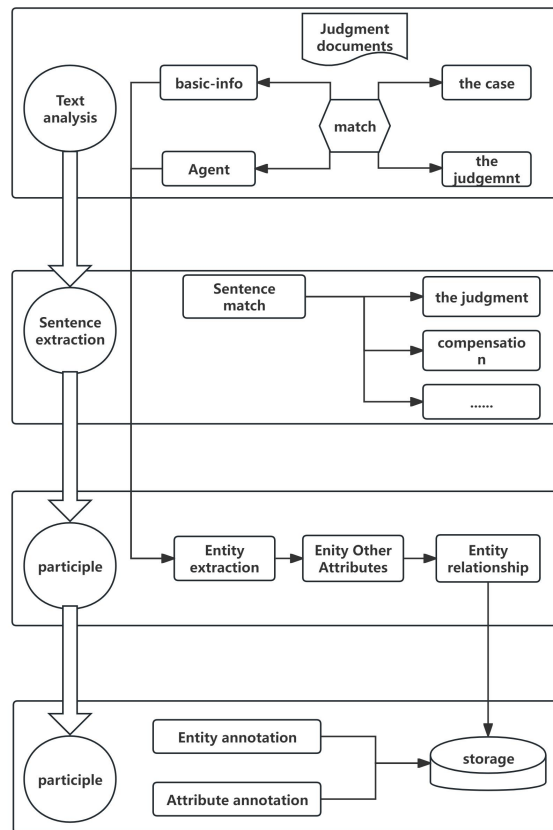


Fig. 2 Extraction algorithm process

After preprocessing the judicial document data, the data obtained is stored in a structured format, providing a solid foundation for subsequent tag generation and lawyer profile modeling. This structured storage ensures that the relevant attributes and entities extracted from the documents are organized in a manner that facilitates efficient retrieval and manipulation, crucial for building accurate and dynamic lawyer profiles.

2.2 Design of Lawyer Portrait Model

Judicial document data contains a wealth of case-related information, but not all data can be used to construct lawyer profiles. Therefore, when recommending the most suitable lawyer to users,

it is necessary to integrate data from various aspects of lawyers and user needs into a whole. This integration is achieved by analyzing the content features of the judicial documents and lawyer attributes, selecting appropriate dimensions to establish lawyer profiles.

This paper summarizes three indicators that can comprehensively encapsulate the characteristics of lawyers, based on field visits to courts, questionnaire surveys, and literature reviews: basic attributes of lawyers, domain attributes, and social attributes.

To establish a complete lawyer profile, it is necessary to collect and integrate multiple indicators including the lawyer's basic attributes, domain attributes, and social attributes. Basic attributes of a lawyer include common demographic characteristics such as gender, age, years of practice, geographical location, and educational background. Domain attributes mainly include the types of cases handled, the level of courts, success rates, and the law firms they belong to. Social attributes primarily encompass user reviews, areas of expertise, activity levels, fees, and the number of services provided.

These indicators are integrated to establish a comprehensive lawyer profile. By collecting and integrating these indicators, a detailed lawyer profile is generated, better assisting clients in finding the appropriate lawyer.

$$\text{Lawyer} = \langle \text{Demographics}, \text{LawyerAttr}, \text{Relation} \rangle \quad (1)$$

Demographics, *LawyerAttr*, *Relation* The three parameters respectively refer to the lawyer's basic information, domain attributes, and social relationships, as shown in Figure 3.

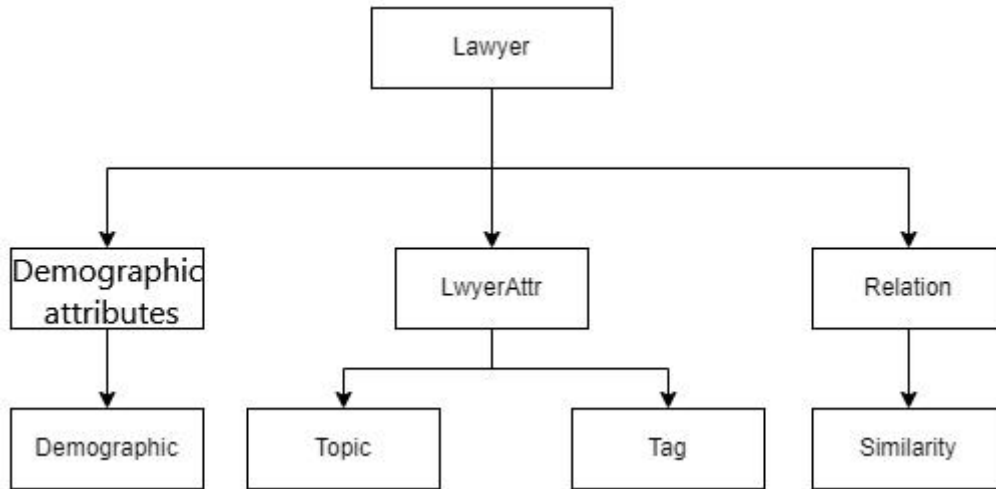


Fig. 3 Lawyer profile model

(1) *Demographics* Static Attribute Vector Model

Demographics = <Unique Identifier, Gender, Age, Education, Affiliated Institution>, where the unique identifier can be represented by a name, account, or an ID reserved in the system. Gender = <Male, Female>; Age = <Under 20, 21-30, 30-40, 40-50, 50-60, Over 60>; Education = <Bachelor's, Master's, Doctorate and above>; The affiliated institution is filled in according to the data mined.

(2) *LawyerAttr* Lawyer Domain Vector Model

This model primarily uses $\langle \text{Topic}, \text{Tag} \rangle$ to represent it. *Topic* is a specific dimension vector of the lawyer's domain, while *Tag* represents the lawyer's tag vector. The weights of these two

vectors are indicated by the size of the dimensions on the vector.

(3) **Relation** Lawyer Social Relationship Vector Model

The lawyer social relationship vector model refers to the degree of connection and similarity between different individual lawyers, denoted as $Relation = \langle Similarity \rangle$. The similarity in lawyers' social relationships is expressed as:

$$\langle Similarityu_1, Similarityu_2, \dots, Similarityu_i \rangle \quad (2)$$

Where $Similarityu_i$ represents the similarity between Lawyer u and Lawyer i .

$Similarityu_i = f(S(Topic), S(Tag))$ Here, $S(Topic)$ denotes the similarity between Lawyer u and Lawyer i in terms of domain vectors, $S(Tag)$ is the similarity based on tag vectors, and $f(S(Topic), S(Tag))$ is the fusion of the $Topic$ domain vector similarity and Tag tag vector similarity of user^[1].

After extensive training, the lawyer recommendation system can profile every lawyer mentioned in the judicial documents, with all attributes of a lawyer being represented by $Lawyer$ vectors. Moreover, when new judicial documents are added, lawyer information can be updated promptly. Then, the similarity among lawyers can be reflected based on the similarity of their profile tag vectors.

3. Collaborative filtering recommendation based on lawyer user profile

3.1 Collaborative filtering and recommendation of user profiles

The core idea of collaborative filtering algorithms is to deduce user preferences by analyzing historical behavior data of users and to make recommendations based on these preferences^[4]. In this project, due to the small number of potential users and the large number of judicial documents with strong timeliness, specifically, if two users have high similarity, their preference tags also tend to be similar. This is because when users choose tags, they usually select those that align with their interests, so similar tags suggest similar preferences. Therefore, this paper uses user profile tag attributes to calculate the similarity between users. Thus, this project proposes a new method of calculating similarity, which blends traditional user rating-based similarity with tag preference-based similarity using a weight α , to better accommodate the needs of different users. This method can help users find suitable lawyers more quickly, increasing the accuracy and efficiency of recommendations.

Additionally, to address the sparsity issue in collaborative filtering, this project employs SVD technology to optimize the sparse matrix, thereby enhancing the recommendation effectiveness. Specifically, in making lawyer recommendations, analysis can be conducted based on the tag attributes of users and lawyers. If user m has similar tag attributes to a lawyer, then every user in the user set mi similar to m should have the same preferences as m . The specific process for collaborative filtering recommendation based on user profiles is as follows:

(1) If T_1, T_2, \dots, T_k represent the tags assigned to users and lawyers, and users are represented as u_1, u_2, \dots, u_m , where W_{ij} represents the weight of user i on tag j , the user-tag scoring matrix, in addition to methods such as the duration of user browsing and the number of similar documents viewed, also implicitly expresses user preferences through the weight of tags, establishing a user tag weight matrix, i.e., a user-tag matrix. Similarly, lawyers are represented as sets V_1, V_2, \dots, V_n , and the value $W_{m,n}$ is the product of the weights of user m and lawyer n on tag i , resulting in the

user-Tag scoring matrix:

Table1. User-Tag Rating Matrix

User-Tag Rating Matrix					
	T ₁	T ₂	T ₃	...	T _k
U ₁	W _{1,1}	W _{1,2}	W _{1,3}	...	W _{1,k}
U ₂	W _{2,1}	W _{2,2}	W _{2,3}	...	W _{2,k}
U ₃	W _{3,1}	W _{3,2}	W _{3,3}	...	W _{3,k}
...
U _n	W _{n,1}	W _{n,2}	W _{n,3}	...	W _{n,k}

Compared to cosine similarity, the Pearson correlation coefficient minimizes the influence of user rating bias by adjusting each independent rating using the average user rating [5]. Thus, the Pearson correlation coefficient is utilized to calculate similarity:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (3)$$

determining the similarity between users and tags Similarity₁ (u, v).

(2) The calculation of user and lawyer tag preference vectors, defining P_{ui} is the preference of user^u to tagⁱ, the calculation of user tag preference is as follows:

$$P_{ui} = \text{weight}_{ui} \times AF \times SF \quad (4)$$

And weight_{ui} is the lawyer^v's score on tagⁱ. AF, SF are weights assigned based on the user's behavior under that tag, among other factors. The formula for calculating the lawyer's tag preference is as follows:

$$P_{vi} = \text{weight}_{vi} \times AF \times SF \quad (5)$$

And weight_{vi} is the lawyer ^l 's score on tag ⁱ. AF, SF represent the capability and expertise of the lawyer ^l on the tag ⁱ.

Table2. User Lawyer Rating Matrix Based on Label Preferences

User Lawyer Rating Matrix Based on Label Preferences					
	V ₁	V ₂	V ₃	...	V _k
U ₁	0	0	P _{1,3}	...	P _{1,k}
U ₂	P _{2,1}	P _{2,2}	0	...	0
U ₃	P _{3,1}	P _{3,2}	P _{3,3}	...	0

Next, calculate the similarity of tag preference amounts, defined Similarity₂ (u, v) as the similarity between the tag preference amounts of two users. ^u and ^v respectively represent the user and the lawyer. The calculation formula is as follows:

$$\text{Similarity}_2(u, v) = \frac{\sum_{i=1}^k P_{ui} P_{vi}}{\sqrt{\sum_{i=1}^k P_{ui}^2} \sqrt{\sum_{i=1}^k P_{vi}^2}} \quad (6)$$

(3) To calculate the final similarity, after obtaining Similarity₁ (u, v) and Similarity₂ (u, v), a dynamic weight is used to measure their ratio. The formula for the final comprehensive similarity is as follows:

$$\text{Similarity}(u, v) = \alpha \times \text{Similarity}_R(u, v) + (1 - \alpha) \times \text{Similarity}_S(u, v) =$$

$$\alpha \times \frac{\sum_{i \in C} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in C} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in C} (r_{v,i} - \bar{r}_v)^2}} + (1 - \alpha) \frac{\sum_{i=1}^k P_{ui} P_{vi}}{\sqrt{\sum_{i=1}^k P_{ui}^2} \sqrt{\sum_{i=1}^k P_{vi}^2}} \quad (7)$$

u and v respectively represent the user and the lawyer. c represents the intersection of the same tags they share. \bar{r}_u and \bar{r}_v represent the average of all tag weights for the user and lawyer, respectively. Then, a α is used to balance the proportion of the two types of similarity calculations. The magnitude of α can be adjusted to optimize the recommendation performance.

(4) The overall similarity is derived from similar users and lawyers with high tag similarity, from which the target user's preferences are predicted, generating precise recommendation results.

3.2 Design of hybrid recommendation algorithm

The collaborative filtering recommendation method based on user profiles, as detailed above, utilizes SVD technology to optimize sparse matrices and addresses the sparsity problem of collaborative filtering, achieving good recommendation results. Based on the user profile tags mentioned earlier, similarities between users and between users and lawyers are calculated and evaluated comprehensively, and lawyers with higher recommendation scores, i.e., the most suitable top k lawyers, are recommended.

(1) When encountering the problem of matrix sparsity, the SVD algorithm is used to fill in the missing matrix. The method for calculating the fill values is based on existing research, and the formula is as follows:

$$P_{ui} = \bar{R}_u + U_f \times \sqrt{S_f}^T (U) \times \sqrt{S_f}^T V(f)^T \quad (8)$$

Where P_{ui} represents the user's predicted rating for the missing matrix, \bar{R}_u represents the average value of user u on rated items. SVD technology simplifies the matrix by decomposing the original matrix into the product of three matrices (U , S , V) and uses it to predict missing item ratings. U and V are orthogonal matrices, while S is a diagonal matrix, the diagonal elements of which represent the singular values of the original matrix, arranged from largest to smallest. The values on the diagonal of matrix S are arranged in descending order, and any rows or columns in the matrix with all zero values are deleted, resulting in a simplified f -dimensional diagonal matrix S_f . After the same processing, U_f , V_f , f are obtained, representing the processed dimensional parameters, and finally, a matrix M_f with complete ratings is obtained. This formula is actually the process of recombining the simplified matrices back into the original matrix [6].

(2) Based on the complete rating matrix M_f , the similarity between users $\text{Similarity}_1(u, v)$ is calculated using the traditional Pearson similarity formula. Through the user profile tag matrix, the users' tag preferences are calculated, and then the similarity between users based on user profiles $\text{Similarity}_2(u, v)$ is calculated.

(3) To calculate the final similarity, a mix of traditional Pearson correlation coefficient similarity calculations and tag preference-based similarity calculations is used [7], resulting in the final similarity $\text{Similarity}(u, v)$.

(4) Taking into account factors such as the location of the user and lawyer, and the lawyer's level of activity, suitable lawyers are recommended to the user.

4. Summary

To better serve citizens seeking legal assistance, this paper proposes a collaborative filtering recommendation technique based on lawyer profiles, aimed at enhancing the precision of lawyer recommendations, deeply matching user needs, and maximizing the protection of user rights. Compared to traditional collaborative filtering algorithms, this paper introduces lawyer profile attributes to improve the similarity calculation method, thereby better meeting users' personalized needs. In calculating similarity, the Pearson correlation coefficient formula is used, which adjusts individual ratings by using the average user rating, reducing the impact of user rating biases and enhancing the accuracy of recommendations.

Furthermore, this paper fully utilizes cutting-edge technologies, such as user profiles and recommendation algorithms, actively responding to the Party Central Committee's call for comprehensively building a "smart rule of law" and contributing to the digitalization, networking, and intelligentization of constructing a rule of law in China. By improving the accuracy of lawyer recommendations and reducing the trial-and-error costs for users, more citizens in need of legal aid can receive timely and effective help, advancing the construction of a rule-of-law society and realizing the people-centered development philosophy.

Reference

- [1] Li Anqi. Research and Implementation of a Lawyer Recommendation System Based on Judgment Documents [D]. East China Jiaotong University, Software Engineering, 2020.
- [2] Liang Nan. Research and Application of Lawyer Recommendation Methods Based on Text Mining [D]. University of Electronic Science and Technology of China, Computer Software and Theory, 2015.
- [3] Xuan Kaizhou. Research on Weibo Rumor Detection and Stance Analysis [D]. Nanjing University of Science and Technology, Pattern Recognition and Intelligent Systems, 2019.
- [4] Ding Li, Fang Xiao. Research on a Recommendation Algorithm Integrating User Interests and Review Text Topic Mining [J]. Journal of Qinghai Normal University (Natural Science Edition), 2022, 38(01): 14-23+96.
- [5] Hu Anming, Chen Huie. Improved Personalized Recommendation Model Based on Time Factor [J]. Software Engineering, 2021, 24(07): 60-62.
- [6] Nie Zhixuan. Research on the Precision Recommendation Method of Peer Review Experts Based on User Profiles [D]. Wuhan University of Science and Technology, 2021.
- [7] Zhang Rong. Research on a Teaching Capability Diagnostic Assistance System Based on Teacher Profiling [D]. Jiangxi University of Finance and Economics, 2021.