

Resumen

Este conjunto de datos (ml-latest-small) describe de texto libre de [MovieLens](#), un servicio de reco y 1296 aplicaciones de etiquetas en 9125 película de enero de 1995 y el 16 de octubre de 2016. Est

Texto original en inglés:



Each user is represented by an id, and no other information is provided.

[+ Proponer una traducción mejor](#)

Los usuarios fueron seleccionados al azar para su inclusión. Todos los usuarios seleccionados habían calificado al menos 20 películas. No se incluye información demográfica. Cada usuario está representado por una identificación, y no se proporciona ninguna otra información.

Los datos están contenidos en los archivos `links.csv`, `movies.csv`, `ratings.csv` y `tags.csv`. Más detalles sobre los contenidos y el uso de todos estos archivos siguen.

Este es un conjunto de datos de *desarrollo*. Como tal, puede cambiar con el tiempo y no es un conjunto de datos apropiado para resultados de investigación compartidos. Vea los conjuntos de datos de *referencia* disponibles si esa es su intención.

Este y otros conjuntos de datos de GroupLens están públicamente disponibles para su descarga en <http://grouplens.org/datasets/>.

Licencia de uso

Ni la Universidad de Minnesota ni ninguno de los investigadores involucrados pueden garantizar la exactitud de los datos, su idoneidad para un propósito particular o la validez de los resultados en función del uso del conjunto de datos. El conjunto de datos se puede usar para cualquier propósito de investigación bajo las siguientes condiciones:

- El usuario no puede afirmar ni implica ningún respaldo de la Universidad de Minnesota ni del Grupo de investigación GroupLens.
- El usuario debe reconocer el uso del conjunto de datos en las publicaciones resultantes del uso del conjunto de datos (ver a continuación la información de las citas).
- El usuario puede redistribuir el conjunto de datos, incluidas las transformaciones, siempre que se distribuya bajo estas mismas condiciones de licencia.
- El usuario no puede usar esta información para fines comerciales o de obtención de ingresos sin antes obtener el permiso de un miembro de la facultad del Proyecto de Investigación GroupLens en la Universidad de Minnesota.
- Las secuencias de comandos del software ejecutable se proporcionan "tal cual" sin garantía de ningún tipo, ya sea expresa o implícita, incluidas, entre otras, las garantías implícitas de comerciabilidad e idoneidad para un fin determinado. El riesgo total en cuanto a la calidad y el rendimiento de ellos está con usted. Si el programa resultara defectuoso, usted asumirá el costo de todos los servicios, reparaciones o correcciones necesarios.

En ningún caso la Universidad de Minnesota, sus afiliados o empleados serán responsables ante usted por los daños y perjuicios que surjan del uso o la imposibilidad de utilizar estos programas (incluida, entre otros, la pérdida de datos o datos que se vuelven inexactos).

Si tiene más preguntas o comentarios, envíe un correo electrónico a grouplens-info@umn.edu

Citación

Para reconocer el uso del conjunto de datos en publicaciones, cite el siguiente documento:

F. Maxwell Harper y Joseph A. Konstan. 2015. Los conjuntos de datos MovieLens: historia y contexto. Transacciones de ACM en sistemas inteligentes interactivos (TiiS) 5, 4, artículo 19 (diciembre de 2015), 19 páginas. DOI = <http://dx.doi.org/10.1145/2827872>

Más información sobre GroupLens

GroupLens es un grupo de investigación en el Departamento de Ciencias de la Computación e Ingeniería de la Universidad de Minnesota. Desde su inicio en 1992, los proyectos de investigación de GroupLens han explorado una variedad de campos que incluyen:

- sistemas de recomendación
- comunidades en línea
- tecnologías móviles y ubicuas
- bibliotecas digitales
- sistemas de información geográfica local

GroupLens Research opera un recomendador de películas basado en el filtrado colaborativo, MovieLens, que es la fuente de estos datos. ¡Le recomendamos que visite <http://movielens.org> para probarlo! Si tiene ideas interesantes para realizar trabajos experimentales en MovieLens, envíenos un correo electrónico a grouplens-info@cs.umn.edu ; siempre estamos interesados en trabajar con colaboradores externos.

Contenido y uso de archivos

Formateo y Codificación

Los archivos del conjunto de datos se escriben como archivos de [valores separados por comas](#) con una sola fila de encabezado. Las columnas que contienen comas (,) se escapan usando comillas dobles ("). Estos archivos están codificados como UTF-8. Si los caracteres acentuados en títulos de películas o valores de etiquetas (por ejemplo, Misérables, Les (1995)) se muestran incorrectamente, asegúrese de que cualquier programa que lea los datos, como un editor de texto, terminal o script, está configurado para UTF-8.

Ids de usuario

Los usuarios de MovieLens fueron seleccionados al azar para su inclusión. Sus identificadores han sido anonimizados. Los `ratings.csv` usuario son consistentes entre `ratings.csv` y `tags.csv` (es decir, el mismo id se refiere al mismo usuario en los dos archivos).

Movie Ids

Solo las películas con al menos una calificación o etiqueta se incluyen en el conjunto de datos. Estos identificadores de películas son consistentes con los utilizados en el sitio web MovieLens (por ejemplo, id 1 corresponde a la URL <https://movielens.org/movies/1>). Los ratings.csv película son consistentes entre ratings.csv , tags.csv , movies.csv y links.csv (es decir, el mismo ID se refiere a la misma película en estos cuatro archivos de datos).

Estructura del archivo de datos de calificaciones (ratings.csv)

Todas las calificaciones están contenidas en el archivo ratings.csv . Cada línea de este archivo después de la fila del encabezado representa una clasificación de una película por un usuario y tiene el siguiente formato:

```
userId,movieId,rating,timestamp
```

Las líneas dentro de este archivo se ordenan primero por ID de usuario, luego, dentro del usuario, por ID de película.

Las calificaciones se realizan en una escala de 5 estrellas, con incrementos de media estrella (0.5 estrellas - 5.0 estrellas).

Las marcas de tiempo representan segundos desde la medianoche hora universal coordinada (UTC) del 1 de enero de 1970.

Etiquetas Estructura del archivo de datos (tags.csv)

Todas las etiquetas están contenidas en el archivo tags.csv . Cada línea de este archivo después de la fila del encabezado representa una etiqueta aplicada a una película por un usuario y tiene el siguiente formato:

```
userId,movieId,tag,timestamp
```

Las líneas dentro de este archivo se ordenan primero por ID de usuario, luego, dentro del usuario, por ID de película.

Las etiquetas son metadatos generados por el usuario sobre películas. Cada etiqueta es típicamente una sola palabra o frase corta. El significado, el valor y el propósito de una etiqueta en particular los determina cada usuario.

Las marcas de tiempo representan segundos desde la medianoche hora universal coordinada (UTC) del 1 de enero de 1970.

Estructura del archivo de datos de películas (movies.csv)

La información de la película está contenida en el archivo movies.csv . Cada línea de este archivo después de la fila del encabezado representa una película y tiene el siguiente formato:

```
movieId,title,genres
```

Los títulos de películas se ingresan manualmente o se importan de <https://www.themoviedb.org/> e incluyen el año de publicación entre paréntesis. Pueden existir errores e inconsistencias en estos títulos.

Los géneros son una lista separada por tuberías y se seleccionan de los siguientes:

- Acción
- Aventuras
- Animación
- Para niños
- Comedia
- Crimen
- Documental
- Drama
- Fantasía
- Film-Noir
- Horror
- Musical
- Misterio
- Romance
- Ciencia ficción
- Novela de suspenso
- Guerra
- occidental
- (no géneros enumerados)

Estructura del archivo de datos de enlaces (links.csv)

Los identificadores que se pueden usar para vincular a otras fuentes de datos de películas se encuentran en el archivo `links.csv`. Cada línea de este archivo después de la fila del encabezado representa una película y tiene el siguiente formato:

```
movieId,imdbId,tmdbId
```

`movieId` es un identificador de las películas utilizadas por <https://movielens.org>. Por ejemplo, la película Toy Story tiene el enlace <https://movielens.org/movies/1>.

`imdbId` es un identificador para las películas utilizadas por <http://www.imdb.com>. Por ejemplo, la película Toy Story tiene el enlace <http://www.imdb.com/title/tt0114709/>.

`tmdbId` es un identificador de las películas utilizadas por <https://www.themoviedb.org>. Por ejemplo, la película Toy Story tiene el enlace <https://www.themoviedb.org/movie/862>.

El uso de los recursos enumerados anteriormente está sujeto a los términos de cada proveedor.

Validación cruzada

Las versiones anteriores del conjunto de datos MovieLens incluían pliegues cruzados precomputados o scripts para realizar este cálculo. Ya no combinamos ninguna de estas características con el conjunto de datos, ya que la mayoría de los kits de herramientas modernos proporcionan esto como una característica incorporada. Si desea obtener información acerca de los enfoques estándar para el cálculo de pliegues cruzados en el contexto de la evaluación de los sistemas de recomendación, consulte [LensKit](#) para obtener herramientas, documentación y ejemplos de código fuente abierto.