



Sistemas de recomendación

PARA UN MARKETPLACE

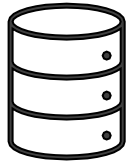
Integrantes:

- Ana María Urán González
- Karen Lizeth Velásquez Moná
- Nicolás Prieto Escobar
- Pablo Saldarriaga Aristizábal
- Victoria Álvarez Restrepo

Objetivo

Crear **sistemas de recomendación** en un contexto de **Big Data** que permitan sugerir productos afines a los intereses de los usuarios en un Marketplace.

Fuentes de datos



Amazon reviews

150 Millones de reviews

21 Millones de productos

33 Millones de clientes

Formato **TSV**

Tamaño **32 GB**

Info: usuario, product, rating

Fuente:

<https://s3.amazonaws.com/amazon-reviews-pds/readme.html>



Metadata

14.7 Millones de productos

Formato **CSV**

Tamaño **9.7 GB**

Info: título, precio, marca, etc

Fuente:

<http://jmcauley.ucsd.edu/data/amazon/>





MODELOS CON INFORMACIÓN HISTÓRICA



Modelos baseline



Modelos de factores latentes

















Modelos de factores latentes con sesgos

115.434 Productos

765.213
Usuarios

								
		5	?	?	3	?	5	?
		?	3	?	2	?	?	3
		3	?	?	?	4	4	?
		1	2	?	?	3	?	?
		?	5	5	?	5	4	2
		?	?	?	?	3	?	?
		5	?	?	?	?	?	?

							
	5	4	2	3	5	5	1
	5	3	4	2	4	3	3
	3	2	4	5	4	4	1
	1	2	5	4	3	2	3
	1	5	5	4	5	4	2
	1	3	4	5	3	5	1
	5	1	2	5	4	3	2

Modelo baseline

Modelo que emplean estadísticos básicos de los ratings para predecir las calificaciones restantes para cada producto.

	Mediana	Media	Media Producto
RMSE (validación)	1.309	1.115	1.056
Tiempo de ejecución (segundos)	180.58	30.51	31.17

Modelos de Factores Latentes

Modelo basado en la **factorización de matrices** con el método ALS (Alternating Least Squares).

Hiperparámetros			RMSE	Tiempo de ejecución (s)
Iteraciones	Factores latentes	Regularización		
10	10	0.5	1.121206	116.2984
10	15	0.5	1.121283	126.8312
10	5	0.5	1.121640	110.6444
10	5	0.1	1.143976	110.0862
10	15	0.1	1.159877	123.3760
10	10	0.1	1.164861	116.7004
10	5	0.01	1.278721	247.1880
10	10	0.01	1.707580	117.4129
10	15	0.01	2.126598	124.0611

Modelos de Factores Latentes (con sesgos)

Se consideran sesgos:

- ✓ Globales
- ✓ A nivel de usuario
- ✓ A nivel de producto

Hiperparámetros			RMSE	Tiempo de ejecución (s)
Iteraciones	Factores latentes	Regularización		
10	15	0.5	1.029867	136.0916
10	10	0.5	1.029874	129.5296
10	5	0.5	1.029883	120.3911
10	15	0.1	1.100043	139.3734
10	10	0.1	1.122282	125.8310
10	5	0.1	1.136316	117.7798
10	15	0.01	1.350831	134.6016
10	10	0.01	1.462684	127.4126
10	5	0.01	1.495363	121.7895

Selección del mejor modelo

	Baseline (Media Producto)	Factores latentes	Factores latentes con sesgos
RMSE (Validación)	1.056	1.121	1.029
Tiempo de ejecución (Segundos)	31.17	116.30	136.09
Hiperparámetros	-	Iteraciones: 10 Factores Latentes: 10 Regularización: 0.5	Iteraciones: 10 Factores Latentes: 15 Regularización: 0.5
RMSE (Test)			1.021

Recomendaciones generadas

+-----+-----+	
customer_id_integer	recommendations
+-----+-----+	
10206	[[108427, 0.03392...]
12566	[[108769, 0.05984...]
13406	[[115007, 0.00467...]
15062	[[109580, 0.01990...]
23123	[[115099, 0.00220...]
23619	[[114162, 0.00708...]
24252	[[112488, 9.75971...]
31264	[[115002, 0.00913...]
37004	[[102704, 0.00950...]
37827	[[110427, 0.00267...]
43548	[[108503, 0.01574...]
44185	[[115007, 0.00881...]
46318	[[115002, 0.01142...]
47067	[[91160, 0.002766...]
48077	[[112968, 0.04526...]
51528	[[115001, 0.06645...]
54110	[[86452, 0.003326...]
60683	[[110369, 0.00369...]
62320	[[111184, 0.00145...]
67630	[[115164, 0.03205...]
+-----+-----+	

-----> Creative Zen V Plus 2 GB Portable Media Player

-----> Summer Infant Deluxe Comfort Booster

-----> The Origin of Species: 150th Anniversary Edition

-----> RFA-188 3-Volt Lithium Battery

-----> Natural Male Enhancement Supplement - 745 MG Potent and High Quality Tablets

-----> Griffin Technology iTrip FM Transmitter for iPod

Modelos Cold Start

¿Qué es cold start?

¿Qué hacer en este caso?

Desarrollo del modelo

Fase

1

- **Entrada:** Título del producto
- **Proceso:** Conversión de textos a shingles (1-gramas).
- **Técnica:** MinHashLSH. 5 tablas de Hash (5 bandas) en el algoritmo.
- **Output:** 15 productos más similares al producto de interés.

Fase

2

- Cálculo de score de afinidad entre cada uno de los 15 artículos similares y el producto de interés.
- El score tiene en cuenta variables propias de los productos para ver qué tan afines son (precio, marca y categorías relacionadas).
- Se eligen los 5 productos con mayor valor de afinidad.

Validación

- Rating promedio de los 5 productos más afines al producto de interés como el rating predicho (oportunidad de mejora en este ítem).
- Tiempo de **ejecución 40 segundos** en hallar los productos más afines para un producto (es decir, en ejecutar ambas etapas).

Ejemplo



Título del producto: 'Girls pink tutu dress'
Precio: '\$19.99'
Marca: 'Hello Kitty'
Categorías relacionadas: 'Dance', 'Sports & Outdoors', 'Clothing','Clothing, Shoes & Jewelry'

1 Los 15 productos más similares al producto nuevo según MinHashLSH son los siguientes:

asin	price	title	brand	category
B00M66FYZC	\$22.00 - \$65.00	Little Girls Birthday Cupcake Pi...	Ella Blu Store	['Clothing, Shoes & Jewelry', 'G...
B00G3JICN4	\$29.89	SOPO Hello Kitty Toddler Dress B...	SoPo	['Clothing, Shoes & Jewelry', 'B...
B00QGM3EU	\$18.60 - \$34.99	Clementine Little Girls' Girls T...	Clementine Apparel	['Clothing, Shoes & Jewelry', 'G...
B006PFJMGW	\$29.99	Rare Editions Little Girls' Tutu...	Rare Editions	['Clothing, Shoes & Jewelry', 'G...
0000031909	\$6.94	Mystiqueshapes Girls Ballet Tutu...	Mystiqueshapes	['Sports & Outdoors', 'Sports & ...
0000031852	\$7.50	Girls Ballet Tutu Zebra Hot Pink	Tutu	['Sports & Outdoors', 'Sports & ...
B00SH78APA	\$10.13 - \$22.99	Hello Kitty Baby Girls' Tutu Dress	Hello Kitty	['Clothing, Shoes & Jewelry', 'N...
B0053WLHEI	\$19.95	Rare Editions Little Girls' Tutu...	Rare Editions	['Clothing, Shoes & Jewelry', 'G...
B00DDVYABS	\$14.99	The Queens Treasures Pink Tutu D...	The Queens Treasures	['Toys & Games', 'Dolls & Access...
B00P835VGC	\$25.99 - \$28.00	Bloch Girls' Toddler Tiffany Dre...	Bloch	['Clothing, Shoes & Jewelry', 'G...
B00VQEW9WG	\$13.99	Coral Pink Rosettes PETS Tutu Pa...	Kirei Sui	['Pet Supplies', 'Dogs', 'Appare...
B003JL8L NK	\$34.99	Rare Editions Little Girls' Tutu...	Rare Editions	['Clothing, Shoes & Jewelry', 'G...
B0072D0GPC	\$9.99	Pink Angel Little Girls' Stripe ...	Pink Angel	['Clothing, Shoes & Jewelry', 'G...
B00DNKCG38	\$29.99	Classykidzshop DS1708 - Pink/Fus...	Classykidzshop	['Clothing, Shoes & Jewelry', 'G...
B00OM1VVTM	\$12.50	Pink Snowflake Tutu Dress-18&quo...	American Fashion World	['Toys & Games', 'Dolls & Access...

Ejemplo

2 Los 5 productos con mejor score entre los 15 más similares son los siguientes:

asin	price	title	brand	category
B00QGMY3EU	\$18.60 - \$34.99	Clementine Little Girls' Girls Ta...	Clementine Apparel	['Clothing, Shoes & Jewelry', 'Gi...
B006PFJMGW	\$29.99	Rare Editions Little Girls' Tutu ...	Rare Editions	['Clothing, Shoes & Jewelry', 'Gi...
B00SH78APA	\$10.13 - \$22.99	Hello Kitty Baby Girls' Tutu Dress	Hello Kitty	['Clothing, Shoes & Jewelry', 'No...
B0053WLHEI	\$19.95	Rare Editions Little Girls' Tutu ...	Rare Editions	['Clothing, Shoes & Jewelry', 'Gi...
B00P835VGC	\$25.99 - \$28.00	Bloch Girls' Toddler Tiffany Dres...	Bloch	['Clothing, Shoes & Jewelry', 'Gi...

A las personas que hayan comprado alguno de estos 5 productos, se les podría recomendar que compren también este producto nuevo.

Trabajo Futuro

1. Modelo baseline empleando la **calificación promedio por usuario**.
2. **Escalamiento** del modelo de factores latentes con sesgo definido teniendo en cuenta la capacidad máxima del clúster.

Conclusiones

1. La **reducción de la dimensionalidad** que realiza el modelo de factores latentes, **facilita el procesamiento y la paralelización** de grandes volúmenes de información en la construcción de un sistema de recomendación.
2. El **proceso de dos fases** que combina un MinHashLSH y el posterior cálculo de scores de afinidad es una buena alternativa para cuando se tenga que dar recomendaciones en presencia de productos nuevos.
3. En el contexto de Big Data, es necesario tener en cuenta las **limitaciones** de los recursos **computacionales** dado que se cuenta con una gran cantidad de información.

Bibliografía

Bianchi, Mattia, Cesaro, Federico, Ciceri, Filippo, Dagrada, Mattia, Gasparin, Alberto, Gratta-rola, Daniele, Inajjar, Ilyas, Metelli, Alberto Maria, & Cella, Leonardo. 2017. Content-based approaches for cold-start job recommendations. Pages 1–5 of: *Proceedings of the Recommender Systems Challenge 2017*.

Boratto, Ludovico, Fenu, Gianni, & Marras, Mirko. 2019. The effect of algorithmic bias on re-commender systems for massive open online courses. Pages 457–472 of: *European Conference on Information Retrieval*. Springer.

Burgess, Stephen, Sellitto Carmine Cox Carmen. 2011. Trust perceptions of online travel information by different content creators: Some social and legal implications. Page 221–235 of: *Information Systems Frontiers*. Springer.

Milano, Silvia, Floridi Luciano. 2020. *Recommender systems and their ethical challenges*. In: *AI SOCIETY*

Panigrahi, Sasmita, Lenka, Rakesh Ku, & Stitipragyan, Ananya. 2016. A Hybrid Distributed Collaborative Filtering Recommender Engine Using Apache Spark. Pages 1000–1006 of: *Procedia Computer Science*. Sciencedirect.

Smirnov A V., Shilov NG, Ponomarev A V. Kashevnik AM Parfenov VG. 2014. *Group context-aware recommendation systems*. Pages 325–334 of: *Scientific and Technical Information Processing*. Springer.