

Propuesta Proyecto IIC 3633 Sistemas Recomendadores

Paula Navarrete Campos
Astrid San Martín J.

Departamento Ciencias de la Computación
Facultad de Ingeniería
Pontificia Universidad Católica de Chile

24 septiembre 2018.

Outline

Problema y Motivación

Alternativas de solución

Set de datos

Bibliografía

Problema y Motivación

- Heurísticas como ayuda para la toma de decisiones organizacionales.
- Heurística:
 1. Enfoque de resolución/aprendizaje/descubrimiento
 2. Emplea un método práctico suficiente para alcanzar un objetivo inmediato
 3. Se basa en experiencias individuales y colectivas.

Problema y Motivación

- Managers deliberadamente desarrollan reglas simples e idiosincráticas para guiar procesos y toma de decisiones
- Están relacionadas positivamente con el desempeño de la empresa

→ Ventajas competitivas.

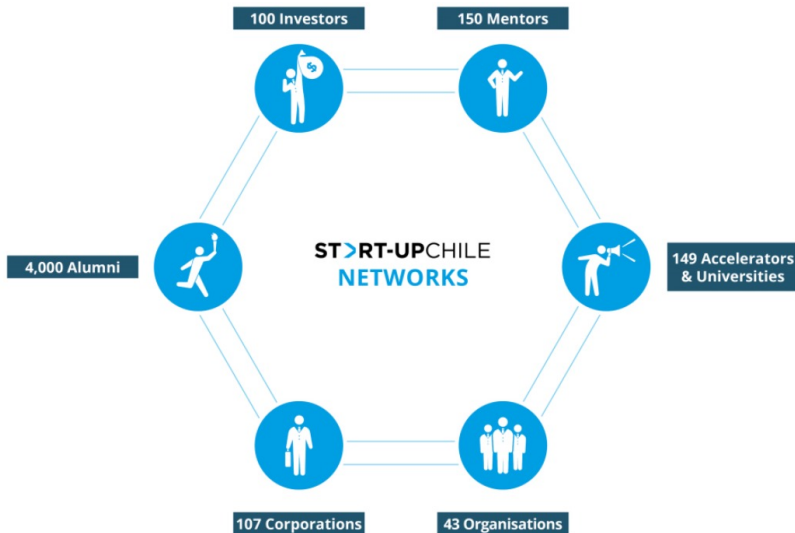
ST>RT-UP CHILE

The following information details the **global, economic and cultural impact** of Start-Up Chile since the beginning of the program in October 2010 to June 2016.

Global ecosystem impact



Problema y Motivación



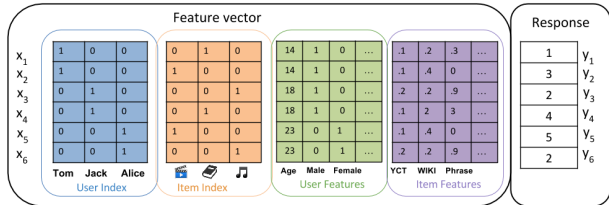
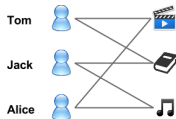
Problema y Motivación

$$f(\langle user, venture \rangle) \longrightarrow \{liked, disliked\}$$



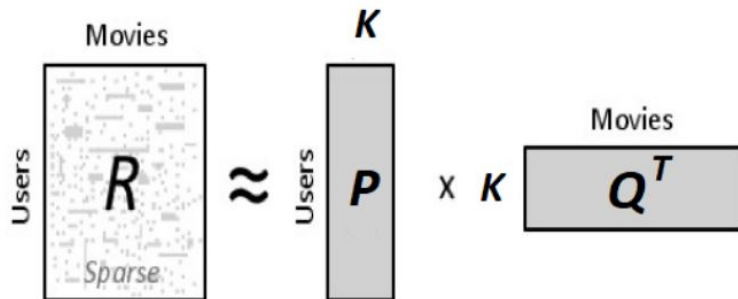
Alternativas de solución

Factorization Machine



Alternativas de solución

Factorization Machine



Set de datos

| | Judge | | | | Venture | | | | Other Ventures rated | | | | Last Venture rated | | | | class | |
|----------|----------|----------|----------|----------|----------|----------|----------|----------|----------------------|----------|----------|----------|--------------------|----------|----------|----------|----------------|---|
| x_1 | 1 | 0 | 0 | ... | 1 | 0 | 0 | ... | 0.3 | 0 | 0.3 | ... | 0 | 0 | 0 | ... | <i>dislike</i> | → |
| x_2 | 1 | 0 | 0 | ... | 0 | 0 | 1 | ... | 0.3 | 0 | 0.3 | ... | 1 | 0 | 0 | ... | <i>like</i> | → |
| x_3 | 0 | 1 | 0 | ... | 0 | 0 | 1 | ... | 0 | 0.5 | 0.5 | ... | 0 | 0 | 0 | ... | <i>dislike</i> | → |
| x_4 | 0 | 1 | 0 | ... | 0 | 1 | 0 | ... | 0 | 0.5 | 0.5 | ... | 0 | 0 | 0 | ... | <i>dislike</i> | → |
| x_5 | 0 | 0 | 1 | ... | 1 | 0 | 0 | ... | 0.5 | 0 | 0.5 | ... | 0 | 0 | 0 | ... | <i>like</i> | → |
| \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | \vdots | |
| x_n | 0 | 0 | 1 | ... | 0 | 0 | 1 | ... | 0.5 | 0 | 0.5 | ... | 0 | 0 | 1 | ... | <i>dislike</i> | → |
| | ↑ | ↑ | ↑ | | ↑ | ↑ | ↑ | | ↑ | ↑ | ↑ | | ↑ | ↑ | ↑ | | | |
| | j^A | j^B | j^C | ... | v^1 | v^2 | v^3 | ... | v^1 | v^2 | v^3 | ... | v^1 | v^2 | v^3 | | | |

Table: Data Structure

Referencias



Bingham, C. B., Eisenhardt, K. M. (2014).

Heuristics in strategy and organizations: Response to Vuori and Vuori.
Strategic Management Journal, 35(11), 1698-1702.



Vuori, N., Vuori, T. (2014).

Comment on “Heuristics in the strategy context” by Bingham and Eisenhardt (2011).
Strategic Management Journal, 35(11), 1689-1697.



Leatherbee, M., del Sol, P. (2016).

Predicting Entrepreneurial Performance: Simple Rules versus Expert Judgment. Working Paper.

Available at: <http://ctie.economia.cl/wp-content/uploads/2017/07/Predicting-Entrepreneurial-Performance-Simple-Rules-2016.pdf> (accessed May 2017).

Referencias



Bingham, C. B., Haleblan, J. (2012).

How firms learn heuristics: Uncovering missing components of organizational learning.

Strategic Entrepreneurship Journal, 6(2), 152-177.



Bingham, C. B., Eisenhardt, K. M. (2011).

Rational heuristics: the 'simple rules' that strategists learn from process experience.

Strategic management journal, 32(13), 1437-1464.



Eisenhardt, K. M., Sull, D. N. (2001).

Strategy as simple rules.

Harvard business review, 79(1), 106-119.

Referencias



Davis, J. P., Eisenhardt, K. M., Bingham, C. B. (2009).

Optimal structure, market dynamism, and the strategy of simple rules.

Administrative Science Quarterly, 54(3), 413-452.



Baum, J. A., Silverman, B. S. (2004).

Picking winners or building them? Alliance, intellectual, and human capital as selection criteria in venture financing and performance of biotechnology startups.

Journal of business venturing, 19(3), 411-436.

Referencias



Basu, C., Hirsh, H., Cohen, W. (1998, July).

Recommendation as classification: Using social and content-based information in recommendation.

In Aaai/iaai (pp. 714-720).



Rendle, S. (2010).

Factorization machines. In Data Mining (ICDM).

2010 IEEE 10th International Conference on (pp. 995-1000). IEEE.



Trofimov, M. Novikov, A. (2016).

tffm: TensorFlow implementation of an arbitrary order Factorization Machine.

GitHub, GitHub repository, [https://github.com/geffy/tffm]

¡Gracias!

Uso de Comentarios en Lenguaje Natural para Recomendaciones

— Proyecto - Sistema
Recomendadores - IIC3633 —

Alain Raymond - Rodrigo Rivera

Problemas a Resolver

- Aprovechar comentarios en lenguaje natural para potenciar predicciones y recomendaciones.
 - Difícil de resumir comentarios a diferencia de valores numéricos.
- Implementar técnicas del curso que no hemos ocupado:
 - LDA
 - BPR
 - FMs
- Solución debe poder responder en un tiempo razonable.
 - Mediremos mAP, nDCG y tiempo Respuesta.


Datasets

2 posibilidades:

- Dataset pequeño de reviews de bebidas alcohólicas.
 - 3.000 reviews aproximadamente.
 - Cientos de usuarios y cientos de ítems distintos.
- Dataset de Yelp!
 - Dominio: restaurantes.
 - Acotado a 1 ciudad (Phoenix), usuarios con más de 15 reviews y restaurantes de más de 15 reviews.
 - 117.346 reviews totales.
 - 3.498 usuarios únicos.
 - 2.191 restaurantes únicos.

Propuestas de Solución

- Sin Comentarios - Línea base FM con BPR
- LDA
- Sentiment Analysis
- doc2Vec
- Transnets - Red Profunda para predicción.



Transformación de
Comentarios a Features
de FM.

Referencias

- [1] Blei, D. (2003). **Latent Dirichlet Allocation**. 3(Jan):993-1022, 2003.
- [2] Rendle, S.(2012). **BPR: Bayesian Personalized Ranking from Implicit Feedback**. Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence (UAI2009).
- [3] Rendle, S. (2010). **Factorization machines**. In Data Mining (ICDM), 2010 IEEE 10th International Conference on (pp. 995-1000). IEEE.
- [4] Rendle, S. (2012). **Factorization machines with libfm**. ACM Transactions on Intelligent Systems and Technology (TIST), 3(3), 57.
- [5] Le, Q. (2014). **Distributed Representations of Sentences and Documents**.
- [6] Turney, P. (2002). **Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews**. Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics (ACL), Philadelphia, July 2002, pp.417-424.
- [7] Covington, P. (2016). **Deep Neural Networks for YouTube Recommendations**.
- [8] Catherine, R. (2017). **TransNets: Learning to Transform for Recommendation**. 11th ACM Conference on Recommender Systems (RecSys 2017).



Propuesta Panchos

Francisco Pérez Páez
Francisco Rencoret Domínguez



Contexto del Problema

- Organizar viajes es una tarea estresante
- Elección de hoteles.
- Empresas de reserva de hoteles masivos pueden ayudar en esta tarea.



Problema y Justificación

- Modelo de negocios de empresas como Booking y Expedia se basan en una buena recomendación de hotel.
- Elegir un hotel dentro de la gran cantidad de opciones que hay es muy difícil para el usuario.
- Mucho flujo de potenciales clientes que no pueden manejar de manera personalizada manualmente.



Datos

- *Dataset* provisto por Expedia a través de Kaggle.
- Datos contextuales (Ej: momento del día de la reserva, días, cantidad de pasajeros, si lo está realizando en un dispositivo móvil, etc.)
- Cluster de hoteles (Conjunto de hoteles agrupados según precio histórico, ubicación geográfica respecto del centro de la ciudad, etc.)



Objetivos

- Generar sistema recomendador basado en datos contextuales que prediga id del cluster de hoteles.
- Utilizar conocimientos de *Deep Learning* para superar rendimiento de métodos basados en memoria como User Knn.
- Rendimiento de recomendación sea comparable con resultados de otros equipos en Kaggle.



Solución Propuesta

- Dada información del contexto, se intentará recomendar priorizadamente los clusters de hoteles probables que el usuario reserve.
- Considerando la naturaleza secuencial de los datos, vamos a usar un modelo de Red Recurrente RNN.
- Posible iteración a modelos de atención.



Descripción de experimentos

- Iterar sobre la estructura de la RNN.
- Optimización de hiper-parámetros de entrenamiento (LR, Optimizador, *EarlyStopping*).
- Evaluación según *Accuracy*, *Precision*, *Recall* y *F1-Score*.



Muchas gracias



Propuesta de Proyecto final

Recomendando Juegos en STEAM

José Guzmán y Germán Cheuque

Índice

- La Industria de los Videojuegos
- Por qué Videojuegos?
- Solución Propuesta
- Datasets
- Referencias



La Industria de los Videojuegos

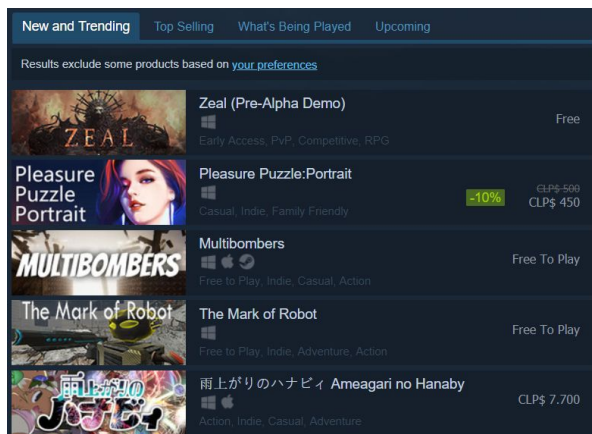
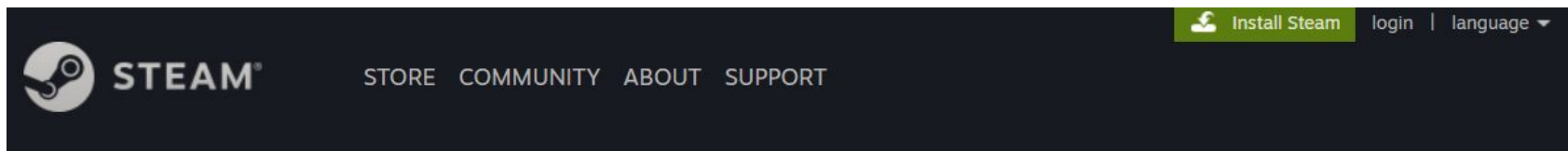


- Más de **2.3 billones de personas** dedican parte de su tiempo a jugar videojuegos (Newzoo. Global Games Market Report 2018)
- Crecimiento exponencial de la Industria en los últimos años
 - En 2017 la industria de los videojuegos creció un 10.7% → \$116 billones de dólares de ganancia
 - Para Q3 de 2018, ya se ha superado al año anterior con \$138 billones de dólares de ganancia

La Industria de los Videojuegos

- Diversificación de la Industria: modos de juego, logros y redes sociales
- **STEAM** es una de las plataformas de distribución de videojuegos más grandes de la actualidad + Interacción Social





10 millones de usuarios ingresan a los servidores de STEAM cada hora

Por qué Videojuegos?

1

Crecimiento y rentabilidad →
Excelente Industria para desarrollar
nuevas soluciones

2

Miles de opciones para escoger →
Necesidad de sistemas
recomendadores eficientes

3

Crecimiento exponencial de la **oferta** de
videojuegos en los últimos 5 años (7,672
en el 2017) → Necesidad de **mostrar los**
nuevos desarrollos

4

STEAM nos ofrece una comunidad
donde los usuarios e ítemes se
relacionan de múltiples formas
(**compras, reviews, recomendaciones**)





Solución propuesta 1

Sistema basado en Collaborative Filtering: ALS

Modelo simple que considerará un número limitado de features en la construcción de las variables latentes.



Solución propuesta 2

Sistema basado en Factorization Machines (FM)

Tomará en consideración un número importante de features de Usuarios e Items para la construcción de las variables latentes.



Solución propuesta 3

Sistema basado en FM + modelo de Deep Learning

Se incluirá el análisis de texto de diferentes reviews y opiniones para extraer sentimiento que será usado como una dimensión más del análisis de recomendación.



Dataset 1: Datos de consumo y preferencias (reviews) de usuarios australianos de STEAM sobre un total de 5,153,209 registros de compra.

| Nombre | Tipo de Dato | # de registros diferentes |
|---------|---------------|---------------------------|
| UserID | <i>int</i> | 8,8310 |
| ItemID | <i>int</i> | 10,978 |
| Reviews | <i>string</i> | 59,305 |
| Bundles | <i>array</i> | 615 |



Dataset 2: Datos de características de los videojuegos de la plataforma STEAM.

| Nombre | Tipo de Dato | Descripción |
|-----------------------|----------------|----------------------------|
| Metacritic | <i>int</i> | Puntaje de Metacritic |
| RecommendationCount | <i>int</i> | # de recomendaciones |
| CategoryMultiplayer | <i>Boolean</i> | Con modo Multijugador |
| CategoryCoop | <i>Boolean</i> | Con modo Cooperativo |
| CategoryInAppPurchase | <i>Boolean</i> | Capacidad de compras InApp |

Referencias



- 01 | Bertens P., Guitart A., Chen P. P., Periañez A. (2018). A Machine-Learning Item Recommendation System for Video Games
- 02 | Sifa R., Bauckhage C., Drachen A. (2014) Archetypal Game Recommender Systems.
- 03 | Global Games Market Revenues 2018 | Per Region Segment. (n.d.). Retrieved from <https://newzoo.com/insights/articles/global-games-market-reaches-137-9-billion-in-2018-mobilegames-take-half/>
- 04 | Takahashi, D. (2017, November 29). Newzoo: Game industry growing faster than expected, up 10.7% to \$116 billion 2017. Retrieved from <https://venturebeat.com/2017/11/28/newzoo-game-industrygrowing-faster-than-expected-up-10-7-to-116-billion-2017/>
- 05 | Orland, K. (2014, April 16). Introducing Steam Gauge: Ars reveals Steam's most popular games. Retrieved from <https://arstechnica.com/gaming/2014/04/introducing-steam-gauge-ars-reveals-steamsmost-popular-games/>
- 06 | Ell, K. (2018, July 20). Video game industry is booming with continued revenue. Retrieved from <https://www.cnn.com/2018/07/18/video-game-industry-is-booming-with-continued-revenue.html>
- 07 | Castells P., Vargas S., Wang J. (2012). Novelty and Diversity Metrics for Recommender Systems: Choice, Discovery and Relevance.



Gracias

Feedback implícito en e-commerce

Carlos Álvarez - Iván Wolf



Contexto

- Grandes cantidades de ventas implican grandes cantidades de logs.
- El desafío es recomendar a usuario de un e-commerce real artículos que les puedan gustar.

Problema

Hay que considerar que cuando trabajamos con feedback implícito:

- No hay feedback negativo
- El feedback implícito es inherentemente ruidoso
- El valor numérico en el feedback explícito refleja preferencia por un producto mientras que en el caso implícito refleja confianza
- La evaluación de un sistema basado en feedback implícito requiere formas de medir adecuadas

Objetivos

- Implementar un sistema recomendador basado en feedback implícito
- Evaluar y comparar calidad de recomendaciones del sistema implementado con otros modelos

Solución

- Implementar el método Alternating Least Squares usando la librería *tensorflow*.
- Adaptar el algoritmo para incluir tanto el feedback de las visitas, adiciones al carro de compras y transacciones.
- Utilizar las métricas que usan en el artículo fuente para evaluar el modelo sobre el dataset

Dataset

Eventos

| | timestamp | visitorid | event | itemid | transactionid |
|---|---------------|-----------|-------|--------|---------------|
| 1 | 1433221332117 | 257597 | view | 355908 | |
| 2 | 1433224214164 | 992329 | view | 248676 | |

Categorías y familias de productos

| | categoryid | parentid |
|---|------------|----------|
| 1 | 1016 | 213 |
| 2 | 809 | 169 |

Propiedades de productos en el tiempo

| | timestamp | itemid | property | value |
|---|---------------|--------|------------|----------------------------|
| 1 | 1435460400000 | 460429 | categoryid | 1338 |
| 2 | 1441508400000 | 206783 | 888 | 1116713 960601 n277.200 |

Experimentos a realizar

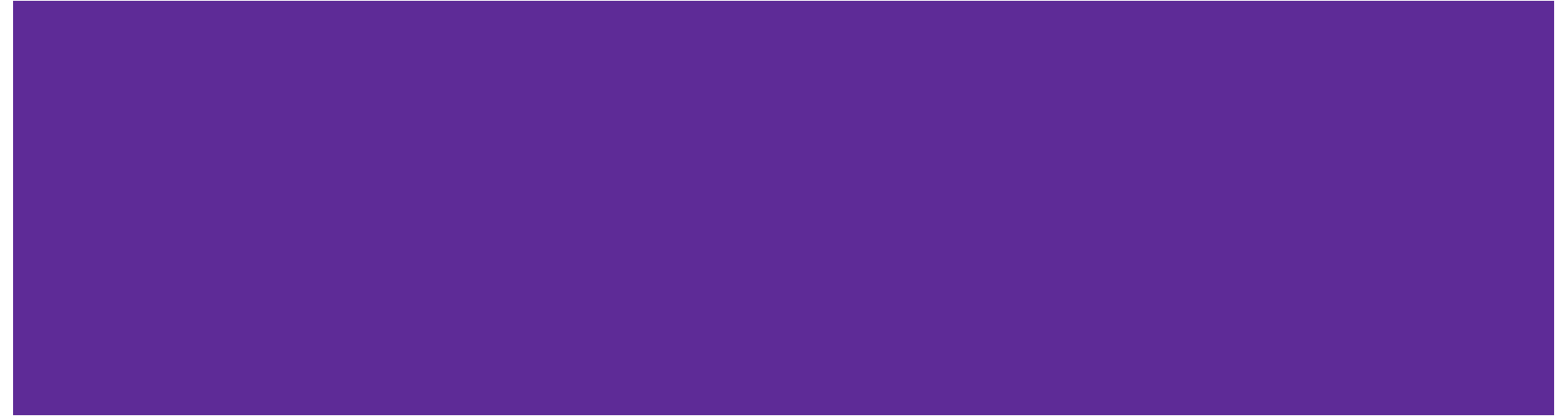
- Implementar algoritmos básicos de sistemas recomendadores como benchmark para nuestro modelo
- Implementar el algoritmo descrito en el paper ‘Collaborative Filtering for Implicit Feedback Datasets’ utilizando datos de logs de vistas de los productos

$$\min_{x_*, y_*} \sum_{u,i} c_{ui} (p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

- Adaptar el modelo a datos del e-commerce utilizando todos los datos disponibles y elegir la mejor metodología que otorgue un mejor rendimiento

Feedback implícito en e-commerce

Carlos Álvarez - Iván Wolf





Implementación de un Sistema Híbrido tipo cascade usando Collaborative Filtering y Content-based Filtering

IIC3633 – Sistemas Recomendadores

Víctor Gálvez Yanjarí

Camilo Ruiz-Tagle Molina

23 de Septiembre de 2018.

1. Problema y Justificación

1. Problema y Justificación

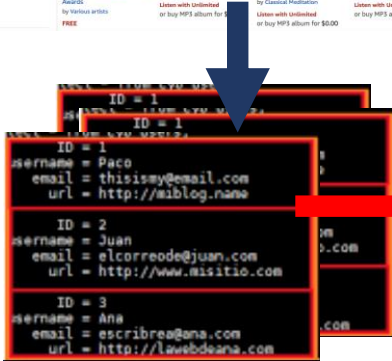
- Para el problema de recomendación de música, es posible contar con información clave (**ratings y reviews**) para el desarrollo de un sistema recomendador.
- Características como el título, álbum, categoría, etc. están disponibles como **metadata**.
- La **mezcla** de ambos inputs puede generar una buena recomendación.

1. Problema y Justificación

- Dado el tipo de input, los sistemas híbridos ocupan **content based (reviews) combinados con sistemas colaborativos**.
- En la literatura, existen sistemas del tipo cascade en la forma content based – colaborativo.
- **No hay evidencia** de sistemas tipo cascade en la forma colaborativo – content based (Burke, 2002).

2. Solución propuesta

2. Solución propuesta

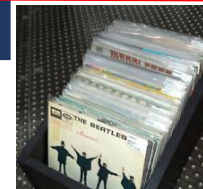


{JSON}

64.706 Registros de
Amazon Music
Review

ALGORITMO
COLABORATIVO

```
19 unsigned int len1 = s1.size(), prevCol[1];
20 const size_t len2 = col[len1+1];
21 vector<unsigned int> col(len2+1, 0);
22 for (unsigned int i = 0; i < len1; i++) {
23     prevCol[i] = i;
24     for (unsigned int j = 0; j < len2; j++)
25         col[j] = i+1;
26     for (unsigned int j = 0; j < len2; j++)
27         col[j+1] = std::min(std::min(prevCol[j] + j + 1, col[j]) +
28                             prevCol[j] + (s1[i]==s2[j]) ? 0 : 1);
29     col.swap(prevCol);
30 }
31 return prevCol[len2];
32 }
```



Subconjunto ítems
recomendados

ALGORITMO
CONTENT-BASED

```
19 unsigned int len1 = s1.size(), prevCol[1];
20 const size_t len2 = col[len1+1];
21 vector<unsigned int> col(len2+1, 0);
22 for (unsigned int i = 0; i < len1; i++) {
23     prevCol[i] = i;
24     for (unsigned int j = 0; j < len2; j++)
25         col[j] = i+1;
26     for (unsigned int j = 0; j < len2; j++)
27         col[j+1] = std::min(std::min(prevCol[j] + j + 1, col[j]) +
28                             prevCol[j] + (s1[i]==s2[j]) ? 0 : 1);
29     col.swap(prevCol);
30 }
31 return prevCol[len2];
32 }
```



Subconjunto ítems
recomendados

2. Solución propuesta

64.706 Registros de Amazon Music Review

- ID de canción.
- Rating.
- ID del reviewer.
- Nombre del reviewer.
- Texto del review.
- Summary del review.
- Rating de utilidad de la review (ej: 2/3)
- Timestamp del review.
- Timestamp del review (unix)

3. Objetivos

3. Objetivos

1. Análisis exploratorio de los datos:
 - Conocer **distribución** de los datos.
 - Identificar registros necesarios que garantice un **adecuado nivel de entrenamiento y test**.
2. Determinar **atributos** relevantes.
3. Determinar **baselines** para comparar los resultados.
4. Determinar forma de **inputs para segundo algoritmo** (words embedding, número de ítems recomendados, etc.).

3. Objetivos

5. De ser necesario, construir una **métrica** que permita:
 - Comparar los resultados.
 - Obtener una medida de error (content based).
6. De ser necesario, usar **otros datasets** similares o complementarios.
7. Seleccionar el **mejor algoritmo** colaborativo y content-based.

4. Descripción de experimentos a realizar

4. Descripción de experimentos a realizar

Entrenamiento
y medición
variables de
entorno colab.



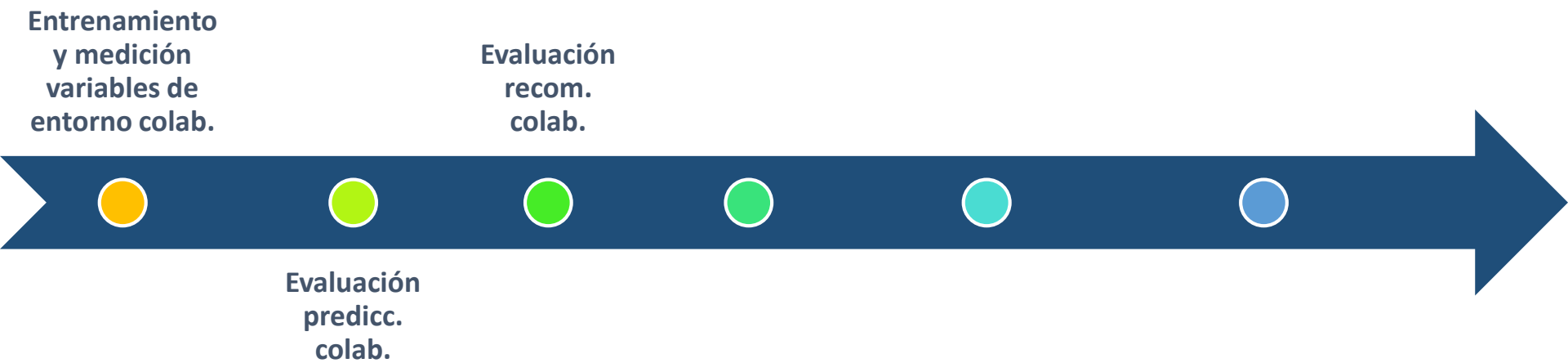
4. Descripción de experimentos a realizar

Entrenamiento
y medición
variables de
entorno colab.

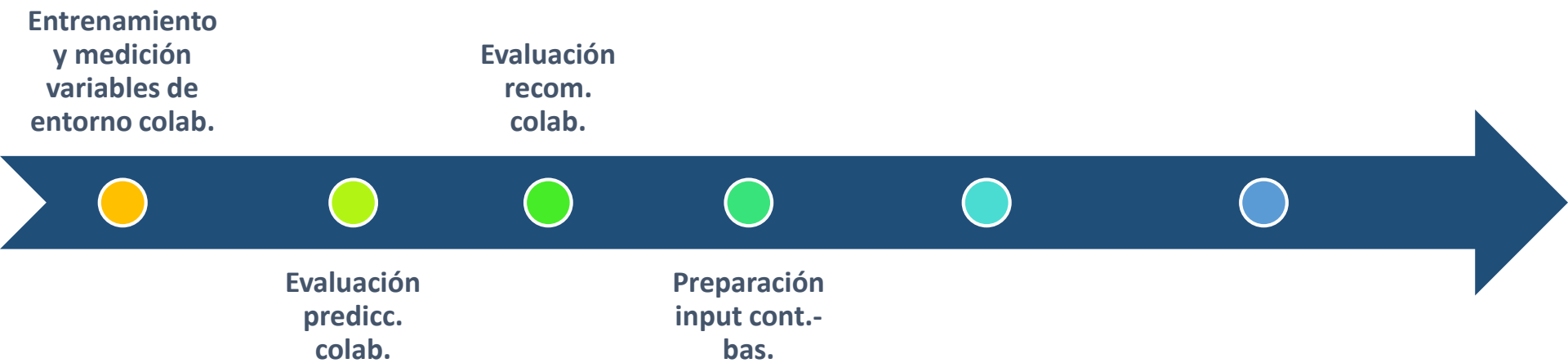


Evaluación
predicc.
colab.

4. Descripción de experimentos a realizar



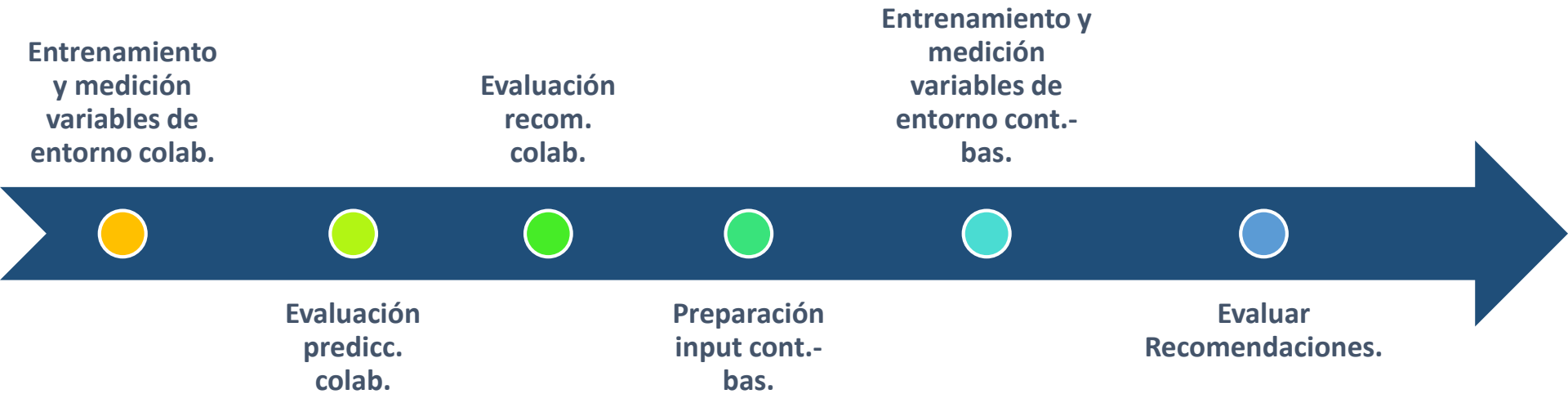
4. Descripción de experimentos a realizar



4. Descripción de experimentos a realizar

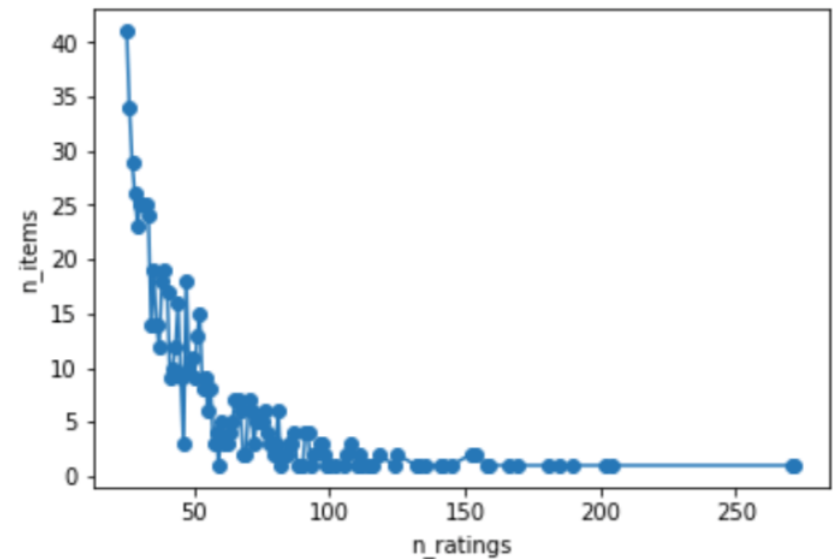
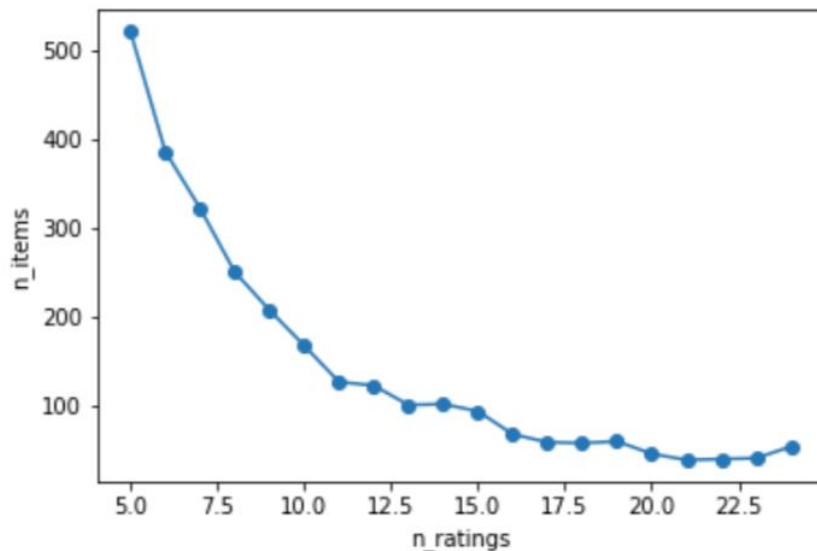


4. Descripción de experimentos a realizar



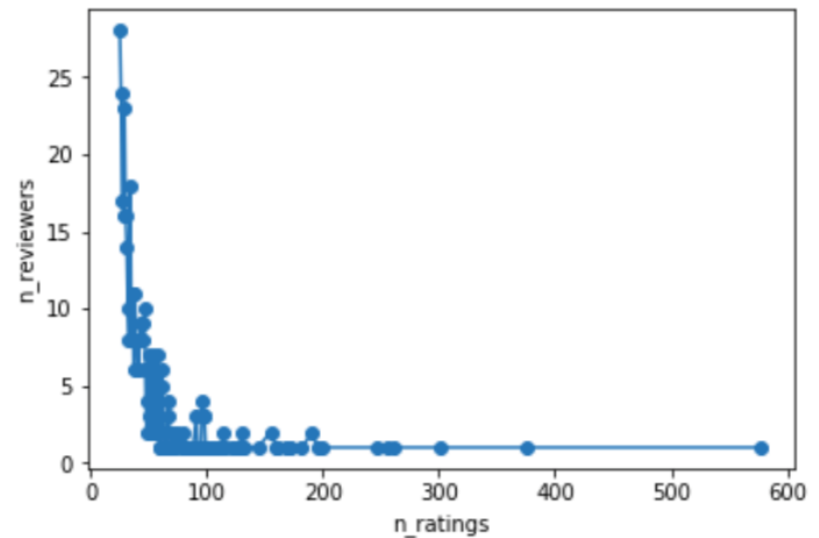
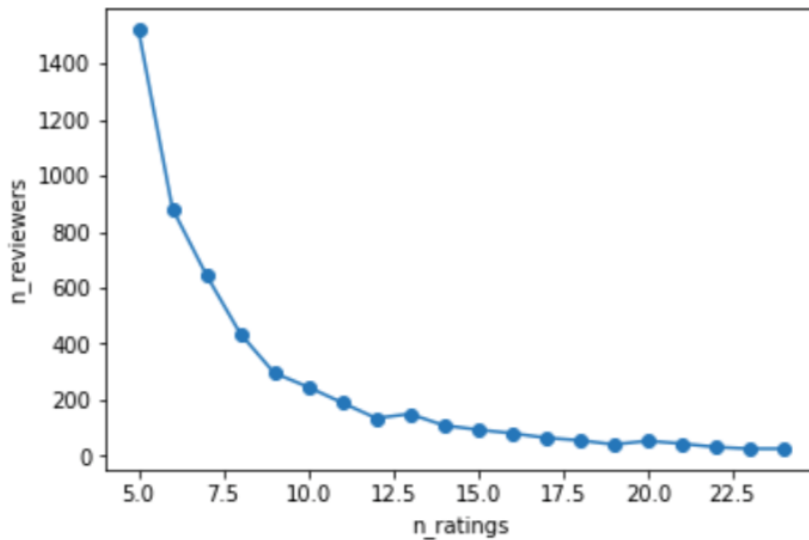
5. Exploración de los datos del experimento

5. Exploración de los datos del experimento



- Más de 500 ítems tienen 5 ratings.
- N° de ítems disminuye a medida que aumenta el número de ratings.

5. Exploración de los datos del experimento



- Son más los usuarios que evalúan poco que los que evalúan más.
- 5 es el mínimo de ratings por ítems y por usuario.

6. ¿Preguntas?

7. Bibliografía

Robin Burke (2002). *Hybrid Recommender Systems: Survey and Experiments*. En: User Modeling and User-Adapted Interaction 12, 4 (November 2002), pp. 331-370. Disponible en <http://dx.doi.org/10.1023/A:1021240730564>

Julian McAuley(2018). *Amazon Product Data*. University of California, San Diego. Disponible en: <http://jmcauley.ucsd.edu/data/amazon/>

Aristomenis Lampropoulos, Paraskevi Lampropoulou George A. Tsihrintzi (2012). *A Cascade-Hybrid Music Recommender System for mobile services based on musical genre classification and personality diagnosis*. Disponible en: <https://link.springer.com/article/10.1007/s11042-011-0742-0>

Aristomenis Lampropoulos, Dionysios Sotiropoulos George A. Tsihrintzis (2013).*Evaluation of a Cascade Hybrid Recommendation as a Combination of One-Class Classification and Collaborative Filtering*. Disponible en: <https://ieeexplore.ieee.org/abstract/document/6495108/>



SmallChaos

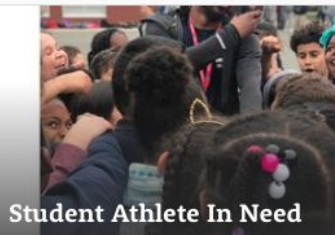
Cristóbal Eyzaguirre
Rolf Skog



Donors Choose

Most urgent projects

HIGHEST ECONOMIC NEED + CLOSEST TO FINISH LINE + FEWEST DAYS LEFT



Student Athlete In Need

My students need running and basketball shoes. A few student-athletes in hardship need athletic shoes.

Mr. Coach O

New Foundations Charter School K-



~~\$160~~ still needed

2X

\$80 for now, thanks to a match



DNA Base-Pairing

My students need beads and the hemp to complete a DNA base-pairing activity.

Ms. Ichiyama

Griffith Senior High School • Griffith, IN



\$57 still needed



We Can Make Beautiful Music Together!

Help me give my students big drums and fun rhythm instruments for music-making.

Mrs. Franks

Southwest Elementary School • Lexington,



\$49 still needed

Problema

- Muchos colegios necesitan donaciones
- Mucha gente dispuesta
- Juntar donantes con su proyecto deseado

Solución

Sistema recomendador:

- Collaborative topic regression
- Deep and Wide
- Next Item Recommendation with Self-Attention

Dataset

Datos de la página

- Donantes:
 - Estado
 - Zip
 - Profesión (profesor o no)
- Donaciones:
 - Monto
 - Fecha
- Proyecto:
 - Descripción
 - Recursos solicitados

Referencias

- Wang, H., Wang, N., and Yeung, D.-Y. Collaborative deep learning for recommender systems. In Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15, ACM (New York, NY, USA, 2015), 1235–1244
- Cheng, H. T., Koc, L., Harmsen, J., Shaked, T., Chandra, T., Aradhye, H., ... Anil, R. (2016, September). Wide deep learning for recommender systems. In Proceedings of the 1st Workshop on Deep Learning for Recommender Systems (pp. 7-10). ACM.
- Shuai Zhang, Yi Tay, Lina Yao, and Aixin Sun. 2018. Next Item Recommendation with Self-Attention. In Proceedings of Conference Submission, , Month 2018/9, 10 pages.



YOLO: You Only Listen Once

Propuesta Proyecto

Thomas Muñoz - Rodolfo Palma

24 de septiembre de 2018

El problema consiste en automáticamente continuar una lista de reproducción

Una **lista de reproducción** es un conjunto de canciones destinadas a ser escuchadas secuencialmente.

Es una variante del problema de generar listas de reproducción de manera automática.

El problema es importante para la academia y la industria. Fue abordado en el *ACM Recys Challenge 2018* y es de interés para servicios como Spotify, Pandora y Apple Music.

La tarea de recomendar música presenta ciertas características que no se presentan en otros dominios

- *Duración de ítems*: usualmente, una canción dura entre 3 y 5 minutos.

La tarea de recomendar música presenta ciertas características que no se presentan en otros dominios

- *Duración de ítems*: usualmente, una canción dura entre 3 y 5 minutos.
- *Cantidad de ítems*: los tamaños de los catálogos de canciones están en el orden de decenas de millones de ítems.

La tarea de recomendar música presenta ciertas características que no se presentan en otros dominios

- *Duración de ítems*: usualmente, una canción dura entre 3 y 5 minutos.
- *Cantidad de ítems*: los tamaños de los catálogos de canciones están en el orden de decenas de millones de ítems.
- *Más emocional*: la música es capaz de provocar emociones fuertes. Por otra parte, la emoción del usuario afecta el tipo de música que escucha.

La tarea de recomendar música presenta ciertas características que no se presentan en otros dominios

- *Duración de ítems*: usualmente, una canción dura entre 3 y 5 minutos.
- *Cantidad de ítems*: los tamaños de los catálogos de canciones están en el orden de decenas de millones de ítems.
- *Más emocional*: la música es capaz de provocar emociones fuertes. Por otra parte, la emoción del usuario afecta el tipo de música que escucha.
- *Contexto*: el contexto afecta fuertemente el tipo de música que se va a escuchar. Las canciones que un usuario escucha luego de terminar una larga relación sentimental son distintas a las que se escuchan en una previa antes de salir a *carretear*.

Se implementará un sistema recomendador simple y luego modelos más elaborados

En primera instancia se utilizará Adaptive KNN con criterio de optimización BPR para feedback implícito.

- *Item-based* KNN.
- Criterio BPR para usar *priors* debido a que los datos son *sparse*.
- Feedback implícito ya que no hay *ratings*, sólo información de interacción.

Se verá la posibilidad de usar la técnica de Máquinas de Factorización en lugar de Adaptive KNN.

El *dataset* está compuesto por 1.000.000 de listas de reproducción

Algunas características relevantes son:

- Toda lista de reproducción tiene un título compartido por un número significativo de otras listas de reproducción.

El *dataset* está compuesto por 1.000.000 de listas de reproducción

Algunas características relevantes son:

- Toda lista de reproducción tiene un título compartido por un número significativo de otras listas de reproducción.
- Las listas de reproducción contienen al menos 3 artistas únicos y 2 álbumes únicos.

El *dataset* está compuesto por 1.000.000 de listas de reproducción



Algunas características relevantes son:

- Toda lista de reproducción tiene un título compartido por un número significativo de otras listas de reproducción.
- Las listas de reproducción contienen al menos 3 artistas únicos y 2 álbumes únicos.
- Existe un mínimo global de cantidad de seguidores y escuchadores.

El *dataset* está compuesto por 1.000.000 de listas de reproducción

Algunas características relevantes son:

- Toda lista de reproducción tiene un título compartido por un número significativo de otras listas de reproducción.
- Las listas de reproducción contienen al menos 3 artistas únicos y 2 álbumes únicos.
- Existe un mínimo global de cantidad de seguidores y escuchadores.
- Las listas de reproducción tienen entre 5 y 250 canciones.

-  Rendle, S., Freudenthaler, C., Gantner, Z. & Schmidt-Thieme L. *BPR: Bayesian Personalized Ranking from Implicit Feedback*, 2009.
-  Schedl, M., Zamani, H., Chen, C., Deldjoo Y. & Elahi M. *Current Challenges and Visions in Music Recommender Systems Research*, 2018.



Propuesta proyecto final: Fake news

Florencia Barrios
Joaquín Tagle

Problemática



Donald J. Trump ✓

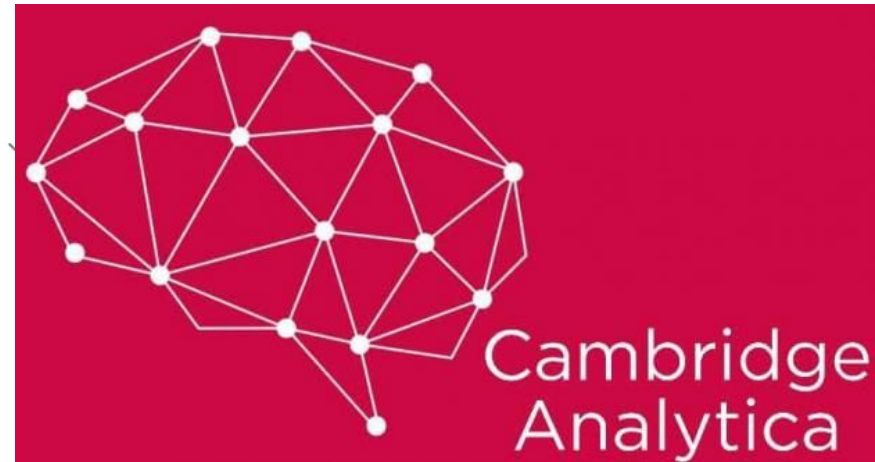
@realDonaldTrump

Seguir

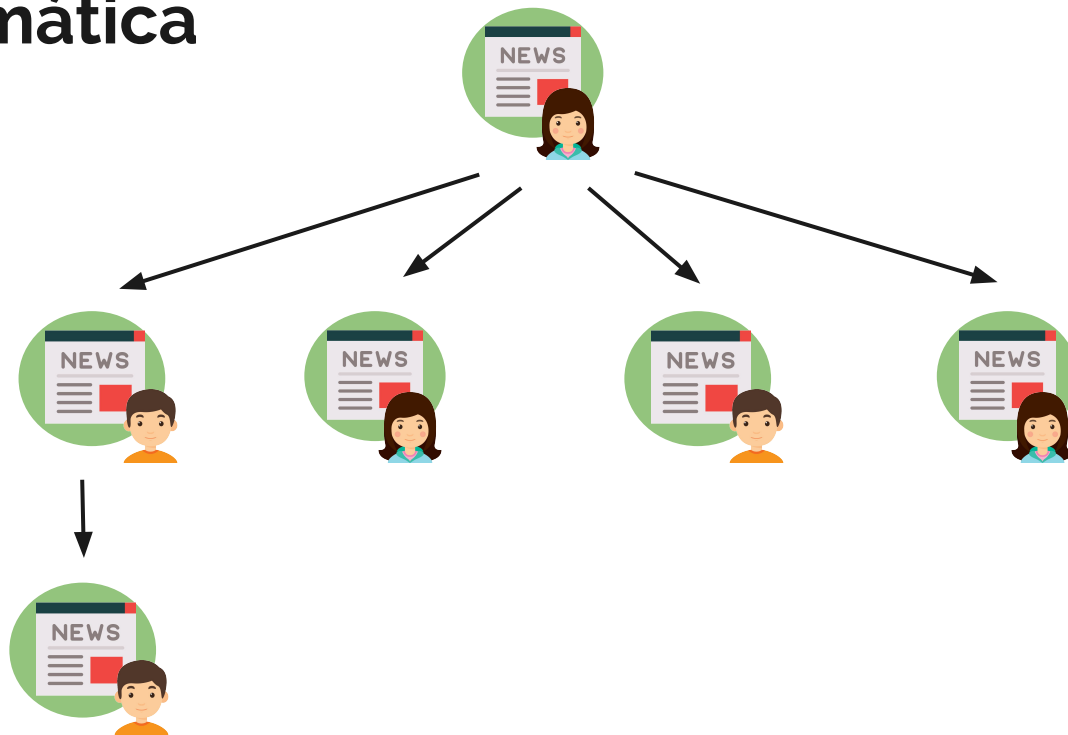
The greatest influence over our election was the Fake News Media "screaming" for Crooked Hillary Clinton. Next, she was a bad candidate!

Traducir Tweet

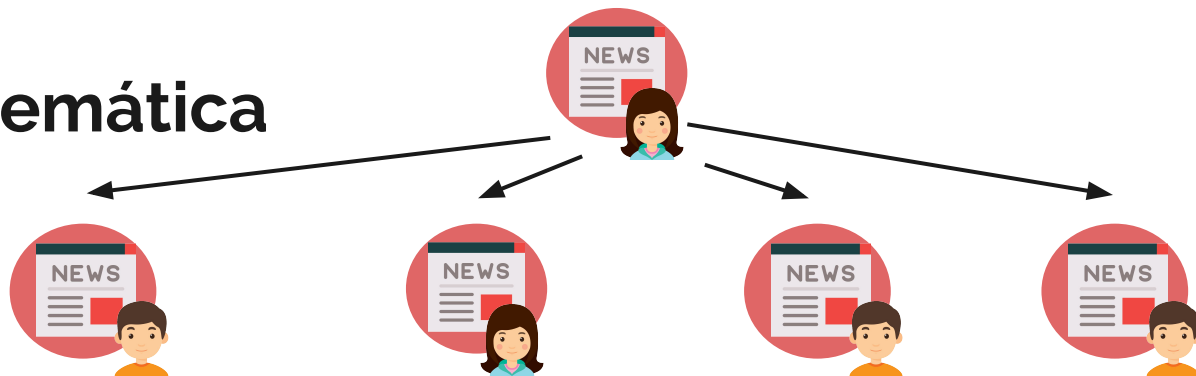
4:26 - 22 sept. 2017



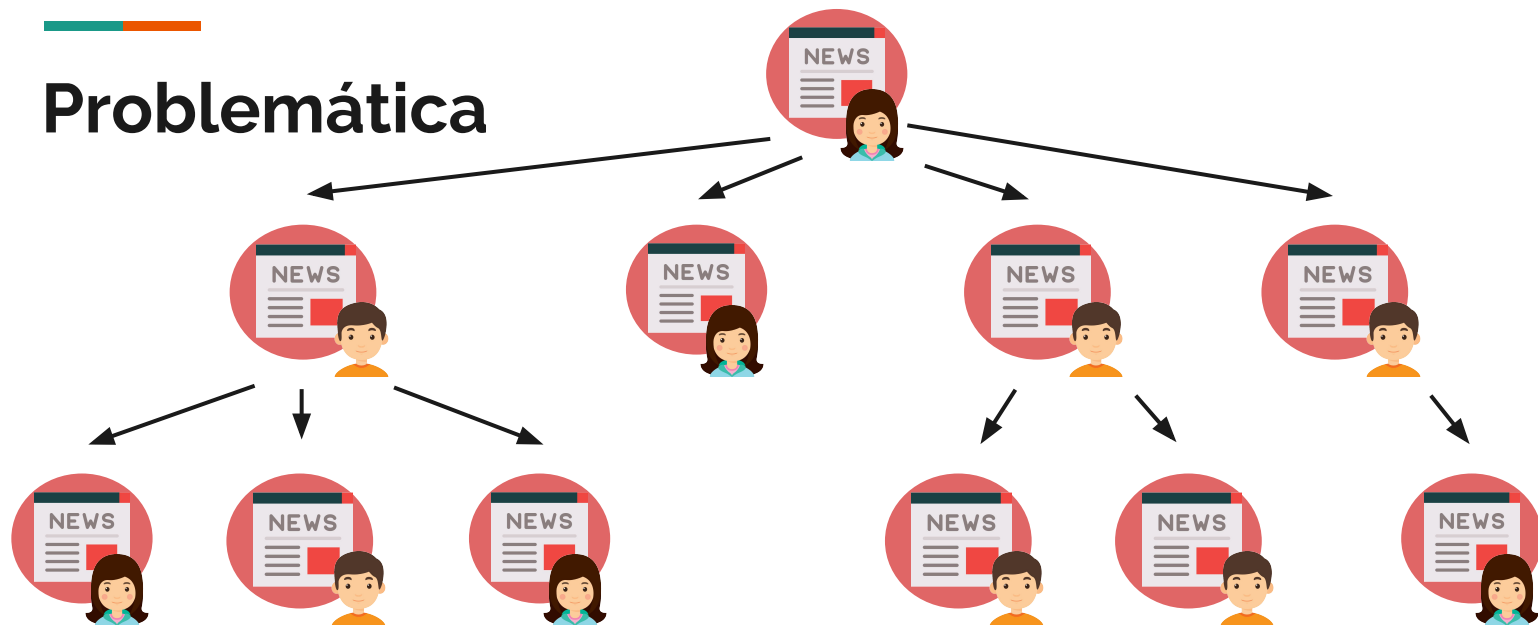
Problemática



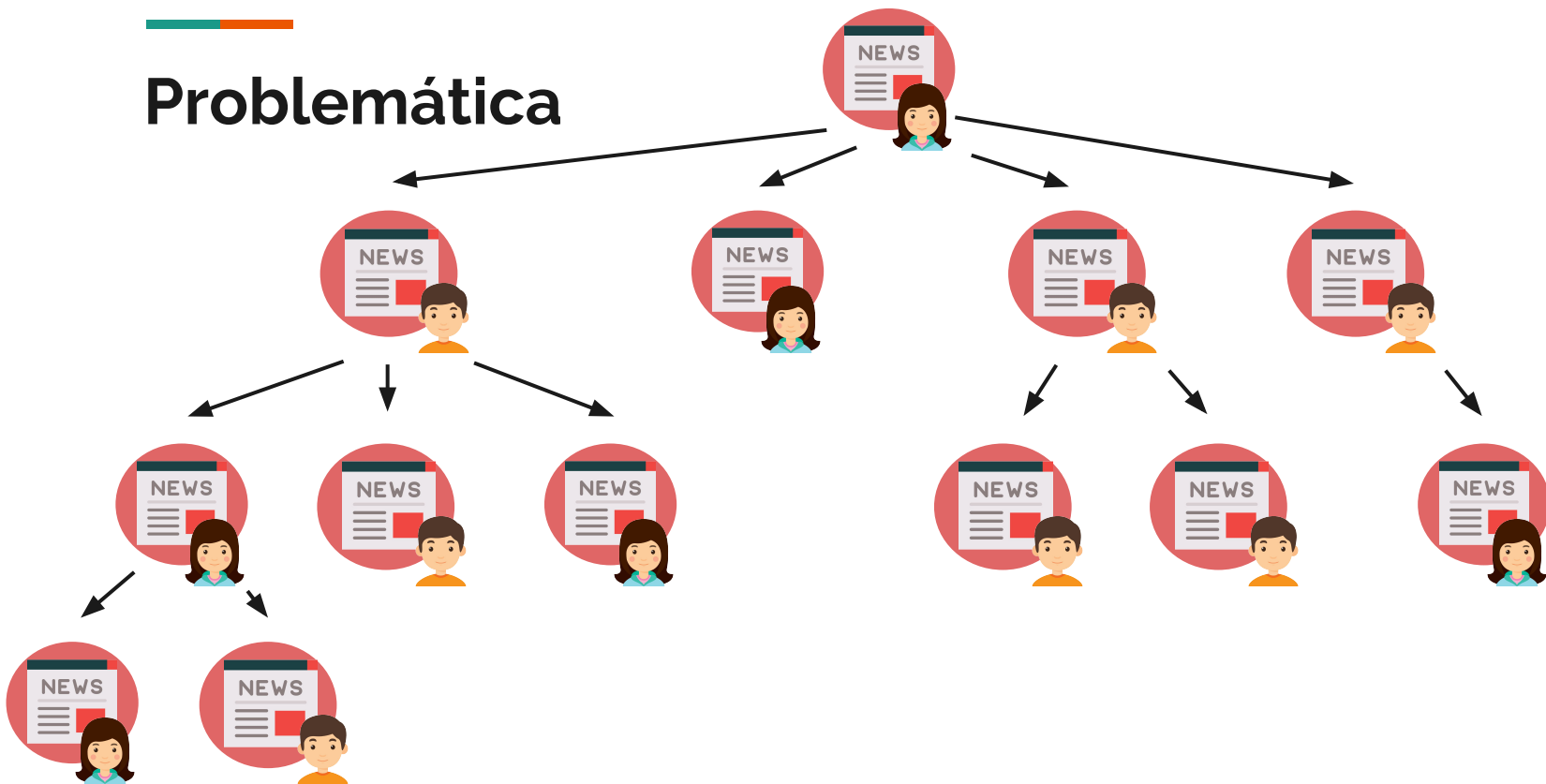
Problemática



Problemática



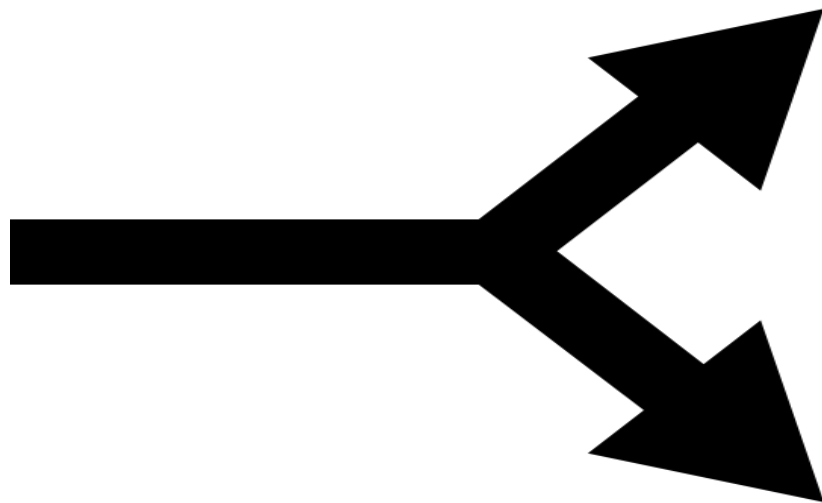
Problemática



Propuesta



Propuesta



FAKE





Metodología

- Cambio de **contexto** en recomendador ya implementado
- Implementación **clasificador**
- Modificación **código** original
- **Evaluar y comparar** resultados



Dataset

<https://www.kaggle.com/c/fake-news/data>

| id | title | author | text | label |
|----|-----------------------------------|--------------------|--|-------|
| 2 | Why the Truth Might Get You Fired | Consortiumnews.com | For those who might wonder why foreign policy makers repeatedly make bad choices, some insight might be drawn from the following analysis. The action here plays out in the United States, but the lessons are probably universal. (...) | 1 |



Referencias

- [1] M. Tambuscio, G. Ruffo, A. Flammini, F. Menczer, “Fact-checking Effect on Viral Hoaxes: A Model of Misinformation Spread in Social Networks”. En WWW’15 Companion Conference, pages 977-982, 2015.
- [2] K. Rapoza, “Can ‘fake news’ impact the stock market?”.
www.forbes.com/sites/kenrapoza/2017/02/26/can-fake-news-impact-the-stock-market/, Forbes, 26 february 2017.
- [3] S. Vosoughi, D. Roy, S. Aral, “The spread of true and false news online”. Science 359, 1146–1151, 2018.
- [4] G. Shani, A. Gunawardana, “Evaluating Recommendation Systems”. En Recommender systems handbook, pp. 257-297. Springer US, 2011.
- [5] Y. Huang, “News4U: A news recommendation system tailored for user communities”.
<https://github.com/huangy22/NewsRecommender>, GitHub, 29 june 2017.