# University major information recommendation based on federated learning

**Yuzhao Song[1,4], Yexin Wu[2] and Yanjie Xue[3]**

[1]NingXia Agricultural School, Yinchuan city, 750021, China
[2]Xinjiang University, Urumqi city, 830046, China
[3]Chang'an University, Xi'an city, 710064, China

[4]2020210829@email.cufe.edu.cn

**Abstract.** This paper focuses on the research of applying federated learning to recommendation systems and proposes a university major recommendation method based on federated learning. Furthermore, an improved knowledge distillation architecture is implemented for university major recommendations. In the collaborative structure of the system based on federated learning, knowledge distillation is used to optimize the recommendation performance. The federated learning algorithm, FedDyn, is employed to aggregate model parameters through weighted averaging, enabling a training mode where only local training data and local models are uploaded to the central server. After reading and studying other papers that apply federated learning to recommendation systems, this paper conducts further speculation and research, aiming to apply relevant knowledge and techniques to establish a system that can recommend specific content to a targeted audience, such as students after the college entrance examination. This includes providing major-related information, predicting students' major preferences, and delivering the latest industry news related to specific majors.This paper also categorizes the technologies used in the creation of recommendation systems and compares them into three categories. The study suggests that the recommendation system utilizing knowledge distillation will be more efficient.

**Keywords:** Federated Learning, Knowledge Distillation, Recommendation System.

## 1. Introduction

The Gaokao (National College Entrance Examination) is a large-scale exam that students worldwide must take. After the Gaokao, many students face the challenge of choosing schools and majors based on their scores, aiming to secure a good major, attend a good university, and ultimately find good employment opportunities. In China, with a wide variety of majors available and significant differences between them, many parents find themselves unfamiliar with numerous professions, not knowing what students should study or what careers they can pursue after graduation. Consequently, an industry related to the Gaokao has emerged, where, after paying a certain fee, professionals recommend schools and majors to students based on their scores, aiming to provide better opportunities for education and employment. However, for most ordinary families may lack access to relevant professionals and sufficient financial resources. To this end, developing an automated professional recommendation method is urgent and of great significance.

On the other hand, with the popularization of computing devices and the rapid development of information technology [1], a large amount of data on the introduction of high school majors, employment prospects and even career development have been accumulated on the Internet. Obviously, analyzing these data, and mining the hidden knowledge behind them can help in the choice of majors, thus bringing economic and social benefits. With the popularity of computing devices and the implementation of applications related to daily life, the speed and volume of digital information have increased dramatically. It is foreseeable that by utilizing these data and employing intelligent computational methods, we can improve the quality of services, thereby bringing about economic and social benefits. This motivates us to create a system that recommends schools and majors to Gaokao students, which can provide intelligent major recommendations to students based on their characteristics, such as Gaokao scores, preferences, and strengths.

However, constructing such a major recommendation system may involve the privacy issues of students' personal information [2]. For example, the model attempts to record the historical search information of students, analyze their knowledge preferences, and identify their weak areas, thereby improving the accuracy of student selection. By analyzing the download volume of student accounts, the system also understands their learning needs and adjust the frequency of book recommendations. Additionally, the system can infer the user's reading preferences, whether they lean towards domestic or international books, and recommend content that aligns with their preferences.

In order to alleviate the above problems, inspired by the rapid development of federated learning in recent years, in this paper, we combine the federated learning and recommendation systems to create a system that recommends schools and majors to Gaokao students. By integrating these two core components, the research aims to provide intelligent major recommendations to students based on their characteristics, such as Gaokao scores, preferences, and strengths, while ensuring the confidentiality of user information. Furthermore, the research aims to incorporate current societal conditions to offer more personalized major recommendations to students.

However, the application of federated learning and recommendation systems faces the following challenges: (1) Addressing the issue of imbalanced data distribution in federated learning by researching differential privacy optimizers that balance accuracy and privacy. This helps mitigate the varying degrees of accuracy reduction caused by differential privacy for different data owners in federated learning. (2) Investigating personalized trust levels in federated learning to support data owners in using differential privacy budgets as needed. This research aims to solve the combination of centralized differential privacy and localized differential privacy. (3) Considering the high computational cost borne by data owners in federated recommendation systems, studying differential privacy protocols that balance the accuracy of local recommendation results for data owners and protect their intentions. This research focuses on specific recommendation models in federated recommendation systems. (4) Addressing the issue of data owners actively deleting data and adjusting models to accommodate the slow response to data deletion requests. Researching federated recommendation problems that satisfy the definition of differential privacy and enable data owners to exercise their right to be forgotten quickly.

These challenges and key issues need to be addressed to effectively combine federated learning and recommendation systems in the proposed system and ensure accurate and privacy-preserving recommendations for Gaokao students. Focusing on the above aspects, in this paper, we propose a major recommendation method based on federated learning. Specifically, we will introduce the basic definitions and research status in Section 2, and give the design details of our recommendation system in Section 3. In Section 4, we will report the result and discuss the limitations of our proposed system.

## 2. Basic Definitions and Research Status

### *2.1. Revisiting the Federated Learning*

Federated learning is a computational framework first introduced by McMahan et al. from Google in 2017. It enables the training of machine learning models without the need for data owners to share their raw data or rely on a central server for distributed training. In federated learning, the training data is distributed across the devices of data owners [3]. Each data owner performs local training and only shares the model parameters with the server. The server learns a global model by aggregating the updates from multiple data owners using model averaging algorithms. The literature has developed secure federated learning frameworks and provided comprehensive reviews on the definitions, architectures, and applications of horizontal federated learning, vertical federated learning, and federated transfer learning. Therefore, federated learning embodies the principle of data minimization.

## *2.2. Overview of Recommendation Systems*

*2.2.1. Definition of Recommendation Systems.* Considering the collaborative filtering-based recommendation problem, the recommendation service [4-5] provider attempts to estimate the low-rank factors of the given local user-item interaction matrix $R \in R^{m\times n}$ for the user (i.e., the data owner), considering the presence of unobserved missing values. This process is known as matrix completion. One approach to matrix completion is to model the user-item interactions in a joint latent space using the local embeddings $U_i \in R^{1\times d}$ for each user and the item embedding $V \in R^{n\times d}$. The estimate for each user-item element $y_{ij}$ is given by $y_{ij} = u_i^T v_j$, where $u_i$ and $v_j$ correspond to the i-th row of $U$ and the j-th column of $V$, respectively. The local embeddings ui are stored in the corresponding local devices of the i-th user, while V is stored in the recommendation server. By using a fixed hyperparameter matrix $W \in R^{m\times n}$, different confidence levels are assigned to the sum terms $y_{ij} - u_i^T v_j$ 2 for each user-item element. Therefore, the objective of this method is to minimize the following loss function:

$$L(U,V) = \frac{1}{2}\sum\sum wij(yij - u_i^T v_j)^2 \tag{1}$$

where ‖·‖ F represents the Frobenius norm. Typically, $wij = 1$ is used to denote observed elements. There are several common strategies for assigning weights to missing data. One straightforward approach is to assign a uniform weight $\mathrm{w}_0 \in [0,1]$ to all missing elements. However, in real-world recommendation system scenarios, if an active user (or a popular item) has no interactions, it is more likely to be considered as negative towards other items (or users). Inspired by the above observation, non-uniform weighting is proposed, for example, by using user activity or item popularity as frequency measures to determine Wij. The goal of federated recommendation systems is to collaboratively train recommendation models in a federated manner, where the recommendation service provider does not directly access users' private data.

*2.2.2. Current Research Status of Recommendation Systems.* In 1997, researchers Resnick and Varian [6] first proposed a descriptive definition of traditional recommendation systems as follows: "They rely on certain e-commerce websites to provide customers with the product information they need, help them make purchase decisions, and simulate the process of salespeople assisting customers in completing purchases in real-time. "From the perspective of disciplinary origins, social recommendation systems are primarily based on the analysis theories associated with social networks. These theories tightly integrate a series of social attribute information that users are involved in with traditional recommendation systems. Social recommendation systems not only effectively address the problems of data sparsity and user cold start but also improve the performance of existing systems. In social network theory, one key tool is social network analysis, which explains the intrinsic relationships and evaluation values among individuals, groups, organizations, computers, and other entities. Social network analysis particularly focuses on the inherent relationships among users, while users and their corresponding information attributes are positioned as subordinates. When numerous individuals interact with each other within a virtual network, they form a specific topology network, which researchers generally define as an online social network. Some researchers point out that this network is part of the broader category of traditional social networks.

In recent years, research on recommendation systems [7-10] has witnessed significant advancements due to the availability of large-scale datasets, advancements in machine learning techniques, and the emergence of new paradigms such as federated learning. Researchers are exploring various approaches to improve the accuracy, diversity, and explainability of recommendations. Additionally, there is a growing emphasis on addressing ethical concerns, privacy protection, and fairness in recommendation systems. The field of recommendation systems continues to evolve, driven by the ever-increasing demand for personalized and relevant recommendations in various domains.

## 3. System Design

### *3.1. Architecture of a Recommendation System using Federated Learning*

*3.1.1. Systems utilizing knowledge distillation.* In the work by Li [11] et al., knowledge distillation is applied to federated learning. They assume a federated system with $\boldsymbol{n}$ clients, where each client $\boldsymbol{i}$ holds $m_i$ private records. The local privacy dataset $D_i = [(x_i^1, y_i^1), ..., (x^i_{mi}, y^i_{mi})]$ is derived from the Cartesian product domain $X \times Y$, $\boldsymbol{y}$ is represented as a vector. The federated server possesses mpub unlabeled public data $D_{pub} = [(x_1, ?), ..., (x_{mpub}, ?)]$. The objective of the federated knowledge distillation algorithm, dependent on the public data, is to label the public dataset $D_{pub}$ using the knowledge from the distributed privacy dataset $D_{priv} = D^1 \cup D^2 \cup ... \cup D^n$ . The student model (federated model) is trained on the labeled public dataset to serve the participants, while ensuring the privacy of the client's local sensitive data during the knowledge transfer process.

To address the limitations of existing federated knowledge distillation algorithms that rely on public data, the paper proposes the Reverse k-Nearest Neighbor Voting-based Federated Knowledge Distillation Algorithm (RKNN). The simplified process of this algorithm is illustrated in Figure 1. The RKNN algorithm utilizes a public representation model to extract data features and uses the cluster centers of the public data as query samples. In the feature space, a single private data point votes for k query samples. This approach not only avoids training the teacher model locally on the client side but also saves the privacy budget required for differential privacy protection. The algorithm described in the paper mainly follows the illustrated process below.

In this study, after applying knowledge distillation techniques to an information recommendation system that incorporates federated learning, certain speculative extensions to the algorithm are made. After implementing knowledge distillation in the creation of the recommendation system, it was assumed that there are four rows named A, B, C, D, corresponding to columns $A_1$, $B_1$, $A_2$, $B_2$, respectively. The resulting numerical table showed column sums of 38, 10, 42, and 11. In the Table 1, weights were assigned to variables 'a' and 'b', with ωa = 0.6 and ωb = 1.2, respectively. Multiple training iterations were performed, and the generated training data was processed by the main server to gradually optimize the model. This led to the convergence of $a_1$ towards 0 and $a_2$towards 1, gradually fitting the model to the desired ideal effect. The workflow at this stage can be visually represented as follows.

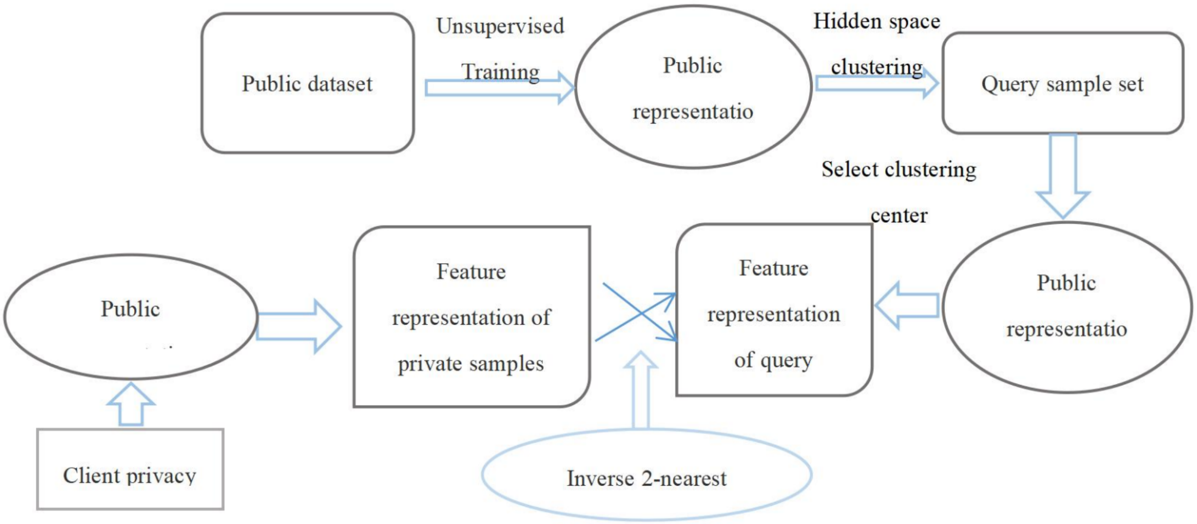

**Figure 1.** The pipeline of the Reverse k-Nearest Neighbor Voting-based Federated Knowledge Distillation Algorithm.

**Table 1.** The training iterations of the main server.

| | $Aa_1$ | $Bb_1$ | $Ca_2$ | $Db_2$ |
|---|---|---|---|---|
| A | 13 | 2 | 6 | 1 |
| B | 6 | 3 | 8 | 6 |
| C | 9 | 1 | 12 | 4 |
| D | 10 | 4 | 16 | 0 |
| | $Y_1$= 38 | $Y_2$= 10 | $Y_3$= 42 | $Y_4$= 11 |

Overall, the system model can be described as follows: the main server connects multiple individual user accounts, each with varying levels and heterogeneous data. The model algorithm is sent to the user devices for training, and the training parameters are sent back to the main server to form a preliminary system model.

The benefits of using this technology for students include the protection of student privacy during model training. Privacy information is not sent to the outside world, eliminating potential risks. FedAvg, as a distributed framework, allows multiple users to train a machine learning model simultaneously. Each user identifies similarities in their data and derives predictive results, forming a preliminary model.

*3.1.2. Recommendation System based on Federated Transfer Learning.* Xun et al. [13] proposed a recommendation system based on federated transfer learning.Registration can be done by users who provide a username, email, and password,granting them access to the system. Within this design, the personal information page includes additional tab pages, primarily the "Profile" page. The Profile page stores information such as bookmarked records, rating history, profile editing, and account security. On the profile editing page, users can select their preferred genres. Additionally, users can set up an email account to receive weekly recommended updates. The movie recommendation page serves as the main page of the system, primarily focusing on recommending 10 movies that users may like based on their historical rating records and bookmarks. On this recommendation page, users can rate the recommended movies on a scale of 1 to 10 and bookmark movies that they haven't previously bookmarked.

*3.1.3. Traditional Recommendation Algorithms.* It is well-known that general recommendation systems [14-15] involve various disciplines such as data mining, machine learning, and predictive algorithms, forming a new research field. Recommendation systems are used to predict user preferences for items. Typically, a recommendation system collects and organizes a user's historical purchase history, search records, or similar preferences to discover items that align with a user's preferences and recommend them. The core technology of a recommendation system lies in the recommendation algorithm. Nowadays, widely used recommendation systems can be roughly categorized into four types: content-based filtering, collaborative filtering based on user preferences, collaborative filtering based on item similarities, and hybrid recommendations. The semi-supervised knowledge distillation technique and self-supervised knowledge distillation technique used in the previous context can improve accuracy

In this study, the research process of the university recommendation system attempted to utilize knowledge distillation techniques and Dyn technology. It started with semi-supervised knowledge distillation, followed by self-supervised knowledge distillation, and finally incorporated the FedDyn algorithm to iteratively optimize the final model, gradually approaching the desired functionality of a university recommendation system. The knowledge distillation technique mentioned earlier was initially developed for model compression purposes. As shown in Figure 2, it involves transferring the content knowledge from a complex and large teacher model to a smaller and simpler student model, thereby enhancing the generalization performance of the student model. The student model, being simplified, is easier to deploy on target devices and can achieve comparable performance to the teacher model. During the distillation process, a specific distillation temperature T is set to soften the data. The softened data contains similar information, providing the student model with specific latent knowledge that facilitates its training.

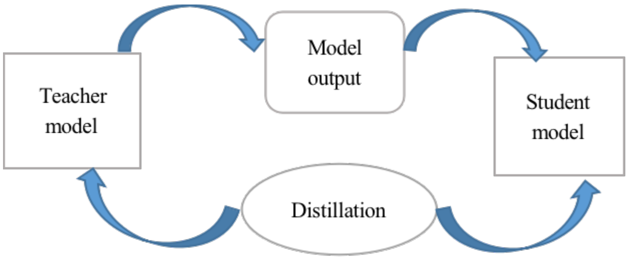

**Figure 2.** The pipeline of Teacher-student model for knowledge distillation.

## *3.2. University System Recommendations and Student Demand Prediction*

The objective of the system is to filter the university information and provide better content recommendations for teachers and students, such as recommending majors and colleges to students and recommending corresponding students to teachers. After collecting this information, the system analyzes various types of information, such as the preferences of students from all departments, the entire university, or even the entire country. By building models, the university information recommendation system can be optimized.

The process is as follows: (1) Semi-supervised Knowledge Distillation. In this step, only a portion of the data is used, such as selecting data from a specific class or department. The data is then subjected to semi-supervised knowledge distillation techniques, which can improve the accuracy of the model in the initial stage. (2) Self-supervised Knowledge Distillation. At this stage, the amount of collected data increases, expanding to the entire university or even the entire country. If using semi-supervised distillation techniques leads to low efficiency, self-supervised knowledge distillation should be employed instead. (3) FedDyn Optimization of the Overall Model. By utilizing FedDyn technology, the overall model's accuracy is improved to a certain extent, creating a more user-friendly system.

The specific functionalities of the system vary based on user roles [15] . For users with a teacher role, the system can recommend relevant students for teachers to facilitate the search for suitable graduate or doctoral students, as well as academic collaborators. For users with a student role, the system adjusts the recommended content to include information about majors, career prospects, and professional directions.

## 4. System Analysis and Discussion

### *4.1. Comparison between Students' Independent Major Selection and the Use of Recommendation Systems*

When students independently choose their majors, they often face confusion and uncertainty when confronted with various options. Some students with sufficient resources may seek assistance from educational institutions, but this often comes with a high cost. Other students may rely on independent research or seek advice from experienced parents. These processes can be time-consuming, costly, and often repetitive for many individuals. By using a recommendation system, this repetitive and similar work can be delegated to an algorithmic program, reducing time and cost while improving the accuracy of the recommendations.

On the other hand, incorporating federated learning into the recommendation system also enhances the protection of users' privacy data. Compared to conventional recommendation systems, a recommendation system that employs federated learning does not require the transmission of users' personal information. Instead, it only transfers the data generated from user model training to the host for processing. This approach effectively safeguards personal privacy while enabling communication among user data.

### *4.2. Analysis of System Limitations*

This report focuses on the feasibility of applying federated learning and recommendation systems to university information. The advantages of this work may be: (1) The accuracy of recommendation results is improved, and the overall analysis speed of the data is increased. (2) The model provides stronger privacy protection capabilities for the data, enabling secure data sharing and communication while ensuring the privacy of individual users.

Compared to other forms of federated learning [16-17], the data analysis speed is slightly lower, and the implementation process can be more complex. The framework presented in this report still has many areas that require further research and improvement.

## 5. Conclusion

To alleviate the major's choosing difficulty due to information asymmetry in the process of choosing majors in college entrance exams, this paper proposes an automatic method of recommending majors based on a recommender system and federated learning. Specifically, our university major recommendation implements an improved knowledge distillation architecture, which can further optimize the recommendation performance. In addition, after reading and studying other papers that apply federated learning to recommender systems, this paper conducts further speculations and research with the aim of applying relevant knowledge and techniques to build a system that can recommend specific content to a target audience, such as post-high school students. This includes providing information related to majors, predicting students' major preferences, and providing up-to-date industry news related to specific majors. This paper categorizes the techniques used to create recommender systems and compares them in three categories. The study shows that recommender systems that utilize knowledge distillation techniques will be more efficient.

## Authors Contribution

All the authors contributed equally and their names were listed in alphabetical order.

## References

[1] Chong Xun. Research and Implementation of Recommendation System Based on Federated Migration Learning[D]. Zhongnan University of Economics and Law, 2022. DOI:10.27660/d.cnki.gzczu.2022.002091.

[2] Zhefu W, Song Z, Agyemang P, et al. Style-aware adversarial pairwise ranking for image recommendation systems[J]. International Journal of Multimedia Information Retrieval,2023,12(2).

[3] Rodrigues S M, Fidalgo F, Oliveira Â. RecipeIS—Recipe Recommendation System Based on Recognition of Food Ingredients[J]. Applied Sciences,2023,13(13).

[4] Wongvilaisakul W,Netinant P, Rukhiran M. Dynamic Multi-Criteria Decision Making of Graduate Admission Recommender System: AHP and Fuzzy AHP Approaches[J]. Sustainability,2023,15(12).

[5] JiHyeok P, JaeDong L. A Customized Deep Sleep Recommender System Using Hybrid Deep Learning.[J]. Sensors (Basel, Switzerland),2023,23(15).

[6] Zoujing Y, Pengyu S, Chunhui Z. Finding trustworthy neighbors: Graph aided federated learning for few-shot industrial fault diagnosis with data heterogeneity[J]. Journal of Process Control,2023,129.

[7] Bo W, Hongtao L,Yina G, et al. PPFLHE: A privacy-preserving federated learning scheme with homomorphic encryption for healthcare data[J]. Applied Soft Computing,2023,146.

[8] Zhiguo Q, Yang L, Bo L, et al. DTQFL: A Digital Twin-Assisted Quantum Federated Learning Algorithm for Intelligent Diagnosis in 5G Mobile Network.[J]. IEEE journal of biomedical and health informatics,2023,PP.

[9] Ye T, Can W, Wee-Chung A L. Dynamic weighted ensemble learning for sequential recommendation systems: The AIRE model[J]. Future Generation Computer Systems,2023,149.

[10] Imran A, Misbah A, Abdellah C, et al. A heterogeneous network embedded medicine recommendation system based on LSTM[J]. Future Generation Computer Systems,2023,149.

[11] Abinash P, Singh D S. Modeling users' preference changes in recommender systems via time-dependent Markov random fields[J]. Expert Systems With Applications,2023,234.

[12] Lina S, Zongpeng L. Incentive-driven long-term optimization for hierarchical federated learning[J]. Computer Networks,2023,234.

[13] Xi C, Hui W, Siliang L, et al. Remaining useful life prediction of turbofan engine using global health degradation representation in federated learning[J]. Reliability Engineering and System Safety,2023,239.

[14] Tianjun W, W.S. T C. FGCR: Fused graph context-aware recommender system[J]. Knowledge-Based Systems,2023,277.

[15] Tingting L, Cheng Y,Cheng L, et al. Efficient one-off clustering for personalized federated learning[J]. Knowledge-Based Systems,2023,277.

[16] Linfeng H, Wei Z,Fengji L, et al. Enhanced contrastive learning with multi-aspect information for recommender systems[J]. Knowledge-Based Systems,2023,277.

[17] Gang S, Zhiqiang F, Yumin G, et al. Efficient and privacy-preserving online diagnosis scheme based on federated learning in e-healthcare system[J]. Information Sciences,2023,647.