

Rapport de Système de Recommandation de Films

Hadj-said Samy

July 7, 2024

1 Introduction

Ce rapport présente le développement et l'évaluation d'un système de recommandation de films basé sur des techniques de filtrage collaboratif. L'objectif est de recommander des films aux utilisateurs en se basant sur leurs préférences passées ainsi que sur les caractéristiques des films eux-mêmes. Les données utilisées proviennent des ensembles de données IMDb et MovieLens, chacun offrant des perspectives uniques sur les films à travers des informations détaillées sur les titres, les genres, les années de sortie et les évaluations des utilisateurs.

2 Chargement et Prétraitement des Données

Pour assurer la qualité des recommandations, il est crucial de prétraiter les données provenant de différentes sources. Les données de MovieLens ont été nettoyées et enrichies en extrayant les années de sortie des titres de films et en normalisant les titres pour une correspondance précise. De même, les données d'IMDb ont été traitées pour gérer les valeurs manquantes et normaliser les titres des films afin de faciliter la fusion et l'analyse ultérieure.

3 Exploration des Données

3.1 Distribution des Années de Sortie des Films

L'exploration initiale des données fusionnées révèle une distribution significative des années de sortie des films. Comme le montre la Figure 1, une forte concentration de films récents est observée, avec une croissance marquée dans les dernières décennies. Cette information est cruciale pour comprendre la dynamique temporelle des préférences cinématographiques des utilisateurs.

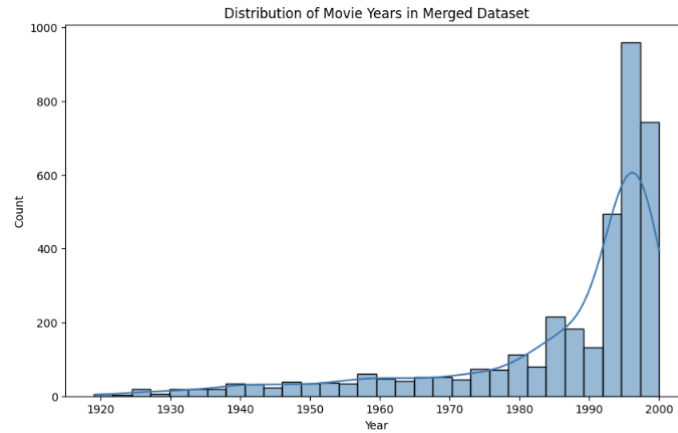


Figure 1: Distribution des Années de Sortie des Films dans l'Ensemble de Données Fusionné

3.2 Distribution des Genres de Films

En examinant la distribution des genres de films (Figure 2), il est clair que certains genres sont plus populaires que d'autres. Les genres comme le drame, la comédie et l'action dominent, reflétant les tendances générales de l'industrie cinématographique et fournissant des insights précieux pour la personnalisation des recommandations.

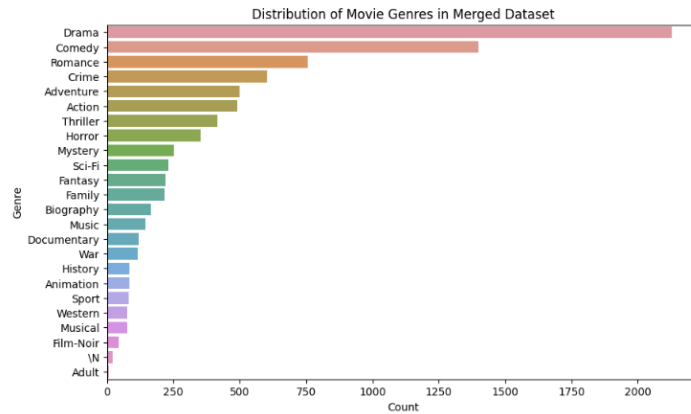


Figure 2: Distribution des Genres de Films dans l'Ensemble de Données Fusionné

4 Développement du Système de Recommandation

Le cœur du projet réside dans le développement d'un système de recommandation efficace. En utilisant les données prétraitées, un modèle de filtrage collaboratif a été mis en place pour prédire les préférences des utilisateurs en fonction de leurs interactions passées avec des films similaires.

4.1 Construction du Modèle de Recommandation

Le modèle de recommandation repose sur une matrice de similarité calculée à l'aide de la similarité cosinus entre les évaluations des utilisateurs pour différents films. Cette approche permet de capturer les similitudes subtiles entre les préférences des utilisateurs et de recommander des films qui correspondent étroitement à leurs goûts.

5 Évaluation du Système de Recommandation

5.1 Génération de Recommandations

Pour illustrer l'efficacité du modèle, des recommandations ont été générées pour des films spécifiques. Par exemple, en entrant "Toy Story", le système identifie et suggère une liste de films similaires qui pourraient intéresser l'utilisateur en fonction des évaluations et des genres des films.

6 Conclusion

Ce rapport a exploré le processus de développement d'un système de recommandation de films en utilisant des techniques de filtrage collaboratif. En intégrant des données détaillées provenant de MovieLens et IMDb, le modèle offre des recommandations personnalisées en se basant sur les préférences des utilisateurs et les caractéristiques des films. L'analyse des données a permis de comprendre les tendances cinématographiques et d'élaborer un modèle robuste pour améliorer l'expérience utilisateur dans la découverte de nouveaux films.