Overview

Weeks 3 & 4 - November 4th - November 24th

- **Focusing on importing, collecting, evaluating, reporting, and processing Amazon datasets into our hands.**
- **Submit corrections as requested.**
- **Keep track of notes about key findings and important relevant articles about how to import datasets and where to import them in the future.**
- **Update our thesis outline based on what methods we used to import data and where we currently store that data for later experiments.**
- **Schedule project meetings with Dr. James Caverlee (faculty advisor) and Yin Zhang (graduate student advisor) weekly to discuss the content of our thesis, research progress, our public presentation information, installments and progress reports. If they request changes, submit those changes.**
- **Make sure that by this time, we should have received all research compliance approvals, been added to all of the necessary protocols, and completed all required training necessary to continue with our project. In this way, we can grab the data.**

Schedule

Monday, November 27th, 2017 Notes:

- **Finished focusing on importing, collecting, evaluating, reporting, and processing Amazon datasets into our hands by writing scripts to load in the datasets and store data.**
- **Submitted corrections for the first installment.**
- **We will update notes on key findings from the dataset and any important information on how to import them here:**
    - **We have began managing the analysis of our data through Jupyter Notebook and manage the contents of the notebook through a Github repository below: https://github.com/KevinJ97/CSCE491-Research**
    - **While reading the Electronic Metadata information, we found the structure of the categories is purely hierarchical for each category such as: Electronics -> Computers & Accessories -> Monitor Accessories.**
    - **We have written the code to analyze the data and create a subcategory dictionary where we can analyze the relationship between categories and subcategories.**
    - **One thing that is interesting is that although we are using McAuley's dataset on electronic metadata, we are also getting information about products that do not entirely belong under technology. This became apparent when traversing the subcategories tree and there is the category of Movies & TV included along with Electronics. We believe that this is due to the fact some items are unique or broad enough to cover more than one unique category.**
    - **We broke down the relationship between the subcategories and began to find that each of the relationships for electronics can almost uniquely be classified by size and**

interconnectivity. The spreadsheet of our analysis is below: https://docs.google.com/a/tamu.edu/spreadsheets/d/1mtD7vA-PlaeL-kDYPZ4gclI8MqrieFjY_Ym9314gGkQ/edit?usp=sharing

- **Another interesting finding is that while analyzing the subcategory relationships, we found that the concept of style is never directly applicable to the two subcategories themselves. For example, a notion of style might be associated with a laptop and a case however the notion of style is actually a user preference and the laptop itself does not necessarily have a style compatibility with the case.**

- **We will be applying for the public presentation next semester when we have more information covered.**
- **Victoria is reading http://cseweb.ucsd.edu/~jmcauley/pdfs/sigir15.pdf and annotating our findings. We understand the difference between previous research on recommender systems and how our research to create a new recommender system based on compatibility is different by noting it in our findings below: https://docs.google.com/presentation/d/13nyDg_nicRqCS8MkdXWe-TBX31x4Gi8yv-KA7nYJ7Ag/edit?usp=sharing**
- **We do not need research compliance approval. We can collect the data freely.**