

Catalogue de Données

1. Aperçu des Données

1.1 Source de Données

Les données utilisées dans ce projet proviennent de l'ensemble de données MovieLens, collecté par le projet de recherche GroupLens à l'Université du Minnesota. Le jeu de données comprend 100 000 évaluations de la part de 943 utilisateurs sur 1682 films.

1.2 Données Utilisées

Les données spécifiques utilisées dans ce projet sont représentées par une entrée unique utilisateur-film-évaluation :

{ userId / movie { movieId / title / genres } / rating / timestamp }

2. Lisez-moi des Sources de Données

2.1 Résumé

L'ensemble de données MovieLens se compose de 100 000 évaluations de la part de 943 utilisateurs sur 1682 films. Chaque utilisateur a évalué au moins 20 films, et le jeu de données inclut des informations démographiques simples sur les utilisateurs, telles que l'âge, le genre, la profession et le code postal.

2.2 Licence d'Utilisation

Les données sont fournies sous certaines conditions :

- Aucune approbation de l'Université du Minnesota ou du GroupLens Research Group n'est implicite.
- La reconnaissance de l'utilisation de l'ensemble de données est requise dans les publications résultant de son utilisation.
- La redistribution des données nécessite une autorisation distincte.
- L'utilisation commerciale ou à but lucratif nécessite l'autorisation du projet de recherche GroupLens.

2.3 Citation

Pour reconnaître l'utilisation de l'ensemble de données dans les publications, veuillez citer l'article suivant :

- F. Maxwell Harper et Joseph A. Konstan. 2015. Les ensembles de données MovieLens : Histoire et Contexte. ACM Transactions on Interactive Intelligent Systems (TiiS) 5, 4, Article 19 (décembre 2015), 19 pages. DOI=<http://dx.doi.org/10.1145/2827872>

2.4 Remerciements

Merci à Al Borchers pour le nettoyage de ces données et la rédaction des scripts associés.

2.5 Travaux Publiés

L'ensemble de données a été utilisé dans des travaux publiés, notamment :

- Herlocker, J., Konstan, J., Borchers, A., Riedl, J. Un cadre algorithmique pour la réalisation de la filtration collaborative. Actes de la Conférence sur la Recherche et le Développement en Recherche d'Information de 1999. Août 1999.

2.6 Informations Supplémentaires

Des informations supplémentaires sur le projet de recherche GroupLens peuvent être trouvées sur [GroupLens Research](<http://www.grouplens.org/>).

3. Descriptions Détaillées des Fichiers de Données

3.1 ml-data.tar.gz

- Fichier tar.gz compressé. Pour reconstruire les fichiers de données, suivez ces étapes :
 - gunzip ml-data.tar.gz
 - tar xvf ml-data.tar
 - mku.sh

3.2 u.data

- Jeu de données complet avec 100 000 évaluations de 943 utilisateurs sur 1682 éléments.
- Liste séparée par des tabulations de user id | item id | rating | timestamp.

3.3 u.info

- Informations sur le nombre d'utilisateurs, d'éléments et d'évaluations dans le jeu de données u.

3.4 u.item

- Informations sur les films, y compris l'ID du film, le titre, la date de sortie, les genres et l'URL IMDb.

3.5 u.genre

- Liste des genres.

3.6 u.user

- Informations démographiques sur les utilisateurs, sous la forme d'une liste séparée par des tabulations de user id | age | gender | occupation | code postal.

3.7 u.occupation

- Liste des professions.

3.8 u1.base à u5.test, ua.base à ub.test

- Ensembles de données pour les séparations d'entraînement et de test pour la validation croisée.

3.9 allbut.pl

- Script générant des ensembles d'entraînement et de test où toutes sauf n évaluations d'un utilisateur sont dans les données d'entraînement.

3.10 mku.sh

- Script shell pour générer tous les ensembles de données u à partir de u.data.

Pour toute question ou commentaire, veuillez contacter GroupLens à l'adresse suivante : [<grouplens-info@cs.umn.edu>](mailto:grouplens-info@cs.umn.edu).

Ce document fournit un aperçu des données utilisées dans le projet, de leurs sources et des informations détaillées sur les fichiers du jeu de données. Pour plus d'informations sur le jeu de données, consultez la documentation originale du jeu de données MovieLens et le site web du projet de recherche GroupLens.

