Pandora Radio

Pandora Radio is an online music subscription service the customizes a users music feed to must appleal to their individual preferences. They are particularily interesting because of the lengths they've gone to to perfect content based recomenders.

Content Based Recomender Systems (CBRS)

These systems can be thought of as a using the atributes of a product to recomend similar ones. For example, if I like *beef bourguignon* a CBRS would recomend *coq au vin* because they are both protiene based, braised French dishes that emphasize red wine and brady along with root vegetables.

Collaberative Filtering Recomender System (CFRS)

CFRS base their recomendations off of what similar users also enjoyed. In this case, I like beef bourguignon and user X like beef *bourguignon* and *coq au vin*, thus it would recommended that.

Costs and Benefits

CBRS benefit from not having a 'cold start'. I can be recomended similar products as soon as I reveal my preference for one. Userprofiles do not need to be ketp which makes this ideal for sites that don't have logins etc. Unfortunetly, these systems require lots of investment to lable the content library. They also don't adapt well to slight changes in individual preference or even just bad products that have similar meta data.

CFRS on the other hand are cheaper to implement and can be highly personalized but require a lot of data to start.

Pandora's Recomender System

Pandora is interesting because they took the CBRS so far. Pandora has something they call the Music Genome Project (MGP). This is a vector of over 450 features such as genera, tempo etc. This is a highly ambitious project given the cost of employing 'musicologists' to classify every song. Rumor has it that each song takes 15-30 minutes to classify.

As far as I can tell, Pandora does not reley on colaberative filtering at all. Hoever, it does build a user profile. If I start playing music by Weezer and the next song

played has similar attributes/meta data but I don't like it, Pandora will update my preferences.

This is interesting because rather unknown bands have an equal chance of matching a user's criteria. While the style may not be original, the bands might be.

While Pandora seems to claim that they don't use any collaborative filtering, I remain a bit skeptical that they haven't at least looked at implementing it in conjunction with the MGP.

Adversarial Attacks on Recommenders

There are many bad actors who might want to influence ratings. This can range from competitors trashing a different store online, to inflated reviews of products by the producer. In the example article the attack was actually done in an attempt to deny the Armenian genocide.

It can be really hard to find these for multiple reasons. First, what even constitutes a fake review is debatable. If I own a restaurant, I could give my friends a discount to come in and rate it well. It's less fake than a review of somebody who has never seen a movie but rates it, but it's less real than a less biased review. Or what if I bought a book on Amazon to purposefully give it a bad review?

Philosophy as side, it is very difficult to classify one to two fake reviews. There are many techniques to classify fake reviews. Similar to email spam, this can be seen as a conflict between interested parties and the review hosts.

These are very interesting, however, I think an even more dangerous application of this is attacks on user information. This would involve changing input data slightly and then being able to tell specific user input based on the changes of the recommender. This is summarized in this post on differential privacy.

Depending on the algorithm and data collected, introducing new data can effect the recomender system in a significant way. Without differential privacy, these effects could be used to undermine privacy.