

LISTA SOBRE FILTRAGEM COLABORATIVA - RECSYS 2021/2

Prof. Filipe Braida do Carmo

15/03/2021

Os exercícios deverão utilizar o *dataset* MovieLens [1]. Essa base de dados vem de um sistema de recomendação de filmes desenvolvido pelo GroupLens e possui 100.000 avaliações. Nela existem 943 usuários e 1682 filmes, e o seu conjunto de preferência é um número inteiro que varia entre um a cinco. Além das avaliações, existem informações sobre usuários, como a idade e o sexo, e sobre os filmes, tais como título do filme e a data de lançamento do filme. Essas informações não serão utilizadas nos exercícios dessa lista.

Questão 1: Gere um gráfico que mostre a quantidade de avaliações de cada usuário. Ordene pela quantidade de avaliações.

Questão 2: Gere o histograma das notas, ou seja, quantos usuários deram nota um, quantos deram nota dois e assim por diante.

Questão 3: Gere a lista dos dez filmes mais bem avaliados do conjunto de dados.

Questão 4: Gere a mesma lista do exercício anterior, mas utilizando a *Damped Mean* com k igual a 30. Faça uma comparação com o exercício anterior.

Questão 5: Divida a base aleatoriamente em 80% e 20%. O primeiro será chamado de base de treinamento e a segunda base de teste. Calcule a média das notas de cada usuário na base de treinamento. Caso não tenha nenhuma avaliação considere a média global. Depois na base de teste você irá prever as notas e utilizará a média do usuário como previsão. Por fim, calcule a média do erro ($MAE = \frac{1}{n} \sum_{u,v} |p_{uv} - r_{uv}|$ - onde p_{uv} é a nota prevista de um usuário u para um item v e r_{uv} é a nota de um usuário u para um item v .) da base de teste utilizando essa nota prevista.

Questão 6: Faça exatamente igual ao exercício anterior, mas utilize a *Damped mean* com k igual a 30. Compare e discuta o erro das duas formas de previsão.

References

- [1] BN Miller, Istvan Albert, and SK Lam. MovieLens unplugged: experiences with an occasionally connected recommender system. *Proceedings of the 8th ...*, pages 263–266, 2003. URL <http://dl.acm.org/citation.cfm?id=604094>.