Overcoming Sparsity in Collaborative Filtering via Social Networks

Allen Tran

November 23, 2014

Suppose there exist users $u \in U$, locations $l \in L$ and votes by users on locations described by the relation $v : U \times L \rightarrow [0, 1]$ times with a relation between users $t : U^2 \rightarrow \mathbb{R}$ denoting strength of relations. Let 0 in the relation v denote the absence of a vote. This describes the environment, now let's talk data.

Each user, u has a set of visited locations $L_u = \{l : \hat{v}(u, l) > 0\}$ where $\hat{v}$ denotes observed votes on locations. The problem in collaborative filtering is to estimate the relation v.

Let's do this via a model where we factor a location into n factors, where each factor is part of a compact set.[1] What are these factors? Whatever is necessary to succinctly describe these locations. For bars, it could be loudness, darkness, cool factor, sports barness, etc. For each of these n factors, each user has a corresponding set of preferences, denoted β_u. Using these two, let's model a user's specific preference for a location as

$$v(u, lt) = \alpha_l' \beta_u \tag{1}$$

Given we have data on these votes, $\hat{v}$, use a minimum distance estimator of your choice to find these parameters, while being aware of the standard issues relating to overfitting and tradeoffs between distance measures.[2]

The well documented problem with this approach is the sparsity problem, where we have many people who have voted very few times. Intuitively, how can recommendations be generated for a user if I have very little user-specific information. There are many approaches to overcoming this. Here I'll document the trust-based approach, which uses linkages between users.

Trust based solutions to the sparsity problem implicitly find links between users based on similarity. However, this does not capture real life relationships between different people.

[1] Why compact? We can guarantee a maximum exists for a continuous function defined with arguments in a compact set. More practically, boundedness reduces the parameter space we search over while continuity smooths out kinks in the loss function, making the optimization behave better in practice.

[2] An EM algorithm is super useful here as holding fixed a set of parameters, we have a loss function which is linear in parameters which has a closed-form solution. In other words, continual iteration of matrix operations until convergence which avoids use of a minimizer, apart from any hyper-parameters.

Often user's friends or trusted relations have quite different preferences which are still valued by users. Just because I visit Chinese restaurants doesn't mean I want to see recommendations for more Chinese restaurants. I want "same but different".

Instead of finding implicit links between users, I use the data to find direct measures of trust between users. In particular, two users trust is modeled as follows, where $t : U \times U \to [0, 1]$ and $m(u, u')$ is a measure of u's connection to u' from the data which need not be symmetric.

$$\hat{t}(u, u') = \frac{\min\left(\hat{m}(u, u'), \hat{m}(u', u)\right)}{\max\left(\sum_{u''} \hat{m}(u, u''), \sum_{u''} \hat{m}(u', u'')\right)} \tag{2}$$

For completeness, let $\forall u, \hat{t}(u, u) = \gamma$ since $\hat{m}(u, u)$ is often undefined. This is a hyper parameter that governs how much a person "listens" to friends recommendations.

Given these trust relations, form an aggregate trust matrix, T, which measures the degree of trust between two users. The element i, j indicates the degree of trust between users i and j. Starting with an initial set of votes from the data, it is easy to propagate votes with users listening to other users in proportion to their trust.

$$V^{n+1} = \lambda T V^n + (1 - \lambda) V^n \tag{3}$$

where λ is the standard PageRank redistribution parameter and V^0 are the initial set of votes from the data with element u, l being $\hat{v}(u, l)$.

Under suitable conditions on the trust matrix, there exists a unique solution to this fixed point problem.[3] The solution with the eigenvector corresponding to the eigenvalue 1 gives the same distribution across the rows of V. That is, the same distributions over the users votes for any location such that for a column of V, we get the distribution over the users in terms of their preferences mattering. But what we want to assess for recommendations is a row, a particular user and the set of locations.

[3] Blackwell's sufficient conditions to show a contraction and the contraction mapping theorem will work.