Resumo
=======

Este conjunto de dados (ml-latest-small) descreve a classificação de 5 estrelas e a atividade de marcação de texto livre do [MovieLens] (http://movielens.org), um serviço de recomendação de filmes. Ele contém 100004 avaliações e 1296 tags em 9125 filmes. Estes dados foram criados por 671 usuários entre 09 de janeiro de 1995 e 16 de outubro de 2016. Este conjunto de dados foi gerado em 17 de outubro de 2016.

Os usuários foram selecionados aleatoriamente para inclusão. Todos os usuários selecionados avaliaram pelo menos 20 filmes. Nenhuma informação demográfica está incluída. Cada usuário é representado por um id e nenhuma outra informação é fornecida.

Os dados estão contidos nos arquivos `links.csv`,` movies.csv`, `ratings.csv` e` tags.csv`. Mais detalhes sobre o conteúdo e uso de todos esses arquivos a seguir.

Este é um conjunto de dados * development *. Como tal, pode mudar ao

longo do tempo e não é um conjunto de dados apropriado para resultados de pesquisa compartilhados. Veja os conjuntos de dados * benchmark * disponíveis se essa for a sua intenção.

Este e outros conjuntos de dados do GroupLens estão disponíveis publicamente para download em <http://grouplens.org/datasets/>.

Licença de Uso
==========

Nem a Universidade de Minnesota nem nenhum dos pesquisadores envolvidos pode garantir a exatidão dos dados, sua adequação a qualquer propósito específico ou a validade dos resultados com base no uso do conjunto de dados. O conjunto de dados pode ser usado para quaisquer fins de pesquisa nas seguintes condições:

* O usuário não pode declarar ou sugerir qualquer endosso da Universidade de Minnesota ou do GroupLens Research Group.
* O usuário deve reconhecer o uso do conjunto de dados em publicações resultantes do uso do conjunto de dados

(veja abaixo as informações sobre citações).
* O usuário pode redistribuir o conjunto de dados, incluindo transformações, desde que seja distribuído sob essas mesmas condições de licença.
* O usuário não pode usar essas informações para fins comerciais ou com fins lucrativos sem primeiro obter permissão de um membro do corpo docente do Projeto de Pesquisa do GroupLens na Universidade de Minnesota.
* Os scripts do software executável são fornecidos "como estão", sem garantia de qualquer tipo, expressa ou implícita, incluindo, mas não se limitando às garantias implícitas de comercialização e adequação a uma finalidade específica. Todo o risco quanto à qualidade e desempenho deles está com você. Caso o programa se mostre defeituoso, você assume o custo de toda a manutenção, reparo ou correção necessários.

Em nenhum caso a Universidade de Minnesota, suas afiliadas ou funcionários serão responsáveis por quaisquer danos decorrentes do uso ou incapacidade de usar esses programas (incluindo, mas não se limitando a, perda de dados ou dados imprecisos).

Se você tiver mais perguntas ou comentários, envie um e-mail para <grouplens-info@umn.edu>

Citação
========

Para reconhecer o uso do conjunto de dados em publicações, cite o seguinte artigo:

> Maxwell Harper e Joseph A. Konstan. 2015. Os conjuntos de dados MovieLens: histórico e contexto. Transações ACM em Sistemas Inteligentes Interativos (TiiS) 5, 4, Artigo 19 (dezembro de 2015), 19 páginas. DOI = <http://dx.doi.org/10.1145/2827872>

Mais informações sobre o GroupLens
==========

GroupLens é um grupo de pesquisa no Departamento de Ciência da Computação e Engenharia da Universidade de Minnesota. Desde a sua criação em 1992, os projetos de pesquisa do GroupLens exploraram uma variedade de campos, incluindo:

* sistemas de recomendação

* comunidades online
* tecnologias móveis e ubíquas
* bibliotecas digitais
* sistemas de informação geográfica local

A GroupLens Research opera uma recomendação de filme baseada na filtragem colaborativa, MovieLens, que é a fonte desses dados. Nós encorajamos você a visitar <http://movielens.org> para experimentá-lo! Se você tiver ideias interessantes para trabalhos experimentais no MovieLens, envie-nos um e-mail para <grouplens-info@cs.umn.edu> - estamos sempre interessados em trabalhar com colaboradores externos.

Conteúdo e Uso de Arquivos

Formatação e Codificação

Os arquivos do conjunto de dados são gravados como arquivos [valores separados por vírgula] (http://en.wikipedia.org/wiki/Comma-separated_values) com uma única linha de cabeçalho. Colunas que contêm vírgulas (`,`) são escapadas usando aspas duplas

(`" `). Esses arquivos são codificados como UTF-8. Se caracteres acentuados em títulos de filmes ou valores de tags (por exemplo, Misérables, Les (1995)) são exibidos incorretamente , certifique-se de que qualquer programa que esteja lendo os dados, como um editor de texto, terminal ou script, esteja configurado para UTF-8.

IDs de usuário

Os usuários do MovieLens foram selecionados aleatoriamente para inclusão. Seus ids foram anonimizados. Ids de usuário são consistentes entre `ratings.csv` e` tags.csv` (isto é, o mesmo id refere-se ao mesmo usuário nos dois arquivos).

IDs de filmes

Somente filmes com pelo menos uma classificação ou tag estão incluídos no conjunto de dados. Esses IDs de filmes são consistentes com aqueles usados no site do MovieLens (por exemplo, id `1` corresponde ao URL <https://movielens.org/movies/1>). Os IDs dos filmes são consistentes entre

`ratings.csv`,` tags.csv`, `movies.csv` e` links.csv` (isto é, o mesmo id refere-se ao mesmo filme nesses quatro arquivos de dados).

Estrutura de arquivos de dados de classificações (ratings.csv)

Todas as classificações estão contidas no arquivo `ratings.csv`. Cada linha desse arquivo após a linha de cabeçalho representa uma avaliação de um filme por um usuário e tem o seguinte formato:

userId, movieId, classificação, timestamp

As linhas nesse arquivo são ordenadas primeiro por userId e, em seguida, por user, por movieId.

As classificações são feitas em uma escala de 5 estrelas, com incrementos de meia estrela (0,5 estrelas - 5,0 estrelas).

Os timestamps representam segundos desde a meia-noite do Tempo Universal

Coordenado (UTC) de 1º de janeiro de 1970.

Estrutura de arquivos de dados (tags.csv)

Todas as tags estão contidas no arquivo `tags.csv`. Cada linha desse arquivo após a linha do cabeçalho representa uma tag aplicada a um filme por um usuário e tem o seguinte formato:

userId, movieId, tag, timestamp

As linhas nesse arquivo são ordenadas primeiro por userId e, em seguida, por user, por movieId.

Tags são metadados gerados pelo usuário sobre filmes. Cada tag é tipicamente uma única palavra ou frase curta. O significado, o valor e a finalidade de uma tag específica são determinados por cada usuário.

Os timestamps representam segundos desde a meia-noite do Tempo Universal Coordenado (UTC) de 1º de janeiro de 1970.

Estrutura de arquivos de dados de filmes (movies.csv)

A informação do filme está contida no arquivo `movies.csv`. Cada linha desse arquivo após a linha de cabeçalho representa um filme e tem o seguinte formato:

movieId, título, gêneros

Os títulos de filmes são inseridos manualmente ou importados de <https://www.themoviedb.org/> e incluem o ano de lançamento entre parênteses. Erros e inconsistências podem existir nesses títulos.

Os gêneros são uma lista separada por pipe e são selecionados entre os seguintes:

* Açao
* Aventura
* Animação
* Crianças
Comédia
* Crime
* Documentário
* Drama
* Fantasia

* Film-Noir
* Horror
* Musical
* Mistério
* Romance
* Ficção científica
* Suspense
Guerra
* Ocidental
* (sem gêneros listados)

Estrutura de arquivos de dados de links (links.csv)

Identificadores que podem ser usados para ligar a outras fontes de dados de filmes estão contidos no arquivo `links.csv`. Cada linha desse arquivo após a linha de cabeçalho representa um filme e tem o seguinte formato:

movieId, imdbId, tmdbId

O movieId é um identificador dos filmes usados pelo <https://movielens.org>. Por exemplo, o filme Toy Story tem o link <https://movielens.org/movies/1>.

imdbId é um identificador para filmes usados por <http://www.imdb.com>. Por

exemplo, o filme Toy Story tem o link <http://www.imdb.com/title/tt0114709/>.

tmdbId é um identificador para filmes usados pelo <https://www.themoviedb.org>. Por exemplo, o filme Toy Story tem o link <https://www.themoviedb.org/movie/862>.

O uso dos recursos listados acima está sujeito aos termos de cada provedor.

Validação cruzada

As versões anteriores do conjunto de dados MovieLens incluíam dobras cruzadas pré-computadas ou scripts para executar esse cálculo. Não agrupamos mais esses recursos com o conjunto de dados, pois a maioria dos kits de ferramentas modernos fornece isso como um recurso interno. Se você deseja aprender sobre abordagens padrão para o cálculo de dobra cruzada no contexto de avaliação de sistemas de recomendação , veja [LensKit] (http://lenskit.org) para ferramentas, documentação e exemplos de códigos de código aberto.