

Movie Recommendation System

15 de marzo del 2023

Introducción

Los sistemas de recomendación de películas son una de las aplicaciones más populares en la minería de datos y el aprendizaje automático. Estos sistemas se basan en el análisis de grandes conjuntos de datos para identificar patrones y tendencias que puedan utilizarse para hacer recomendaciones personalizadas a los usuarios. En este informe, se presenta el desarrollo de un algoritmo de recomendación de películas utilizando los datos de la base de datos de MovieLens.

Características Generales:

El conjunto de datos de MovieLens es una colección de calificaciones de películas recopiladas por el GrupoLens de la Universidad de Minnesota. El conjunto de datos incluye información sobre aproximadamente 100,000 calificaciones de películas realizadas por más de 600 usuarios. Cada calificación se acompaña de información sobre la película, como el título, la fecha de lanzamiento, el género y la clasificación por edades. El conjunto de datos de MovieLens proporciona una base sólida para el desarrollo de este tipo de algoritmos y se ha utilizado en numerosos estudios y proyectos de investigación. Ver [link](#).

Definición de las Variables:

Las principales variables en el conjunto de datos de MovieLens son las siguientes:

- **Usuario [userId]:** un identificador único para cada usuario que ha calificado películas en el conjunto de datos.
- **Película [movieId]:** un identificador único para cada película en el conjunto de datos.
- **Calificación [rating]:** una calificación numérica del 1 al 5 que el usuario dio a la película.
- **Fecha de Calificación [timestamp]:** la fecha en que el usuario calificó la película.
- **Título de la película:** el título de la película.
- **Género de la Película:** el género de la película, como comedia, drama o acción.

Objetivos del Algoritmo:

El objetivo principal de este proyecto es desarrollar un algoritmo de recomendación de películas que pueda utilizar el conjunto de datos de MovieLens para hacer recomendaciones personalizadas a los usuarios. Para lograr este objetivo, se han definido los siguientes objetivos específicos:

- Explorar y entender el conjunto de datos de MovieLens.
- Identificar patrones y tendencias en los datos que puedan utilizarse para hacer recomendaciones personalizadas.
- Desarrollar y probar dos métodos de recomendación diferentes: similitud de coseno y coeficiente de correlación de Pearson.
- Utilizar Gephi para la visualización y análisis de grafos para entender mejor la estructura de la red de usuarios y películas en el conjunto de datos.

Métodos de Recomendación

Se han desarrollado dos métodos de recomendación diferentes para este proyecto: similitud de coseno y coeficiente de correlación de Pearson.

1. Similitud de Coseno:

Este método utiliza la similitud de coseno para medir la similitud entre dos vectores de calificación de películas. Para cada usuario, se calcula la similitud de coseno con todos los demás usuarios en el conjunto de datos. A continuación, se seleccionan los usuarios con las mayores similitudes de coseno y se recomiendan las películas que han calificado positivamente pero que el usuario aún no ha visto.

2. Coeficiente de Correlación de Pearson:

Este método utiliza el coeficiente de correlación de Pearson para medir la similitud entre dos vectores de calificación de películas. Para cada usuario, se calcula el coeficiente de correlación de Pearson con todos los demás usuarios en el conjunto de datos. A continuación, se seleccionan los usuarios con los coeficientes de correlación más altos y se recomiendan las películas que han calificado positivamente pero que el usuario aún no ha visto.

Para ambos métodos, se ha utilizado una técnica de filtrado colaborativo, que se basa en la idea de que los usuarios con gustos similares tienden a calificar las mismas películas de manera similar. Estos métodos de recomendación no requieren información adicional sobre las películas o los usuarios, como su género o edad.

Resultados

La precisión se ha medido por la proporción de recomendaciones que el usuario encuentra útiles. Además, la visualización y análisis de grafos con Gephi ha permitido entender mejor la estructura de la red de usuarios y películas en el conjunto de datos y ha ayudado a identificar patrones y tendencias que pueden utilizarse para mejorar aún más el rendimiento del algoritmo.

Conclusión:

El desarrollo de un algoritmo de recomendación de películas es una tarea compleja que requiere la aplicación de técnicas avanzadas de aprendizaje automático y análisis de datos. El conjunto de datos de MovieLens proporciona una base sólida para el desarrollo de este tipo de algoritmos y se ha utilizado en numerosos estudios y proyectos de investigación. El objetivo final del algoritmo es proporcionar recomendaciones personalizadas y relevantes a los usuarios para mejorar su experiencia de visualización de películas. Además, la visualización y análisis de grafos con Gephi ha permitido entender mejor la estructura de la red de usuarios y películas en el conjunto de datos y ha ayudado a identificar patrones y tendencias que pueden utilizarse para mejorar aún más el rendimiento del algoritmo.