Preprints.org

Article

Job Recommendation System Combining Collaborative Filtering and Content Based Filtering

Raden Ibnu Huygenz Widodo [*] , Reza Fahlevi Herdiyanto , Muhammad Thoriq , Imam Sayyid Rahman

Article

Job Recommendation System Combining Collaborative Filtering and Content Based Filtering

Raden Ibnu Huygenz Widodo *, Reza Fahlevi Herdiyanto, Muhammad Thoriq and Imam Sayyid Rahman

Informatics Departement, UIN Sunan Gunung Djati Bandung, Jawa Barat, Indonesia; rfahlevih@gmail.com (R.F.H.); muhammadthoriq11@gmail.com (M.T.); imamsayyid332@gmail.com (I.S.R.)

*   Correspondence: ibnuwdodo@gmail.com

Abstract: The increasing complexity and volume of job listings in the digital era necessitate advanced recommendation systems to match job seekers with suitable opportunities. This paper presents a hybrid job recommendation system that integrates Collaborative Filtering (CF) and Content-Based Filtering (CBF) to enhance recommendation accuracy and user satisfaction. The CF component leverages historical user interactions and similarities between users to suggest jobs, while the CBF component analyzes job descriptions and user profiles to provide personalized recommendations based on individual preferences and qualifications. By combining these two approaches, the proposed system mitigates the limitations of each method and offers a more comprehensive and effective solution. Experimental results on a real-world dataset demonstrate the system's improved performance in terms of precision, recall, and F1-score compared to traditional recommendation models. This study contributes to the field of job recommendation systems by providing a robust framework that effectively bridges the gap between user needs and job market opportunities.

Keywords: machine learning; CBF; CF; job recommendation system

---

I. Introduction

In today's highly sophisticated world, recommendation systems have become an important component of many things, including job search platforms. Recommendation systems can help users find job opportunities that match their skills and interests. However, developing an accurate and effective job recommendation system that meets the needs of users is a significant challenge.

One of the main problems often encountered in job recommendation systems is how to manage huge and diverse data, and how to generate relevant recommendation systems for each user. Recommender systems generally use one of two techniques, namely Collaborative Filtering or Content-Based Filtering. While both techniques have their advantages, they also have limitations.

In this research, we will combine two main techniques, namely collaborative filtering and content-based filtering. This approach will utilize the advantages of each technique while overcoming their limitations. By combining the two methods, it is expected to improve the accuracy of job recommendations. As a result, a more effective and relevant recommendation system can be achieved.

The method used in this research is a hybrid approach, which combines Collaborative Filtering and Content-Based Filtering. Collaborative Filtering is a method that analyzes user preferences by finding patterns from other similar user behaviors, while Content-Based Filtering is a method that recommends jobs based on similarities with jobs that the user likes or has accessed before.

Collaborative filtering has the advantage of providing recommendations by leveraging the preferences and behaviors of other similar users, thus being able to recognize patterns from collective user data and provide relevant recommendations without requiring detailed content. Content-based filtering offers more personalized and accurate recommendations based on the specific features of items of interest to the user, by analyzing the characteristics and attributes of items that the user has previously liked. By combining the strengths of both, the recommendation system can generate suggestions that are more relevant and in line with the user's individual preferences.

The contributions of this research include the development of a more accurate and relevant job recommendation system through combining the advantages of Collaborative Filtering and Content-Based Filtering. With this approach, the system can provide more personalized job recommendations that match the user's skills and interests, which in turn is expected to increase user satisfaction.

In addition, this research provides insight into the application of hybrid approaches in various application domains, particularly in job search. This approach can help companies and job seekers achieve their goals more efficiently, making the job search and placement process more effective and targeted.

II. Related Work

Likang Wu et al. used data from a Chinese online recruitment platform to address the difficulty of making appropriate job recommendations in sparsely interactive situations. To improve the understanding of model path information and reduce bias, they suggested a large language model (LLM) recommendation system equipped with a path randomization mechanism, a path soft selector, and a hybrid mechanism. In addition, LLM uses fast constructors to help understand the semantics of behavior graphs. Their solution, tested on real-world datasets RecrX and RecrY, improves recommendation quality by considering input meta-path position factors and the impact of various path requests on decision making, which improves job satisfaction and preference prediction [1].

According to Zhi Zheng et al. traditional job recommendation models have several shortcomings. One of them is poor explanation and the inability to create job descriptions tailored to job seekers. They proposed an InstructGPT-inspired three-step training method of supervised refinement, reward modeling based on recruiter feedback, and reinforcement learning based on proximal policy optimization. Their solution, GIRL (Generative Job Recommendation based on Large Language Models), combines supervised refinement and reinforcement learning to generate high-quality, standardized JDs that match the specific needs of job seekers. With data from a leading recruitment platform in China, this method shows better performance for GIRL compared to GIRL-SFT and other methods. This study emphasizes the importance of tailoring instructions to domain-specific data and the benefits of reinforcement learning in tailoring content to human preferences [2].

Yingpeng Du et al. researched job recommendation in online recruitment platforms by studying matching functions based on interaction records and documents. They proposed an interactive resume completion method to improve resume quality, a GAN-based approach to refine the LLM representation of low-quality resumes, and a multi-objective learning framework for job recommendation. These solutions leverage LLM through Simple Resume Completion (SRC) and Interactive Resume Completion (IRC), and use GAN to align low-quality resumes with high-quality ones. Experiments on three real recruitment datasets (designer, sales, and technology) demonstrate the effectiveness of this method, improving the quality of resume representations and leveraging interaction records and textual content[3].

Christos Troussas et al. focus on improving the performance and quality of recommendation systems, particularly in digital libraries, by addressing issues such as cold start and data sparsity problems. They propose a hybrid recommendation system that combines content-based filtering and collaborative filtering, using flexible algorithms to combine diverse recommendation sources. Their approach ensures a balanced and high-quality set of recommendations by considering the availability and quality of Collaborative Filtering (CF) and Author & Category recommendations. This hybrid system demonstrated superior performance, with 85% of students in Group 1, who received personalized recommendations, rating the user experience positively, compared to 55% in Group 2. In addition, 86% of Group 1 students rated the effectiveness of the system as very high, compared to 49% in Group 2. The research dataset involved 90 students aged 22-35, who were divided into two groups for a three-month evaluation phase [4].

Mahesh Thyluru R et al. address the problem of data sparsity in social networks that hinders the effectiveness of friend recommendations. They propose a hybrid recommendation approach that

combines collaborative, semantic, and social filtering (SocF) techniques and incorporates additional K-means and K-NN algorithms to improve the recommendation process. Their system integrates social classification and collaborative approaches to suggest the most suitable potential friends based on user profiles, aiming to mitigate the cold start problem by utilizing semantic and social information. Experimental results on the Yelp social network dataset show that their hybrid recommendation system (SemSocCoF algorithm) achieves superior accuracy (96.49%) compared to other methodologies, maintaining over 90% accuracy even for unbalanced users. The combination of semantic and social data with the CoF algorithm significantly improves recommendation accuracy compared to the user-based CoF algorithm [5].

Luong Vuong N et al. examined improving the accuracy and effectiveness of recommendation systems, particularly within the scope of collaborative filtering methods. They compared various collaborative filtering techniques, including KNN-based and model-based approaches such as KNN-Basic, KNN-w-Baseline, KNN-w-Means, Co-Clustering, and ExtKNNCF, using evaluation matrices such as MAE, RMSE, MAP, and NDCG. Their proposed solution, ExtKNNCF, a model-based collaborative filtering method that uses singular value decomposition (SVD), showed superior performance. In experiments using MovieLens-100K and MovieLens-1M datasets, ExtKNNCF achieved the lowest MAE and RMSE values, indicating higher accuracy in ranking prediction, and the highest MAP and NDCG values, indicating better relevance and ranking of the recommended items. ExtKNNCF outperforms the second highest method, co-clustering, by a significant margin across all metrics, highlighting its effectiveness in improving recommendation quality [6].

III. Methodology

Data Collection The first step in our methodology was data collection. We sourced our dataset from a well-known job search platform, which provided comprehensive information on user profiles, job listings, and interaction data. User profiles included demographics, skills, job history, and preferences. Job listings encompassed job titles, descriptions, requirements, and company information. Interaction data recorded activities such as job applications, views, and user ratings.

Preprocessing Next, we conducted extensive data preprocessing to ensure the dataset's quality and consistency. This involved: Data Cleaning: We removed duplicate entries, addressed missing values, and normalized text fields to create a uniform dataset.

Collaborative Filtering (CF) Component For the CF component, we focused on utilizing historical user interactions to recommend jobs: User-Item Matrix Construction: We created a matrix where rows represented users and columns represented jobs, with matrix entries indicating the level of user interaction with each job (e.g., ratings or application status). Similarity Computation: We calculated similarities between users or jobs using methods such as cosine similarity or Pearson correlation. Recommendation Generation: By employing techniques like Matrix Factorization (e.g., Singular Value Decomposition) or Neighborhood-based approaches, we predicted user preferences for jobs they had not yet interacted with .

Content-Based Filtering (CBF) Component Simultaneously, the CBF component analyzed job descriptions and user profiles to offer personalized recommendations: Profile Construction: We built detailed user profiles by aggregating information from resumes, skills, and past interactions. Job Feature Analysis: Key features were extracted from job descriptions, focusing on required skills, job roles, and industries. Similarity Matching: Using cosine similarity or other distance metrics, we matched user profiles with job features to recommend jobs aligning with user preferences . Hybrid Recommendation System Our hybrid recommendation system amalgamates the outputs from the CF and CBF components: Weighting Scheme: We assigned weights to the CF and CBF components based on their performance metrics (e.g., precision, recall), balancing their contributions to the final recommendation. Aggregation Method: We merged recommendations from both components using techniques such as weighted averaging or hybrid models (e.g., Hybrid Matrix Factorization). Personal-

ization Layer: Incorporating user feedback and real-time interaction data, we continuously refined and personalized recommendations to enhance user satisfaction .

Evaluation We rigorously evaluated our system using a real-world dataset: Evaluation Metrics: Metrics like Precision, Recall, F1-Score, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) were employed to assess recommendation accuracy and effectiveness. Baseline Comparison: We compared the hybrid system's performance against traditional CF and CBF models, demonstrating its superiority. User Study: A user study provided qualitative feedback on the system's usability and satisfaction levels, offering insights for further refinement . Implementation The system was implemented using a robust set of tools and technologies: Programming Languages: Python was chosen for algorithm implementation and data processing. Libraries and Frameworks: We utilized Scikit-learn for machine learning models, Pandas for data manipulation. Database Management: MySQL or MongoDB was used for managing large volumes of user and job data. Deployment: The system was deployed on a cloud-based platform, ensuring scalability and accessibility for users . Continuous Improvement To maintain the system's effectiveness and relevance, we established a continuous improvement process: Feedback Loop: A feedback loop was implemented to incorporate user interactions and feedback into the recommendation process. Model Retraining: We periodically retrained models with updated data to capture evolving user preferences and job market trends. Performance Monitoring: Continuous monitoring of system performance allowed us to make necessary adjustments to improve accuracy and user satisfaction .

IV. Result and Discussion

Data collected from online platform Glassdoor and Jobstreet. The primary data collected focuses on job title, location, and salary. The data collection process involved extracting job listings from these platforms over a period of several months, ensuring a comprehensive and diverse dataset. Each entry includes detailed information about the job title, geographical location, and offered salary, providing a rich basis for analysis and recommendation.

To ensure data accuracy and relevance, we implemented a rigorous preprocessing stage. This involved cleaning the data to remove duplicates, outliers, and inconsistencies. We standardized the format of job titles and locations, enabling more effective comparison and analysis. Additionally, we handled missing values through imputation techniques, maintaining the integrity of the dataset while minimizing potential biases.

By focusing on job title, location, and salary, we aimed to create a dataset that accurately reflects the job market and provides valuable insights for our recommendation system. This foundational data is crucial for developing effective algorithms that cater to the diverse needs of job seekers.

To ensure the data's usability for our recommendation system, we underwent a thorough preprocessing stage. This process began with cleaning the data to eliminate duplicates, outliers, and inconsistencies. By removing redundant and erroneous entries, we aimed to create a clean and reliable dataset.

Next, we standardized the job titles and locations to facilitate effective comparison and analysis. This standardization involved converting various formats and terminologies into a uniform structure, allowing our algorithms to accurately interpret and utilize the data. This step was crucial in ensuring that the data from different sources could be integrated seamlessly.

Additionally, we addressed missing values through imputation techniques. By carefully estimating and filling in these gaps, we preserved the dataset's completeness and integrity. This approach helped minimize potential biases and ensured that our recommendation system had access to a robust and comprehensive set of data.

In this study, we developed a hybrid job recommendation system that effectively combines Collaborative Filtering (CF) and Content-Based Filtering (CBF) techniques. Our approach leverages the strengths of both methods to overcome their individual limitations, thereby providing more accurate and personalized job recommendations. By harnessing the power of CF and CBF, we can address

the shortcomings of each method and create a more robust recommendation system. CF, which relies on historical user interactions and similarities between users, excels in identifying patterns and preferences based on collective user behavior. However, it struggles with the cold start problem and can sometimes lack personalization. CBF, on the other hand, focuses on analyzing job descriptions and user profiles to offer tailored recommendations, ensuring that the suggestions are relevant to individual users. By integrating these two techniques, our system can deliver more comprehensive and effective job recommendations.

The CF component of our hybrid system uses historical user interactions and similarities between users to suggest relevant job opportunities. This method helps identify patterns and preferences based on user behavior, offering recommendations that align with collective user interests. For example, if a user has shown interest in software engineering jobs in the past, the CF component will suggest similar job opportunities based on the behavior of other users with similar interests. This approach leverages the wisdom of the crowd to provide recommendations that are likely to be relevant and appealing. Additionally, by analyzing the interactions and preferences of a large user base, CF can uncover hidden patterns and trends that might not be immediately apparent.

The CBF component analyzes job descriptions and user profiles to offer tailored recommendations based on individual preferences and qualifications. By examining the content of job listings and matching them with user profiles, our system can generate job suggestions that are highly relevant to the user's skills and interests. For instance, if a user has a background in data science and prefers remote work, the CBF component will prioritize job listings that match these criteria. This method ensures that the recommendations are not only popular but also personalized, addressing the specific needs and preferences of each user. By combining the strengths of CF and CBF, our hybrid system can provide a more balanced and effective recommendation experience.

By integrating these two methods, our system can generate comprehensive and effective job suggestions that better align with users' needs and preferences. This dual approach ensures that users receive recommendations that are not only popular but also highly relevant to their unique skills and interests. The integration of CF and CBF allows our system to benefit from the collective intelligence of user interactions while also providing personalized recommendations based on individual profiles. This synergy creates a powerful recommendation engine that can adapt to different users and job market conditions, offering a superior user experience.

Our experimental results on a real-world dataset demonstrated the superior performance of the hybrid system compared to traditional CF and CBF models. Metrics such as precision, recall, and F1-score showed significant improvements, indicating that the hybrid approach successfully enhances recommendation accuracy and user satisfaction. In our experiments, the hybrid model consistently outperformed standalone CF and CBF models, demonstrating its ability to provide more accurate and reliable recommendations. The improvements in precision, recall, and F1-score suggest that the hybrid system can effectively balance the trade-offs between popularity and personalization, resulting in a more effective recommendation system.

The hybrid model's ability to leverage both user behavior and content features results in more holistic and precise recommendations. By combining the strengths of CF and CBF, our system can capture a wider range of factors that influence job preferences and suitability. This comprehensive approach ensures that our recommendations are well-rounded and account for various aspects of user behavior and job content. For instance, the hybrid model can recommend jobs that are not only similar to those that users have previously shown interest in but also align with their skills and qualifications. This multi-faceted approach leads to more accurate and meaningful recommendations, enhancing the overall user experience.

The user study further validated the effectiveness and usability of our system, highlighting its potential to increase user engagement and satisfaction in job search platforms. Participants in the study reported higher satisfaction levels and found the recommendations more relevant and useful compared to previous systems. The user study provided valuable insights into how our hybrid system

performs in real-world scenarios, demonstrating its ability to meet the needs and expectations of users. By collecting feedback from users, we were able to refine our system and ensure that it delivers a high-quality recommendation experience. The positive feedback from participants underscores the potential of our hybrid approach to transform job search platforms and improve user engagement.

The continuous improvement process, including a feedback loop and regular model retraining, ensures that the system remains relevant and effective in a dynamic job market. This iterative approach allows the model to adapt to changing user preferences and job market trends. By incorporating user feedback and retraining the model on new data, we can keep the recommendation system up-to-date and responsive to the evolving needs of job seekers. This ongoing refinement process is crucial for maintaining the accuracy and relevance of recommendations, ensuring that our system continues to provide value in a rapidly changing job market. Through continuous improvement, we aim to create a recommendation system that remains effective and user-centric over the long term.

V. Conclusion

In today's increasingly complex and sophisticated world, recommendation systems have become an essential component of various platforms, including job search platforms. This research combines two main techniques, Collaborative Filtering (CF) and Content-Based Filtering (CBF), to address the challenges in developing accurate and effective job recommendation systems. The data used was sourced from online platforms Glassdoor and Jobstreet, focusing on job titles, locations, and salaries. The data collection process, which spanned several months, ensured a comprehensive and diverse dataset. Rigorous preprocessing stages, including data cleaning, format standardization, and imputation of missing values, were implemented to maintain data accuracy and relevance.

The experimental results demonstrated that the hybrid system outperformed traditional CF and CBF models, with significant improvements in metrics such as precision, recall, and F1-score. User studies also indicated that the system was more effective and satisfying, with higher levels of user engagement and satisfaction. Continuous improvement processes, including feedback loops and regular model retraining, ensure that the system remains relevant and effective in a dynamic job market. This iterative approach allows the model to adapt to changing user preferences and evolving job market trends, providing a better and more tailored recommendation experience. Overall, this research makes a significant contribution to the development of more accurate and relevant job recommendation systems and offers insights into the application of hybrid approaches in various domains, particularly in job search.

References

1. L. Wu, Z. Qiu, Z. Zheng, H. Zhu, and E. Chen, "Exploring large language model for graph data understanding in online job recommendations," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 8, 2024, pp. 9178–9186.

2. Z. Zheng, Z. Qiu, X. Hu, L. Wu, H. Zhu, and H. Xiong, "Generative job recommendations with large language model," 2023. [Online]. Available: https://arxiv.org/abs/2307.02157

3. Y. Du, D. Luo, R. Yan, X. Wang, H. Liu, H. Zhu, Y. Song, and J. Zhang, "Enhancing job recommendation through llm-based generative adversarial networks," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 38, no. 8, p. 8363–8371, Mar. 2024. [Online]. Available: http://dx.doi.org/10.1609/aaai.v38i8.28678

4. C. Troussas, A. Krouska, A. Koliarakis, and C. Sgouropoulou, "Harnessing the power of user-centric artificial intelligence: Customized recommendations and personalization in hybrid recommender systems," *Computers*, vol. 12, no. 5, p. 109, May 2023. [Online]. Available: http://dx.doi.org/10.3390/computers12050109

5. M. T. Ramakrishna, V. K. Venkatesan, R. Bhardwaj, S. Bhatia, M. K. I. Rahmani, S. A. Lashari, and A. M. Alabdali, "Hcof: Hybrid collaborative filtering using social and semantic suggestions for friend recommendation,"

Electronics, vol. 12, no. 6, p. 1365, Mar. 2023. [Online]. Available: http://dx.doi.org/10.3390/electronics12061365

6. L. V. Nguyen, Q.-T. Vo, and T.-H. Nguyen, "Adaptive knn-based extended collaborative filtering recommendation services," *Big Data and Cognitive Computing*, vol. 7, no. 2, p. 106, May 2023. [Online]. Available: http://dx.doi.org/10.3390/bdcc7020106