Online Collaborative Filtering

Joshua Morton

05/15/2014

This outlines some of the mathematics used in the online collaborative filter used in other areas of this tool. This is both for my sanity and a way to outline and think through some of the mathematics used elswhere in the codebase. To begin with, I'll outline the collaborative filtering code as a whole.

The estimated opinion, r, by user u of an item i is defined below, where U is the set of all users that are not u.

$$r_{u,i} = k \sum_{u' \in U} simil(u, u')r_{u,i}$$

This function makes use of two additional functions, $simil(u, u')$ and k. $simil(u, u')$ operates over the set I, which is defined as the intersection of the sets of objects that have previously been rated by both u and u'

$$k = \frac{1}{\sum\limits_{u' \in U} |simil(u, u')|}$$

$$simil(u, u') = \frac{\sum\limits_{i' \in I} (r_{u,i})(r_{u',i})}{rss(u) \times rss(u')}$$

In this case, rss signifies the root sum squares of all ratings for that user. Here, S refers to the set of all items rated by a given user. To put this mathematically:

$$rss(u) = \sqrt{\sum_{i \in S} r_{u,i}^2}$$

Putting this all together, we get that the formula to calculate the rating for a given item at any time can be expressed as

$$r_{u,i} = \frac{\sum\limits_{u' \in U} \left(simil(u, u') \cdot r_{u,i} \right)}{\sum\limits_{u' \in U} |simil(u, u')|}$$

This expands to the awful looking and disgusting piece of math that is

$$r_{u,i} = \frac{\sum\limits_{u' \in U} \left(\dfrac{\sum\limits_{i' \in I} (r_{u,i})(r_{u',i})}{\sqrt{\sum\limits_{i \in S} r_{u,i}^2} \sqrt{\sum\limits_{i \in S} r_{u',i}^2}} \cdot r_{u,i} \right)}{\sum\limits_{u' \in U} \left| \dfrac{\sum\limits_{i' \in I} (r_{u,i})(r_{u',i})}{\sqrt{\sum\limits_{i \in S} r_{u,i}^2} \sqrt{\sum\limits_{i \in S} r_{u',i}^2}} \right|}$$

Much of this is repetetive code and can be made into functions, but this is still obiviously worst case $\mathcal{O}(n^2)$