Explanation of the Algorithm

Collaborative Filtering for Implicit Feedback Datasets

On the surface, recommendation systems process information through some explicit feedback. However, in real-life applications, recommendation systems need to be centred on implicit feedback, inferring user preferences from implicit feedback to users, and reflecting opinions indirectly by observing users' behaviours.

In the paper "Collaborative Filtering for Implicit Feedback Datasets", the researchers designed their implicit feedback model by analyzing the latent factor model, which represents users' preferences by introducing binary variables.

First, they formalized the notion of confidence which the r_{ui} variables measure. They introduced a set of binary variables p_{ui}, which indicates the preference of user u to item i. The p_{ui} values are derived by binarizing the r_{ui} values:

$$p_{ui} = \begin{cases} 1 & r_{ui} > 0 \\ 0 & r_{ui} = 0 \end{cases}$$ [1]

If user u consumes item i, it is the first case. Otherwise, it is the second case. However, a user not consuming an item does not necessarily mean that the user does not like it; the user may be unaware of the existence or price limitations of the item. Therefore, the researchers introduced a set of variables C_{ui} to indicate the user's stronger liking mood for the item.

$$C_{ui} = 1 + \alpha r_{ui}$$ [1]

The researchers set the rate of mood increase α to 40 and obtained good results.

Users' preferences are assumed to be inner products: $P_{ui} = x_u^T y_i$. These vectors are referred to as user factors and item factors, respectively. In essence, these vectors strive to map users and items to a common space of potential factors where they can be directly compared. The researchers computed the factors by minimizing the following cost functions.

$$\min_{x_\star, y_\star} \sum_{u,i} c_{ui}(p_{ui} - x_u^T y_i)^2 + \lambda \left(\sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$ [1]

The exact value of the parameter λ is data-dependent and determined by cross-validation.

The m-n of the cost function can easily reach several billion so the researchers propose an effective optimization scheme.

When the user factor or project factor is fixed, the cost function becomes quadratic and the global minimum can be easily computed. this leads to an alternating least-squares optimization process. the steps of recalculating the user factor and project factor are alternated, and each step will reduce the value of the cost function. The researchers address the sparsity of the objective function that can result from displaying feedback by using the structure of the variables. The researchers simplified the latent factor model to a linear model that predicts preferences as a linear function of past behaviour.

Preference and confidence level should be the two factors that should be observed by the user.

The researchers provide a latent factor algorithm that directly addresses the preference-confidence paradigm. Unlike the explicit dataset, the model here should take as input all user-item preferences, including those that are unrelated to any input observation. The algorithm is linear in the size of the inputs while addressing the full range of user-item pairs without resorting to any subsampling. An interesting feature of the algorithm is that it allows the interpretation of recommendations to the end-user, which is rare in latent factor models.

How it integrates with the project

The algorithm mentioned in the paper belongs to a technique of collaborative filtering.
The algorithm mentioned in the paper is related to implicit feedback, while the ALS algorithm used in the project is related to display feedback, although they are not directly related, they both belong to the category of collaborative filtering.
The dataset used in the project is the display feedback dataset with rating values from 1 to 5, with 1 being disliked and 5 being very much like. For the display feedback dataset, alternating least squares is used, treating the unknown values as missing, which can lead to sparse objective functions. And the new algorithm proposed by the researchers tries to solve these problems by using the structure of the variables so that the process can be highly scalable. This gave me a deeper understanding of my project.
The latent factor algorithm proposed by the researchers allowed me to understand how to optimize my project. the disadvantage of the ALS algorithm is not only that it is an offline algorithm, but also that it cannot accurately evaluate newly added users or goods. The optimization method is to fix the other dimensions and optimize one of them.
The content of the paper gave me an understanding of the limitations of recommender systems using explicit data sets. Many times, users do not evaluate movies not because they do not like them, but they may not be aware of their existence. The project would have been more practical if the algorithm focused on implicit feedback as the algorithm in the paper does. For example, evaluating users based on the type of movie they watched.

Reference:

[1] Y. Hu, Y. Koren and C. Volinsky, "Collaborative Filtering for Implicit Feedback Datasets," *2008 Eighth IEEE International Conference on Data Mining, 2008*, pp. 263-272, doi: 10.1109/ICDM.2008.22.