

Ideas de Proyectos Finales RecSys 2020

Denis Parra
IIC3633

Plazo para entrega de propuesta

- Lunes 5 de octubre 2020, a las 20:00 para enviar su propuesta

Musica: Datos Spotify para RecSys challenge 2018

- El siguiente paper describe el RecSys challenge 2018, los equipos que enviaron y sus repos, datasets y métricas

Table 1. Basic statistics of the Million Playlist Dataset.

Property	Value
Number of playlists	1,000,000
Number of tracks	66,346,428
Number of unique tracks	2,262,292
Number of unique albums	734,684
Number of unique artists	295,860
Number of unique playlist titles	92,944
Number of unique normalized playlist titles	17,381
Average playlist length (tracks)	66.35

- <https://arxiv.org/pdf/1810.01520.pdf>

- Idea de proyecto: Investigar una forma novedosa de recomendar con este mismo dataset

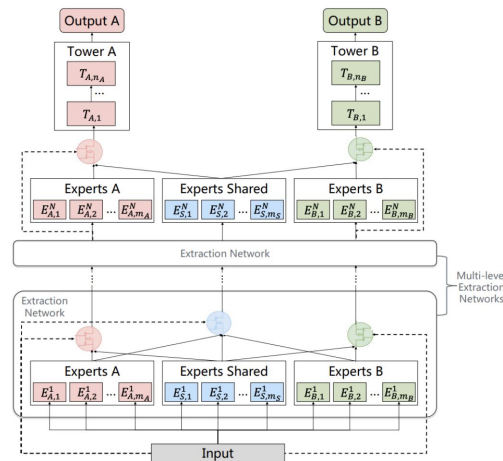
Reproducir un paper

- No todos los paper publicados tienen una implementación pública
- Una muy buena idea de proyecto es implementar el modelo del paper ver si es posible replicar resultados
- Posteriormente, si la replicación fue posible, testear lo mismo pero con otros datasets para ver si el resultado es consistente (reproducible)

Ejemplos: DVBPR y Attentive Collaborative Filtering para recomendación de imágenes

Multi-stakeholder recommendation

- Recomendar no sólo para un objetivo (ejemplo: rating o ranking) sino que considerar un approach multiobjetivo o multicriterio
- Ejemplo de Dataset: Beer dataset del repositorio de Julian McAuley UCSD
- Best paper en RecSys 2020 „Progressive Layered Extraction (PLE): A Novel Multi-Task Learning (MTL) Model for Personalized Recommendations“



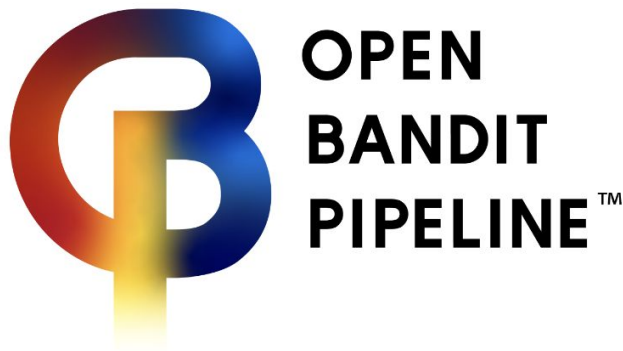
Fairness vs. Bias

- Investigar métricas de sesgo (ranking) para recomendaciones
- Comparar modelos no sólo a nivel de accuracy y diversity, sino que también a nivel de bias
- El dataset debe tener “grupos protegidos” para poder medir sesgo (por ejemplo, el género de los usuarios o de los productores del contenido)
- O se debe identificar clientes/productores y medir que si son similares, tienen recomendaciones y resultados similares

Online Learning / Bandits

- Online Learning y bandits se han tomado la última versión de la conferencia RecSys. Se compartió una biblioteca y dataset para experimentación:
- <https://github.com/st-tech/zr-obp>

README.md



[Documentation](#) | [Overview](#) | [Installation](#) | [Usage](#) | [References](#) | [Quickstart](#) | [Open Bandit Dataset](#) | [日本語](#)

Adversarial RecSys

Jiaxi Tang, Hongyi Wen, and Ke Wang. 2020. Revisiting Adversarially Learned Injection Attacks Against Recommender Systems. In Fourteenth ACM Conference on Recommender Systems (RecSys '20).

Konstantina Christakopoulou and Arindam Banerjee. 2019. Adversarial attacks on an oblivious recommender. In Proceedings of the 13th ACM Conference on Recommender Systems (RecSys '19).

Adversarial Generative RecSys

Homanga Bharadhwaj, Homin Park, and Brian Y. Lim. 2018. RecGAN: recurrent generative adversarial networks for recommendation systems. In Proceedings of the 12th ACM Conference on Recommender Systems (RecSys '18).

[Interpretable User Profiles] Murium Iqbal, Kamelia Aryafar, and Timothy Anderton. 2019. Style conditioned recommendations. In Proceedings of the 13th ACM Conference on Recommender Systems (RecSys '19)

Datasets

- Kaggle

<https://www.kaggle.com/search?q=recommender+in%3Adatasets+sortBy%3Adate>

- Pagina de Julian McAuley UCSD

<https://cseweb.ucsd.edu/~jmcauley/datasets.html>

Algunas Ideas

Recomendación de imágenes con datos de pinterest, wikimedia, behance, ugallery (tenemos estos datasets):

- + Implementar metodos VisRank, VBPR, DVBPR, ACF como baselines para luego proponer alguna mejora sustancial
- + Utilizar noción de estilo / contenido para generar recomendaciones
- + Generar explicaciones usando Network Dissection

Resúmenes de la conferencia RecSys

<https://www.inovex.de/blog/recsys-2020-highlights/>

- Saito: Doubly Robust Estimator for Ranking Metrics with Post-Click Conversions
- Wang et al.: Causal Inference for Recommender Systems
- Guo et al. (Twitter): Deep Bayesian Bandits: Exploring in Online Personalized Recommendations

Dominant Topics: Biases, Fairness, Causality, Bandits and Reinforcement Learning

In my personal view, the outcry of recent years regarding the narrow-minded focus on accuracy was heard by the community. Acknowledging biases and developing de-biasing techniques, looking beyond correlation and trying to model causal effects as well as addressing fairness and accountability are among the dominant topics of the conference. I believe that the RecSys research community is much more aware of these topics than general society gives it credit for. However, my view is biased towards what I saw at the conference and not what happens in general behind each system. But there is also evidence that addressing these issues drives beneficial long-term business goals and is therefore grounded in industry's own interest.

Resumenes de la conferencia RecSys

<https://www.urinieto.com/2020/09/recsys2020/>



**ORIO (URI)
NIETO**

PhD in Music Data Science



Carousel Personalization in Music Streaming Apps with Contextual Bandits

Cool application of contextual bandits in a "Carousel" music application, tested in a large-scale scenario by the good people at Deezer. They not only report how effective their model is, but they also release a publicly available dataset and an open-source carousel personalization environment to play with. The code and dataset should be [here](#) "by the end of September 2020." This makes me both happy and sad: happy because I see that I'm not alone in terms of doing research at the very last minute (it's September 28th and the repo is still empty); sad because the code isn't available yet and the world would be a much better place if researchers wouldn't do things at the very last minute.

Content-Collaborative Disentanglement Representation Learning for Enhanced Recommendation

Do you want to combine content and collaborative features together? Disentangle them first! This paper proposes a two-level disentanglement model that makes use of the KL divergence to ensure that the learned features are statistically independent. Results obtain state-of-the-art on 3 different datasets.

Exploring Longitudinal Effects of Session-based Recommendations

Interesting "reinforcement effect" investigation on music recommendation. The authors run several algorithms in a repeated manner and report how, in the long term, recommenders tend to pick smaller pools of songs because they keep recommending more and more similar tracks due to the reinforcement effect. By running a re-ranking method they are able to reduce this potential bias in the recommendations.

KRED: Knowledge-Aware Document Representation for News Recommendations