

Imputing Trust Network Information in NMF-based Recommendation Systems

Fatemah Alghamedy, Jun Zhang
Department of Computer Science, University of Kentucky

Introduction

With the emergence of E-commerce, recommendation system becomes a significant tool which can help both sellers and buyers. It helps sellers by increasing the profits and advertising items to customers. In addition, recommendation systems facilitate buyers to find items they are looking for easily.

In recommendation systems, the rating matrix R represents users' ratings for items. The rows in the rating matrix represent the users and the columns represent items. If particular user rates a particular item, then the value of the intersection of the user row and item column holds the rating value. The trust matrix T describes the trust relationship between users. The rows hold the users who create a trust relationship - trustor - and the columns represent users who have been trusted by trustors - trustee -.

Nonnegative Matrix Factorization (NMF) has been applied in the collaborative filtering to learn the missing values in the rating matrix. A nonnegativity constraint is enforced in the linear model to guarantee that all users' ratings can be represented as an additive linear combination of canonical coordinates.

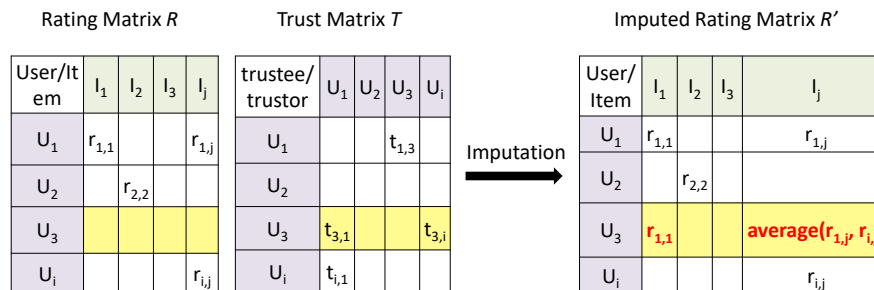
$$R_{m \times n} \approx U_{m \times k} \cdot V_{n \times k}^T$$

Many studies show the relationship between the users' taste and their friends' taste. Users trust each other at the beginning because they agree with their ratings and reviews. However, after a while, the trustee influences the trustor even on some topics that they did not agree on in the past.

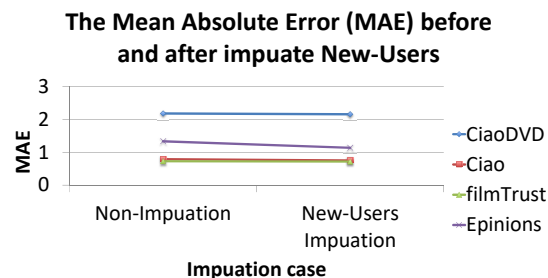
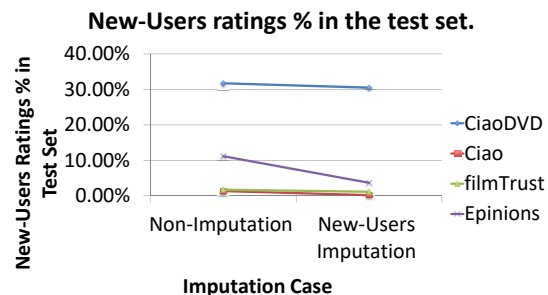
Imputation is an approach to complete missing data which is the process of replacing missing data with substituted values

Proposed Method

We propose an Nonnegative Matrix Factorization - based approach in collaborative filtering based recommendation systems to handle the cold-start users issue, especially for the New-Users who did not rate any items, by utilizing the trust network information to impute missing ratings before NMF is applied.



Results and Discussion



The Epinions dataset gets the most improvement because 7.5% of the ratings of New-Users group in test set do not belong to the New-Users group after the imputation process, which is the largest percentage among other datasets. The lowest ratio improvement is with the FilmTrust dataset because it has the lowest difference in the New-Users group ratings percentage between Non-Imputation and New-Users imputation cases. The improvement ratio with CiaoDVD is in between the FilmTrust and Ciao datasets. The difference in the New-Users group ratings percentage is more in CiaoDVD than FilmTrust and slightly more than Ciao datasets. However, CiaoDVD still suffers from the highest percentage of the New-Users group ratings after the imputation process.