

Modelo de Aprendizaje para Recomendación de Películas Mediante una Red Neuronal

Redes Neuronales con Embedding

1st Jonnathan Oswaldo Matute Curillo
Universidad
Politécnica Salesiana
Ecuador, Cuenca
jmatutec2@est.ups.edu.ec

2nd Edison Javier Yunga Tacuri
Universidad
Politécnica Salesiana
Ecuador, Cuenca
eyungat@est.ups.edu.ec

Abstract—Actualmente los sistemas de recomendaciones son herramientas que nos generan recomendaciones sobre un determinado objeto de estudio, a partir de las preferencias y opiniones dadas por los usuarios. El uso de estos sistemas se está poniendo cada vez más de moda en Internet debido a que son muy útiles para evaluar y filtrar la gran cantidad de información disponible en la Web con objeto de asistir a los usuarios en sus procesos de búsqueda y recuperación de información. La idea principal de este trabajo es presentar un estudio sobre la recomendación de películas utilizando la tecnología basada en Redes neuronales con Embedding mejorando la trama de entrenamiento y pérdida de validación utilizando el método mae y obteniendo el accuracy, lo que se pensó posteriormente es comparar con otro sistema de recomendación sin usar una red neuronal con el objetivo de ver cual es el mejor sistema de recomendación con el único fin de recomendar las mejores películas para un usuario, para ello se utiliza un conjunto de datos llamada MovieLens, el objetivo que se tiene como meta es dejar un precedente para futuros investigadores que retomen nuestro avance y con ello llegar a elaborar una aplicación para el servicio de todos.

Index Terms—Red Neuronal, sistema de recomendación, conjunto de datos, filtrado colaborativo.

I. INTRODUCCIÓN

En el mundo de hoy, cada usuario o cliente de las diferentes plataformas online se enfrenta a múltiples opciones de recomendación. Como es el caso de (Netflix) que me recomienda películas basado en lo que he visto anteriormente. Esto es lo que hacen los sistemas de recomendación y su poder está siendo aprovechado por la mayoría de las empresas en estos días. Desde Amazon hasta Netflix, desde Google hasta Spotify, los sistemas de recomendación son una de las aplicaciones más utilizadas de las técnicas de machine learning (aprendizaje automático).

En esta investigación abordaremos conceptos básicos para el entendimiento total del modelo propuesto, demostrando el filtrado colaborativo utilizando el conjunto de datos MovieLens para recomendar películas a los usuarios. El objetivo principal es predecir las clasificaciones de las películas que un usuario aún no ha visto y al final se recomienda películas a base del tiempo y se muestra 10 películas pronosticadas con

las calificaciones más altas las cuales también se recomienda al usuario.

A. Fundamentos de los sistemas de recomendación

Sistemas de recomendación: Los sistemas de recomendación basados en algoritmos de filtrado colaborativo utilizan las valoraciones de los usuarios sobre ciertos elementos del conjunto total para predecir valoraciones en el resto de los elementos y recomendar los de mayor valoración predicha. [1] [2] Los sistemas de recomendación son herramientas cuyo principal objetivo ayudar a los usuarios en la toma de decisiones para seleccionar un ítem que más se adecue a sus preferencias (gustos e intereses). [3] Un sistema de recomendación se puede definir, de manera formal, como aquel sistema que tiene como principal tarea seleccionar ciertos objetos de acuerdo con los requerimientos del usuario. Estos sistemas son muy atractivos en situaciones donde la cantidad de información que se ofrece al usuario supera ampliamente cualquier capacidad individual de exploración. [4]

B. Filtrado colaborativo neuronal.

El filtrado colaborativo neuronal: Se da al reemplazar el producto interno con una arquitectura neural que puede aprender una función arbitraria de los datos, presentamos un marco general llamado NCF, abreviatura de Filtrado colaborativo basado en redes neuronales. NCF es genérico y puede expresar y generalizar la factorización matricial en su marco. Extensos experimentos en dos conjuntos de datos del mundo real muestran mejoras significativas con NCF sobre los métodos de vanguardia. La evidencia muestra que el uso de capas más profundas de redes neuronales ofrece un mejor rendimiento de recomendación. [5]

C. Filtrado colaborativo.

Filtrado Colaborativo se da por ejemplo. Si a la persona A le gustan 3 películas, digamos Indiana Jones, Rambo y Duro de Matar, y a la persona B le gusta Indiana Jones, Rambo y Terminator, entonces tienen intereses casi similares. Podemos decir con cierta certeza que A debería gustar a Terminator y B

a Duro de Matar. El algoritmo de filtrado colaborativo utiliza “Comportamiento del usuario” para recomendar elementos.[5]

D. ¿Cómo funciona un motor de recomendación?

Los sistemas de recomendación, a veces llamados en inglés “recommender systems” son algoritmos que intentan “predecir” los siguientes ítems (productos, canciones, etc.) que querrá adquirir un usuario en particular.

E. Tipos de motores

Existen algunos tipos de motores de recomendación aquí unos ejemplos:

Popularity: Aconseja por la “popularidad” de los productos. Por ejemplo, “los más vendidos” globalmente, se ofrecerán a todos los usuarios por igual sin aprovechar la personalización. Es fácil de implementar y en algunos casos es efectiva.

Content-based: A partir de productos visitados por el usuario, se intenta “adivinar” qué busca el usuario y ofrecer mercancías similares.

Colaborative: Es el más novedoso, pues utiliza la información de “masas” para identificar perfiles similares y aprender de los datos para recomendar productos de manera individual.

F. Que es embeddings?

Hay otro tipo de recomendadores que almacenan un perfil para cada usuario en el que, de alguna manera, están contenidos sus gustos o preferencias [7]. Posteriormente, estos perfiles podrán combinarse para hacer recomendaciones personalizadas, con lo que se conseguirá una mayor satisfacción del usuario ante la recomendación. La forma en la que se calcule ese perfil podrá ser más o menos elaborada. Una manera de tener en cuenta ese perfil de consumo consiste en representar cada usuario con un vector que codifique la información que se dispone gracias al conjunto de datos movielens. Esto puede lograrse mediante, por ejemplo, la construcción de embeddings utilizando factorización de matrices [8].

En el presente trabajo esta estructurado de la siguiente manera: En la sección II se especifica el trabajo relacionado, en la sección III se presenta el método propuesto, en la sección IV se expone el diseño de experimentos partiendo de la descripción general del conjunto de datos, los parámetros de las técnicas utilizadas y las propiedades para la evaluación del método. En la sección V se presenta los resultados y el análisis correspondiente, y finalmente en la sección VI se puede encontrar las conclusiones y trabajo futuro de esta investigación.

II. TRABAJO RELACIONADO

Unos de los trabajos que más influyo en nuestra investigación es del autor Siddhartha Banerjee que realizo una investigación sobre la recomendación de películas usando conjunto de datos MovieLens, lo que pretende es poder predecir las clasificaciones de las películas que un usuario aún no ha visto para poder recomendarle al usuario. Emplean un mapeo

del ID del usuario a un “vector de usuario” a través de una matriz de incrustación, luego asigna un ID de película a un “vector de película” a través de una matriz de incrustación, después calcula el producto de puntos entre el vector del usuario y el vector de la película, para obtener una puntuación de coincidencia entre el usuario y la película (calificación prevista) y finalmente entrena las incrustaciones a través del descenso de gradiente utilizando todos los pares de películas de usuario conocidos.[6]

III. MÉTODO PROPUESTO

A. Conceptos previos

Los sistemas de recomendación son algoritmos que intentan “predecir” los siguientes ítems (productos, canciones, películas, imágenes, etc.) que querrá adquirir un usuario en particular, a su vez los embeddings son una manera de dar valoración útil a datos categóricos. Para ello asignaremos una profundidad a cada “identificador”, es decir un vector con valores continuos inicialmente aleatorios.

Con ello también se propone crear otro sistema de recomendación usando el filtrado colaborativo, donde se utilizara pivot con el objetivo de comparar cual es el mejor sistema de recomendación.

B. Diseño del método

A continuación vamos a presentar el primer diseño del método propuesto basado en un sistema que recomienda películas utilizando una red neuronal con embeddings, obteniendo el error medio de train t test y luego procediendo a graficar los modelos mae y del accuracy obtenido.

TABLE I
ALGORITMO DEL SISTEMA DE RECOMENDACIÓN

1: Cargamos el dataset de Movielens.
2: Visualizamos los datos del usuario ratings y películas.
3: Normalizamos los datos.
4: Dividimos los datos en train y test.
5: Realizamos un preprocesamiento para codificar usuarios y películas como índices enteros.
6: Creamos el modelo con 50 dimensiones.
7: Entrenamos el modelo.
8: Graficamos el modelo con diferentes epochs.
9: Calculamos el error medio.
10: Graficamos el modelo de mae.
11: Graficamos el modelo del accuracy
12: Mostramos las 10 mejores películas recomendadas para el usuario.

Presentamos el segundo diseño del método propuesto basado en filtrado colaborativo.

TABLE II
ALGORITMO DEL SISTEMA DE RECOMENDACIÓN

1: Cargamos el dataset de Movielens.
2: Visualizamos los datos del usuario ratings y películas.
3: Realizamos el preprocesamiento de datos.
4: Clasificamos las películas
5: Graficamos la clasificación
6: Realizamos la distribución de edad de los evaluadores
7: Creamos el Pivot Table
8: Mostramos las mejores películas con el motor de recomendación.

C. Diseño de Trama de entrenamiento y pérdida de validación

El diseño de trama de entrenamiento y pérdida de validación realizado por el Autor: Siddhartha Banerjee. [6]

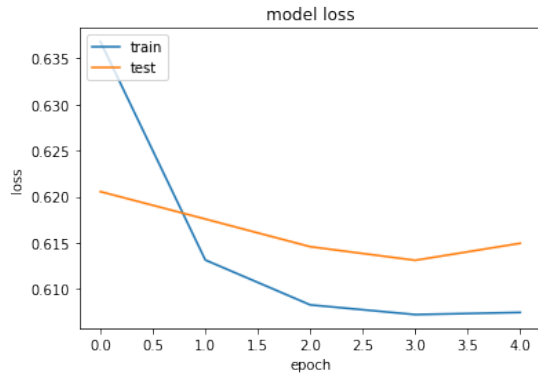


Fig. 1. Modelo Original

IV. DISEÑO DE EXPERIMENTOS

A. Experimentos establecidos

Para nuestra investigación utilizamos un dataset llamado MovieLens 100 k ml, en las siguientes tablas se puede observar el contenido del dataset.

TABLE III
PROPIEDADES DEL DATASET UTILIZAMOS EN LA INVESTIGACIÓN.

Dataset	MovieLens
Numero de usuarios	943
Numero de Ratings	100,000
Numero de peliculas	1,682
Valor del Rating	(1-5)

B. Medidas de calidad

La medida de calidad que se utilizo es el RMSE como Root Mean Squared Error: El error cuadrático medio (RMSE) es la desviación estándar de los residuos (errores de predicción). Los residuos son una medida de que tan lejos están los puntos de datos de la linea de regresión; RMSE es una medida de la dispersión de estos residuos. En otras palabras, le dice que tan concentrados están los datos alrededor de la línea de mejor ajuste. El error cuadrático medio se usa comúnmente en

climatología, predicción y análisis de regresión para verificar resultados experimentales. [1]

$$RMSE = \sqrt{\frac{\sum_{k=0}^{ltd} |R_{vi} - P_{vi}|^2}{ltdx}}$$

Se define de la siguiente manera:

- R_{vi} : Voto real del usuario a un item específico
- i : Item
- v : Voto real
- P_{vi} : Prediccion del voto del item
- ltd : Longitud de datos del rest

V. RESULTADOS

A. Resultados del error medio obtenido.

Para calcular el erro medio se tuvo que utilizar las métricas mae y obtener el accuracy teniendo en cuenta la cantidad de epochs.

TABLE IV
EVALUACIÓN CON MODELO DE EMBEDDING

Métrica	Error Medio
Train	0.16387768
Test	0.16690567

B. Modelo mejorado.

Para mejorar al modelo original se entreno la red neuronal con la métrica mae, al igual se opto por sacar el accuracy teniendo en cuenta la cantidad de epochs.

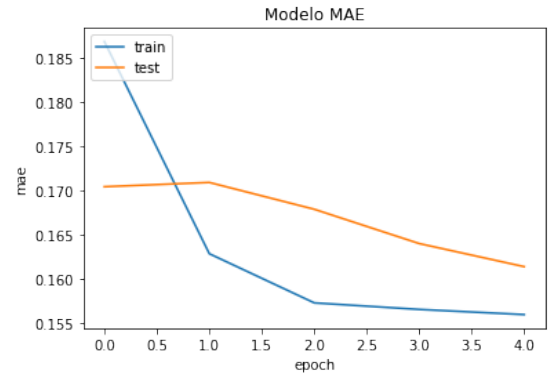


Fig. 2. Modelo mejorado mediante técnica mae

C. Gráfica del accuracy obtenido.

También se gráfico el accuracy obtenido al momento de entrenar teniendo en cuenta los epochs.

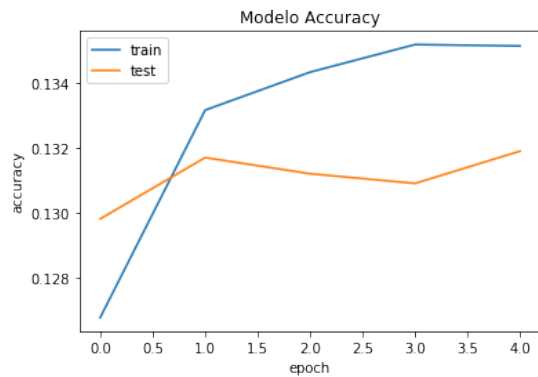


Fig. 3. Modelo del accuracy

D. Recomendación de Películas.

Finalmente mostramos las 10 mejores recomendaciones de películas a un usuario.

```
Mostrando recomendaciones para el usuario: 414
-----
Películas con altas calificaciones del usuario
-----
Dr. Strangelove or: How I Learned to Stop Worrying and Love the Bomb (1964) : Comedy|War
Some Like It Hot (1959) : Comedy|Crime
Wallace & Gromit: The Wrong Trousers (1993) : Animation|Children|Comedy|Crime
L.A. Confidential (1997) : Crime|Film-Noir|Mystery|Thriller
Rushmore (1998) : Comedy|Drama
-----
Las 10 mejores recomendaciones de películas
-----
Hoop Dreams (1994) : Documentary
Philadelphia Story, The (1940) : Comedy|Drama|Romance
Sunset Blvd. (a.k.a. Sunset Boulevard) (1950) : Drama|Film-Noir|Romance
Secrets & Lies (1996) : Drama
12 Angry Men (1957) : Drama
Third Man, The (1949) : Film-Noir|Mystery|Thriller
Psycho (1960) : Crime|Horror
Miller's Crossing (1990) : Crime|Drama|Film-Noir|Thriller
Touch of Evil (1958) : Crime|Film-Noir|Thriller
Shining, The (1980) : Horror
```

Fig. 4. Sistema de Recomendación

E. Gráfica de clasificación de películas con filtrado colaborativo.

Esta gráfica se realizó en el sistema de recomendación con filtrado colaborativo con el objetivo de realizar una distribución por el número de usuarios teniendo en cuenta el rating.

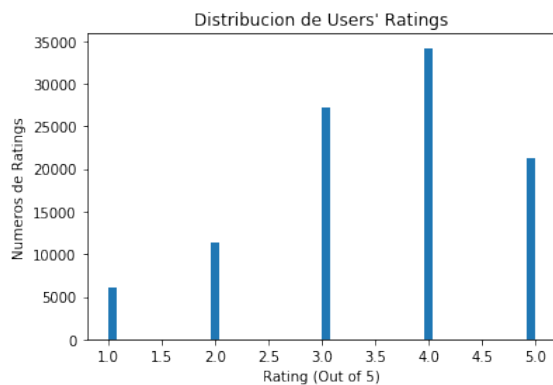


Fig. 5. Distribución de usuarios

F. Recomendación de películas con filtrado colaborativo.

Recommended movies based on your choice of Speed (1994) :				
movie_id		title	release_date	similarity
785	786	Bad Moon (1996)	01-Nov-1996	5.0
314	315	Apt Pupil (1998)	23-Oct-1998	5.0
152	153	Fish Called Wanda, A (1988)	01-Jan-1988	5.0
330	331	Edge, The (1997)	26-Sep-1997	5.0
344	345	Deconstructing Harry (1997)	01-Jan-1997	5.0
352	353	Deep Rising (1998)	30-Jan-1998	5.0
374	375	Showgirls (1995)	01-Jan-1995	5.0
379	380	Star Trek: Generations (1994)	01-Jan-1994	5.0
394	395	Robin Hood: Men in Tights (1993)	01-Jan-1993	5.0

Fig. 6. Recomendación con filtrado colaborativo

VI. CONCLUSIONES Y TRABAJO A FUTURO

Los modelos de recomendación cada vez son más populares debido a que aplicaciones como Netflix, Spotify, Google, han tomado una gran acogida por diferentes usuarios a nivel mundial. Por ello en este trabajo se implementó una red neuronal con Embedding llegando a recomendar 10 mejores películas a un usuario. También es importante recalcar que se encuentra el enlace del material adicional en la referencia [9], donde encontrarán el código que se elaboró para este trabajo.

Como trabajo futuro queremos llegar a elaborar una aplicación comercial con el objetivo de brindar un servicio para diferentes tipos de usuarios, recomendando según sus gustos y así no perder el tiempo buscando o navegando por internet.

REFERENCES

- [1] Nieto, S.M.G.: Filtrado colaborativo y sistemas de recomendación. Inteligencia en Redes de Comunicaciones. Madrid (2007).
- [2] Herlocker, J.L., Konstan, J.A., Terveen, L.G., Riedl, J.T.: Evaluating collaborative filtering recommender systems. ACM Transactions on Information Systems (TOIS) 22(1), 5-53 (2004).
- [3] Matos, V.L.: Sistemas de recomendación y personalización (2006).
- [4] Morales, C.A.R.: Algoritmo svd aplicado a los sistemas de recomendación en el comercio. Tecnología Investigación y Academia 6(1), 18-27 (2018).
- [5] JAYWRKR:Guía para construir un sistema de recomendación. Medium. (2009).
- [6] Xiangnan He, Lizi Liao, Hanwang Zhang, Liqiang Nie, Xia Hu y Tat-Seng Chua. 2017. Filtrado colaborativo neuronal. En Actas de la 26ª Conferencia Internacional sobre la World Wide Web (WWW '17). Comité Directivo de las Conferencias Internacionales de la World Wide Web, República y Cantón de Ginebra, CHE, 173-182. DOI: <https://doi.org/10.1145/3038912.3052569>.
- [7] J. Díez, D. Martínez-Rego, A. Alonso-Betanzos, O. Luaces, and A. Bahamonde, "Metrical representation of readers and articles in a digital newspaper," in 10th ACM Conference on Recommender Systems (RecSys 2016) Workshop on Profiling User Preferences for Dynamic Online and Real-Time Recommendations (RecProfile 2016), ACM, 2016.
- [8] O. Luaces, J. Díez, T. Joachims, and A. Bahamonde, "Mapping preferences into euclidean space," Expert Systems with Applications, vol. 42, no. 22, pp. 8588 – 8596, 2015.
- [9] Abad, C., Kam Paw Molina, A. R., Calva Paucar, L. E. (2010). Sistema de recomendación de películas.
- [10] Cortés, C. J. B., Vargas, S. G. J., Osorio, F. A. G., Gelbukh, A. (2015). Recomendación de productos a partir de perfiles de usuario interpretables. Tecnura, 19(45), 89-100.
- [11] Pérez, A. A., Close, J. G. (2015). Sistemas de recomendación. Robots Autónomos: Navegación.
- [12] Sanz, D. M. Valoración de Películas Basada en Análisis de Redes Sociales.
- [13] seo Ahn, K., Lin, J., Lu, M., Neath, L., Liang, J. Lecture 12: Face Recognition Dimensionality Reduction.

- [14] Rahutomo, R., Perbangsa, A. S., Soeparno, H., Pardamean, B. (2019, August). Embedding Model Design for Producing Book Recommendation. In 2019 International Conference on Information Management and Technology (ICIMTech) (Vol. 1, pp. 537-541). IEEE.
- [15] Qian, M. (2013). Prediction of the Type for Web Page: A Practical Application in Classification.
- [16] Nieto, S. M. (2007). Filtrado colaborativo y sistemas de recomendación. Inteligencia en Redes de Comunicaciones. Madrid.
- [17] Matich, D. J. (2001). Redes Neuronales: Conceptos básicos y aplicaciones. Universidad Tecnológica Nacional, México.
- [18] Pérez-Núñez, P., Luaces Rodríguez, Ó., Bahamonde Rionda, A., y Díez Peláez, J. (2018). Representaciones basadas en redes neuronales para tareas de recomendación. In XVIII Conferencia de la Asociación Española para la Inteligencia Artificial (CAEPIA 2018). I Workshop en Deep Learning (DEEPL 2018).
- [19] Díez, R. P., Gómez, A. G., y de Abajo Martínez, N. (2001). Introducción a la inteligencia artificial: sistemas expertos, redes neuronales artificiales y computación evolutiva. Universidad de Oviedo.
- [20] Tarazona, G. M., Chávez, J. S., Ferro, R. (2013). Modelación de sistemas de recomendación aplicando redes neuronales artificiales. Visión electrónica, 7(2), 45-56.
- [21] Enlace del material adicional
<https://github.com/JavierYungaT/SistemaRecomendacion>.