2025 Fall ParAlgo HPC Project Submissions

Task 01

Group Members:
Md. Golam Kibria
21201553
Aryan Shrestha
16201074

Paper Title:

A Privacy-Preserving Distributed Movie Recommendation System

Paper Link:

1. https://ieeexplore.ieee.org/document/9215973
2. https://journals.sagepub.com/doi/10.3233/WEB-230346
3. https://dl.acm.org/doi/abs/10.5555/3504035.3504067
4. https://dl.acm.org/doi/full/10.1145/3708982

1 Summary:

1.1 Motivation

The primary drivers for this project are the inherent flaws in traditional, centralized recommendation systems. The authors identify three specific pain points:

* Privacy Concerns: Centralized servers often require users to upload sensitive preference data.

* Scalability Issues: Centralized systems face bottlenecks as user bases grow.

* Fairness and Security: Traditional models may lack transparency in suggestions and are vulnerable to data breaches.

1.2 Contribution

The project contributes a novel framework that shifts the computation from a central authority to the "edge" (the user's side). Key contributions include:

*Privacy-Preserving Architecture: Preferences are processed locally, ensuring data never leaves the user's control.

*Offline Functionality: Unlike most systems, this can generate recommendations without an active internet connection.

*Ethical Alignment: Explicit implementation of FAIR principles (Findable, Accessible, Interoperable, and Reusable) and ethical data practices.

*Fault-Tolerant Infrastructure: A design that remains functional even if multiple nodes fail.

1.3 Methodology

The system uses a multi-faceted technical approach involving containerization and content-based filtering.

* System Architecture: The system is divided into three Dockerized components:
1. Scraping Server: Collects data from IMDb.
2. Recommendation Server: Handles the user interface and local query processing.
3. Rebuild Server: Updates the model periodically with new data.

* The Recommendation Algorithm: The project utilizes a Content-Based Filtering approach:

1. Vectorization: It uses CountVectorizer to turn movie metadata into numerical vectors.

2. Similarity: It applies Cosine Similarity to measure the distance between a user's preferred movie profile and other movies in the 25,000-movie dataset.

* Decentralization Logic:

1. Discovery: Uses UDP broadcasts for peer-to-peer (P2P) discovery.

2. Propagation: Updates are shared across nodes to ensure the local datasets remain current.

1.4 Conclusion

The authors conclude that the system is a viable alternative to centralized models. It achieves:

* High Performance: 85% positive user rating and a precision score of 0.82.

* Resilience: The system proved robust, showing only a 5% increase in response time despite a 20% node failure rate.

* Sustainability: It demonstrates near-linear scalability, proving that adding more users does not crash the system.

2 Limitations

2.1 First Limitation : Lack of Serendipity (Filter Bubble)

While not explicitly stated as a failure, the methodology reveals a limitation: the use of a Content-based recommendation algorithm. This type of algorithm only recommends items similar to what the user has liked before. Unlike "Collaborative Filtering," it cannot suggest movies that are different but liked by "similar users," potentially trapping the user in a "filter bubble" where they never see diverse content.

2.2 Second Limitation: Resource Intensity on Local Nodes

The architecture requires each node to store data locally and run a Rebuild Server periodically. For users with low-end hardware or limited mobile battery life, running Docker containers and performing text vectorization/cosine similarity calculations on 25,000 records may lead to significant computational overhead and battery drain, which could hinder widespread adoption on mobile devices.

3 Synthesis

The Decentralized Movie Recommendation System represents a significant step toward "User-Centric AI." By combining Dockerized microservices with a P2P discovery mechanism, the system successfully decouples personalization from data surveillance. While the reliance on content-based filtering may limit the "surprise" factor of recommendations, the system’s ability to function offline and its 82% precision rate make it a powerful proof-of-concept for private, ethical, and resilient distributed computing in the entertainment industry.