Midterm review for APSTA - GE-2017

Jing Xie

4/3/2017

Understanding the questions

The goal of the project is to built a music reccomander based on user's listening history. You have a very well defined question with well designed research project. So far you have made a lot of progress and you already have a reccomander built. I would tag your project as green since I believe you can learn a lot from this well designed project.

The project is largely based on the package called reccomanderlab and you are using the collabertive filtering algorithms for building the system.

My review Questions

- You can easily convert the data from raw data format to the utility matrix of shown in your slide using group_by and spread command in dplyr package. It might be helpful to speed up the formatting as dplyr is a c++ based language.
- I am not sure if this is something you woule like to do , but this is just some idea that you can try out:
  - As the data set you obtain is very large. working with the whole dataset can be very challenging. One thing I am thinking might be working is to preprocess the artists into multiple groups. Since you can obtain the tags for each artistics you might be able to classified the aritsts into several groups and reduce the demention of the utilty matrix.
  - I personally believe users listen to music by genres. The algorithm shall reccomand to the users music 's genre that is prorpotional to the original genre pool. For example if a user like 3 jazz artistics and 5 classic, the reccomander shoud reccomand similar porortion of jazz : classic in stead of heavily reccomanding one. This could be also a way of testing the algorithm.

Conclusion

I think you are working on a very meaningful project and I beilieve it woule yield some pretty nice result.