Instituto Tecnológico y de Estudios Superiores de Monterrey

Maestría en Inteligencia Artificial Aplicada

Materia:

Análisis de grandes volúmenes de datos

Actividad:

6.2 Avance de proyecto 2: Sistema de Recomendación

Equipo 40:

Cecilia Acevedo Rodríguez- A01793953

Francisco Xavier Bastidas Moreno - A01794188

Ricardo Mar Cupido – A01795394

Edgar Gerardo Rojas Medina - A00840712

Tabla de contenido

Introducción

En esta actividad aplicaremos una técnica avanzada para la creación de un sistema de recomendación, existen varios enfoques para la realización de esta, pero el que se usará será el de factorización matricial.

La factorización matricial es un proceso mediante el cual una matriz se descompone en el producto de dos o más matrices más simples. Estas descomposiciones son fundamentales en ámbitos como el álgebra lineal, la optimización numérica, el análisis de datos y el procesamiento de señales. De acuerdo con Ng and Tan (2021) algunos de los procesos de descompresión de matrices más usados son los siguientes:

1. Descomposición LU: Consiste en descomponer una matriz cuadrada A en el producto de una matriz triangular inferior L y una matriz triangular superior U. Es comúnmente utilizada para resolver sistemas de ecuaciones lineales y para calcular la inversa de una matriz.

2. Descomposición QR: Se descompone una matriz A en el producto de una matriz ortogonal Q y una matriz triangular superior R. Es útil para resolver sistemas de ecuaciones lineales y para encontrar el polinomio característico de una matriz.

3. Descomposición de Cholesky: Esta descomposición es específica para matrices simétricas y definidas positivas. Consiste en descomponer una matriz A en el producto de una matriz triangular inferior L y su traspuesta. Se utiliza en problemas de optimización convexa y simulaciones Monte Carlo.

4. Descomposición en Valores Singulares (SVD): Esta es una de las descomposiciones más utilizadas. Consiste en descomponer una matriz A de tamaño $m \times n$ en tres matrices U, Σ y V^T, donde U y V son matrices ortogonales (sus columnas y filas son ortogonales entre sí) y Σ es una matriz

diagonal con los valores singulares de A. Es útil para tareas como reducción de dimensionalidad, compresión de imágenes y sistemas de recomendación.

Modelo SVD

Para esta actividad se optó utilizar el modelo de SVD para crear nuestro sistema de recomendación, ya que de acuerdo con el artículo de Nguyen (2016), uno de los principales problemas que sufren los sistemas de recomendación es la escasez, causada por una cantidad insuficiente de datos de calificación de los usuarios, y la escalabilidad, causada por datos grandes e intratables, los cuales son desafíos que el modelo de SVD puede abordar eficazmente.

De acuerdo con Sarwar et al. (2002), una de las principales características que tiene el SVD es la capacidad que tiene de reducir la dimensionalidad de la matriz de interacciones usuario-elemento, lo que facilita el cálculo de las preferencias de los usuarios y las características de los elementos.

Esto resulta muy útil para el proceso de filtrado colaborativo, ya que SVD produce un conjunto de vectores propios no correlacionados. Básicamente, cada cliente y producto está representado por su correspondiente vector propio. Estos vectores propios representan las características latentes de los usuarios y los elementos, lo que facilita la identificación de similitudes entre usuarios y elementos en un espacio de características común.

Investigaciones han demostrado que los algoritmos basados en SVD pueden hacer que el proceso de formación de vecindarios de los sistemas de Filtrado Colaborativo sea altamente escalable, a su vez que producen mejores resultados en la mayoría de los casos. A pesar de la buena calidad y el excelente rendimiento en línea, los algoritmos basados en SVD sufren de un inconveniente serio: el paso de descomposición SVD fuera de línea es computacionalmente muy costoso.

Métricas

Para examinar la diferencia entre el método del sistema de recomendación que refleja solo los datos de calificación y el método que integra los datos de calificación con las puntuaciones de sentimiento, se utilizan el error absoluto medio (MAE) y el error cuadrático medio (RMSE) para el método de evaluación (Chai & Draxler, 2014).

Otra de las métricas mas utilizadas para evaluar la precisión de las predicciones de los sistemas de recomendación es el Error Cuadrático Medio (MSE). Básicamente, el MSE mide cuán cerca están las predicciones del modelo de las calificaciones reales proporcionadas por los usuarios.

A continuación, se ofrece una breve descripción de estas métricas:

Error cuadrático medio (RMSE): mide la diferencia entre los valores predichos por el modelo y los valores observados reales. Es la raíz cuadrada de la media de los errores o diferencias de los valores mencionados, al cuadrado.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

$$\begin{aligned} \hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n &\text{ are predicted values} \\ y_1, y_2, \ldots, y_n &\text{ are observed values} \\ n &\text{ is the number of observations} \end{aligned}$$

Ilustración 1: Fórmula RMSE

MSE: similar al RMSE pero sin tomar la raíz cuadrada. Mide la media de los errores al cuadrado entre los valores predichos y los valores observados

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \tilde{y}_i)^2$$

Ilustración 2: Fórmula MSE

MAE: mide la media de las diferencias absolutas entre los valores predichos y los valores observados. A diferencia del MSE y RMSE, el MAE no eleva al cuadrado las diferencias, lo que lo hace menos sensible a valores atípicos

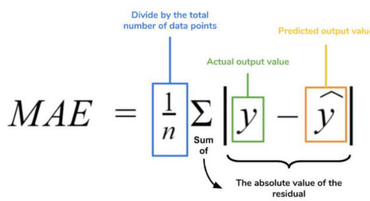

Ilustración 3: Fórmula MAE

En nuestro modelo nuestras métricas obtuvieron los siguientes resultados:

Resultados

Una vez explicadas las métricas y su relevancia para evaluar el desempeño de los sistemas de recomendación, podemos comprender los resultados obtenidos en la implementación de nuestro algoritmo.

RMSE: 1.3640
MAE: 1.0669
MSE: 1.8606

En resumen, estas métricas sugieren que el algoritmo de recomendación tiene un desempeño aceptable, pero aún hay margen para mejorar la precisión de las predicciones, especialmente si se desean minimizar los errores cuadráticos y absolutos.

Conclusión

En conclusión, los sistemas de recomendación avanzados representan una evolución significativa sobre los enfoques más simples, ofreciendo una serie de ventajas clave. Su capacidad para analizar datos complejos y encontrar patrones sutiles permite recomendaciones más precisas y personalizadas, lo que mejora la experiencia del usuario.

Además, su escalabilidad, flexibilidad y capacidad de mejora continua los convierten en herramientas poderosas para empresas que buscan maximizar el valor de sus recomendaciones y adaptarse a un entorno digital en constante cambio.

En particular, la Descomposición en Valores Singulares (SVD) es una técnica popular en sistemas de recomendación avanzados debido a su capacidad para reducir la dimensionalidad de los datos y capturar relaciones latentes entre usuarios e ítems. En resumen, el SVD es una herramienta poderosa en el arsenal de los sistemas de recomendación avanzados, ofreciendo una manera eficaz de modelar la complejidad de las interacciones usuario-ítem y mejorar la calidad de las recomendaciones ofrecidas.

Enlace de ejercicio GitHub de esta actividad:

https://github.com/BigDataEquipo40/MNA-BigData-40/blob/main/Proyecto_avance_2/Avance_de_proyecto_2_Sistema_de_Recomendacio%CC%81n.ipynb

Enlace ejercicio complementario:
https://github.com/BigDataEquipo40/MNA-BigData-40/blob/main/Proyecto_avance_2/Avance_de_proyecto_2_Actividad_Complementaria.ipynb

Referencias

- Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. Geoscientific Model Development, 7(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

- Ng, W. S., & Tan, W. W. (2021). Some properties of various types of matrix factorization. ITM Web of Conferences, 36, 03003. https://doi.org/10.1051/itmconf/20213603003

- Nguyen, A. (2016). SINGULAR VALUE DECOMPOSITIONS IN RECOMMENDER SYSTEMS. Departmental Honors in the Department of Mathematics Texas Christian University. https://repository.tcu.edu/bitstream/handle/116099117/11320/Nguyen__Anh-Honors_Project.pdf?sequence=1&isAllowed=y

- Oracle® Fusion Cloud EPM Trabajo con Planning. (n.d.). https://docs.oracle.com/cloud/help/es/pbcs_common/PFUSU/insights_metrics_RMSE.htm#PFUSU-GUID-FD9381A1-81E1-4F6D-8EC4-82A6CE2A6E74

- Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. (2002). Incremental Singular Value Decomposition Algorithms for Highly Scalable Recommender

Systems. Fifth International Conference on Computer and Information Science.