Universidad Politécnica de Madrid

Escuela Técnica Superior de Ingenieros Informáticos

Information retrieval, extraction and integration

Assignment 1: Comparison and discussion of two Profile-based Retrieval Systems

Authors:

José Antonio Ruiz Heredia

Joseph Tartivel

Álvaro Honrubia Genilloud

Teacher:

Miguel García-Remesal

Date:
March 26, 2025

Contents

1 Introduction

LinkedIn and Coursera are two prominent profile-based retrieval systems that demonstrate how user profiles can drive personalized content delivery. Both platforms excel in collecting, analyzing, and leveraging user data to facilitate connections between professionals and job opportunities and learners and educational content.

LinkedIn operates in the domain of Job Portals, serving as the world's largest professional networking platform. Its sophisticated Job Matching system connects job seekers with relevant opportunities based on their skills, experience, and professional connections. This intelligent matching capability enhances job search efficiency and provides better quality applicants for specific roles to employers.

Coursera, on the other hand, operates in the domain of Online Learning Platforms, providing accessible education from leading universities and organizations worldwide. Its Skill Development system delivers personalized courses tailored to individual learning goals and career aspirations, enhancing access to education regardless of geographical or financial barriers. Coursera also improves career opportunities through specialized courses and certifications recognized by employers globally, allowing learners to acquire in-demand skills and credentials.

2 LinkedIn

LinkedIn's job recommendation system is designed to help users discover relevant job opportunities efficiently and effectively. To achieve this, LinkedIn employs a multi-stage process combining advanced machine learning techniques and data-driven strategies. This system operates in three key stages: retrieval, ranking, and forecasting — each contributing to delivering personalized, balanced, and timely job recommendations.

2.1 Retrieval Stage

The first step in the recommendation pipeline is the Retrieval Stage, where LinkedIn filters a massive pool of job listings to a smaller subset. This is done using a candidate selection model that rapidly eliminates irrelevant jobs. To achieve low search latency and high responsiveness, LinkedIn employs decision trees to partition the data based on user profiles and job characteristics. Additionally, query construction optimizes the retrieval process by dynamically crafting queries that reflect user preferences, past behavior, and job attributes.

2.2 Ranking Stage

Once a smaller set of jobs is retrieved, LinkedIn's system ranks them according to their relevance to the user. This Ranking Stage leverages both explicit and implicit data signals. Explicit user data includes profile details, job preferences, and past applications, while implicit signals capture user behavior, such as clicks and engagement with job listings. To combine these signals effectively, LinkedIn applies advanced machine learning models like *Generalized Linear Mixed Models (GLMix)*, which balances fixed patterns with user-specific behavior.

2.3 Forecasting

The final stage, Forecasting, ensures a balanced distribution of applications across job listings. LinkedIn's *LinkedIn Job-Applied Rate, (LiJAR System)* prevents certain job posts from receiving too many or too few applications. This system forecasts the likelihood of a job receiving sufficient applications and adjusts the visibility of job listings accordingly. By promoting under-applied jobs and limiting over-saturated ones, the system helps recruiters attract a fair number of candidates while ensuring job seekers have a diverse and balanced set of opportunities.

3 Coursera

Coursera's sophisticated recommendation system lies at the heart of its ability to connect millions of learners with appropriate educational content. By implementing a multi-faceted approach to understanding both user preferences and course attributes, the platform creates personalized learning pathways that maximize engagement and learning outcomes. The system employs several complementary methodologies that work in concert to deliver increasingly refined recommendations as users interact with the platform. Below is an examination of the key components that power Coursera's recommendation engine.

3.1 Collaborative Filtering (CF)

Coursera's recommendation engine leverages collaborative filtering approaches to connect learners with appropriate courses. The platform implements: User-based collaborative filtering that analyzes patterns among similar users to make course recommendations. When a learner shares behavioral patterns with others, the system suggests courses that those similar users have successfully engaged with or completed. This creates a community-driven recommendation ecosystem where course discoveries are informed by collective user experiences.

Item-based collaborative filtering that identifies relationships between different courses based on user interaction patterns. By recognizing that certain courses are frequently taken together or in sequence, Coursera can recommend logical next steps in a learning journey. This approach helps learners build coherent skill pathways rather than disconnected learning experiences.

3.2 Profiling

Coursera employs dual profiling strategies to understand learner needs: Explicit profiling enables users to directly communicate their learning objectives through questionnaires, career goals selection, and subject interest indicators. Learners can specifically state what skills they hope to develop or which career paths they aim to pursue, allowing for immediately relevant recommendations without waiting for behavioral data to accumulate. Implicit profiling works behind the scenes by tracking user behavior such as course viewing patterns, completion rates, time spent on specific topics, and quiz performance. This content-based filtering system continuously updates user preference profiles dynamically, ensuring recommendations evolve in alignment with changing interests and progress.

3.3 Content-Based Filtering (CBF)

Coursera's content understanding capabilities form another pillar of its recommendation system:

Analysis of metadata allows the platform to comprehensively catalog courses through keywords, topics, difficulty levels, institutional affiliations, and detailed descriptions. This rich taxonomic approach ensures recommendations can be matched to highly specific learning needs based on content characteristics.

User profile building occurs as learners interact with the platform, creating detailed maps of their knowledge domains, preferred learning styles, and content engagement patterns. These profiles become increasingly sophisticated over time, enabling more precise content matches based on previous positive learning experiences.

3.4 Hybridisation

Coursera implements multiple hybrid approaches to maximize recommendation quality:

Weighted hybrid systems combine collaborative and content-based filtering outputs with dynamically adjusted weighting. For new users with limited history, content-based recommendations might receive higher weighting, while established users benefit from stronger collaborative influences.

Switching hybrid mechanisms intelligently select the most appropriate filtering method based on data availability. For example, the system might default to content-based recommendations for niche subjects with sparse user data, while leveraging collaborative insights for popular course categories.

Feature augmentation strategies enable one recommendation model to enhance another by generating intermediate features. Collaborative filtering might identify relevant skill clusters that then augment content-based recommendations, creating a more nuanced understanding of both content relationships and user needs.

4 Comparison

Coursera and LinkedIn are two platforms that use profile-based information retrieval techniques to deliver personalized experiences. Coursera's recommendation system focuses on educational development, creating personalized learning pathways through a hybrid approach that combines collaborative filtering with content-based recommendations. Its strength lies in adaptability, as it constructs multidimensional user profiles through explicit declarations and implicit signals. However, Coursera's system has limitations, such as a "*filter bubble*" effect, which can limit exposure to interdisciplinary topics.

On the other hand, LinkedIn's recommendation engine optimizes for professional advancement and network development within a larger ecosystem. Its system demonstrates remarkable scalability, efficiently processing massive datasets of professional interactions, skill endorsements, and career transitions while maintaining high availability for job seekers and employers. LinkedIn's implicit profiling capability extends beyond on-platform interactions to incorporate broader job market trends, industry patterns, and professional peer activities.

However, LinkedIn struggles with rapid personalization adjustments, particularly when users undergo significant career pivots that diverge from their established professional history. Additionally, its network-effect driven recommendations can amplify existing biases in the professional landscape, potentially over-representing popular job roles while limiting visibility for more diverse career paths. Both platforms demonstrate the power of profile-based retrieval in delivering personalized experiences, but their different implementation priorities reflect their core objectives.

5 Conclusions

The comparison between LinkedIn and Coursera highlights the power of profile-based retrieval systems in tailoring user experiences across different domains. While LinkedIn optimizes professional networking and job matching through structured profiling and predictive modeling, Coursera focuses on adaptive learning pathways by leveraging hybrid recommendation strategies. Despite their effectiveness, both platforms face challenges such as bias in recommendations, limited adaptability to rapid user changes, and the risk of reinforcing existing patterns rather than fostering exploration. Addressing these issues through explainable AI, external data integration, and dynamic profiling techniques could significantly enhance their personalization capabilities. As these systems evolve, the balance between automation and user control will be crucial in ensuring relevance, fairness, and user satisfaction.

6 Improvements proposals

Both LinkedIn and Coursera need to implement new techniques to enhance their recommendation systems to improve user experience, tackle bias, and improve personalization.

6.1 LinkedIn proposals

LinkedIn can focus on transparency and personalization by using **Explainable AI** to ensure that users and recruiters understand the recommendations, reducing hidden bias. Methodologies such as **NLP on user interactions**, make inferences of job preferences more intuitively. In addition, LinkedIn can implement a **bidirectional recommendation system**, helping job seekers and recruiters find better matches. To further improve personalization, it is suggested to integrate **external labor market data**, allowing the system to reflect broader trends in the labor market.

6.2 Coursera proposals

On the other hand, Coursera could face bias through **Adversarial Debiasing**, preventing over-reliance on past user behavior to encourage more diverse learning paths. To gather better user feedback, **NLP-based chatbots** are an efficient solution, collecting insights directly from learners to refine recommendations. Coursera can also improve content understanding by applying **NLP** to go beyond simple course titles and tags, ensuring that recommendations align with the deeper context of what users want to learn. For a more adaptive learning experience, **AI-driven adaptive learning** integrates user profiles more effectively to tailor course recommendations based on progress, speed, and individual goals.

References

[1] Viet Ha-Thuc, Ganesh Venkataraman, Mario Rodriguez, Shakti Sinha, Senthil Sundaram, Lin Guo "Personalized Expertise Search at LinkedIn" https://arxiv.org/abs/1602.04572

[2] Krishnaram Kenthapadi, Benjamin Le, Ganesh Venkataraman "Personalized Job Recommendation System at LinkedIn: Practical Challenges and Lessons Learned" https://www.researchgate.net/publication/319285276_Personalized_Job_Recommendation_System_at_LinkedIn_Practical_Challenges_and_Lessons_Learned

[3] Jianqiang Shen, Yuchin Juan, Shaobo Zhang, Ping Liu, Wen Pu, Sriram Vasudevan, Qingquan Song, Fedor Borisyuk, Kay Qianqi Shen, Haichao Wei, Yunxiang Ren, Yeou S. Chiou, Sicong Kuang, Yuan Yin, Ben Zheng, Muchen Wu, Shaghayegh Gharghabi, Xiaoqing Wang, Huichao Xue, Qi Guo, Daniel Hewlett, Luke Simon, Liangjie Hong, Wenjing Zhang "Learning to Retrieve for Job Matching" https://arxiv.org/abs/2402.13435

[4] Jiarui Rao, Conrad Borchers, Jionghao Lin "Coursera-REC: Explainable MOOCs Course Recommendation using RAG-facilitated LLMs" https://www.researchgate.net/publication/381725424_Coursera-REC_Explainable_MOOCs_Course_Recommendation_using_RAG-facilitated_LLMs

[5] Deepani B. Guruge, Rajan KadelORCID, Sharly J. Halder "The State of the Art in Methodologies of Course Recommender Systems—A Review of Recent Research" https://www.mdpi.com/2306-5729/6/2/18

[6] Kalyan Kumar Jena, Sourav KUMAR Bhoi, Biju Patnaik, Tushar Kanta Malik, Kshira Sagar Sahoo "E-Learning Course Recommender System Using Collaborative Filtering Models" https://www.researchgate.net/publication/366717794_E-Learning_Course_Recommender_System_Using_Collaborative_Filtering_Models

[7] Fazel Keshtkar, Candice Burkett, Haiying Li, Arthur C. Graesser "Using Data Mining Techniques to Detect the Personality of Players in an Educational Game" https://link.springer.com/chapter/10.1007/978-3-319-02738-8_5