Discussion 2

Albert Gilharry

June 20, 2019

Spotify's Recommender Discussion

First of all, I found it real interesting that the speaker was a PHD student in Machine Learning, a good motivation those that are aspiring to do a PHD related to the field of data science. This is also a bit intimidating I might add, but the opportunity to learn from people with such background is a pleasure.

I found it quite interesting but not surprising that Spotify was able to frame the matrix factorization problem in way that ALS can be used instead of more "fancy" methods. ALS is very useful in this case cause because it converages quite quickly which is very important when dealing with millions of items or users. Creating ratings by weighting items based on the number of times streamed is a very neat way to provide ratings when the users did not explicitly rate the items. It shows that recommenders can do just fine without explicit user-defined ratings.

Coming from more of a computer science background, I had to implement many algorithms and techniques from scratch in the academic setting but rarely had to do so in the professional setting. Spotify had to implement their own serializers because of performance compatibility issues and it was interesting find another scenario where one must provide their own implementation of out of the box functions. Even with all the improvements made by Spotify, they were still unable to run their system on the entire dataset. This shows that when dealing with big data we should aim to provide the best acceptable performance rather than the best possible performance. As long as the users are happy the resource overhead that comes with gaining every second of improvement is not always worth it.