

Marcelo Sartori Locatelli - 2019006949

Link para a apresentação: <https://youtu.be/sMuPkOkjbww>

Matrix Factorization

- Três testes: NMF, SVD, SVD++.
- Biblioteca Surprise.
- Execução rápida e eficiente.
- Otimização bayesiana para hiperparâmetros.
- Resultados no kaggle:
 - NMF: 0.25
 - SVD: 0.23
 - SVD++: 0.32

Recomendações CB com BERT

- Representações do conteúdo (plot, título, etc) com BERT.
- Recomendação Rocchio.
- Problemas:
 - Custo de tempo para gerar representações.
 - Precisava de muita RAM.
 - Reduzir o custo de RAM aumentava ainda mais o tempo gasto.
- Abandonada após mais de 6 horas de execução.

LightFM

- Recomendador híbrido.
- Permite incorporar dados do conteúdo.
- Implementa losses WARP e BPR -> Learning To Rank.
- Treinamento e previsões rápidas.
- Pouco gasto de espaço.

Testes com LightFM

- Descobrir quais features do item valem a pena.
- Features numéricas convertidas em categóricas.
- Votos do imdb: essenciais -> 0.54 ndcg.
- Gênero (one-hot encoded), rating (metascore e imdb) e diretores -> 0.63.
- Vitórias em awards e línguas -> 0.65, quase batendo todos baselines.
- Alguns atributos pioraram o modelo.
 - Nomeações, atores, escritores, etc.

Modelo final

- Cold start de usuario: ordenar itens pelo valor dos baselines.
- Popularidade é importante.
- $\text{Score} = \text{predição} + 0.25 * \text{popularidade}$.
 - Predições e popularidades padronizadas (standardized).
- Erro final: 0.663
- Outros testes
 - Recomendadores da microsoft;
 - Ensembles.