تمرین امتیازی

**Recommender system**

نام و نام‌خانوادگی: روژینا کاشفی، هانیه ظاهری شماره دانشجویی: ۹۸۳۱۱۱۸-۹۸۳۱۰۹۰

----------------------------------------------------------------------------------------

با توجه به صورت سوال ما دو نوع فیلتر در recommender system ها داریم یکی بر اساس کاربران و یکی بر اساس شباهت item ها با یکدیگر است. در شباهت بر اساس کاربران مشاهده می‌کنیم کاربرانی که به مواردی که امتیاز داده‌اند، چه امتیازی لحاظ کرده‌اند و آن کاربرانی که شبیه یکدیگر امتیاز داده‌اند امتیاز مشابه‌ای دارند.

1. ابتدا کتاب‌خانه‌های مورد نیاز از اسپارک را صدا می‌کنیم
2. یک session اسپارک درست می‌کنیم
3. سپس فایل‌های csv را توسط اسپارک مطالعه می‌کنیم
4. مشاهده می‌کنیم ماکیسمم rateموجود ۵ است.
5. سپس برای دستور‌های کويری با استفاده از registerTempTable دیتاست لود شده را به عنوان table به spark. اضافه می‌کنیم.
6. سپس دو جدول مورد نیاز را join می‌کنیم
7. پس از آن داده‌های مورد نیاز را به دو دسته train و test تقسیم می‌کنیم تا ALS را اموزش دهیم.
8. ALS یک روش است که در collaborative filtering استفاده می‌شود. در collaborative filtering هدف پر کردن جا‌های خالی ماتریس userو game است. برای اینکار از یک سری فاکتور‌ استفاده می‌شود و این فاکتور‌ها توسط الگوریتم ALS قابل یادگیری هستند.
9. پیاده سازی als چند پارامتر ورودی دارد که توضیحات هر یک را مشاهده می‌کنیم.

* *numBlocks* is the number of blocks the users and items will be partitioned into in order to parallelize computation (defaults to 10).
* *rank* is the number of latent factors in the model (defaults to 10).
* *maxIter* is the maximum number of iterations to run (defaults to 10).
* *regParam* specifies the regularization parameter in ALS (defaults to 1.0).
* *implicitPrefs* specifies whether to use the *explicit feedback* ALS variant or one adapted for *implicit feedback* data (defaults to false which means using *explicit feedback*).
* *alpha* is a parameter applicable to the implicit feedback variant of ALS that governs the *baseline* confidence in preference observations (defaults to 1.0).
* *nonnegative* specifies whether or not to use nonnegative constraints for least squares (defaults to false).

1. روشی که از user-item استفاده می‌کند explicit است.
2. Cold start strategy برابر با drop قرار داده شده تا زمانی که یک itemدرون مدل قرار ندارد مقدار NaN جایگذاری نشود. و این مفید است زمانی که می‌خوایم prediction جدید داشته باشیم.
3. تنها معیار ارزیابی که برای ارزیابی ALS استفاده می‌شود RMSE است.
4. مشاهده می‌کنیم RMSE مدل ما .۰۹ که خیلی کم است.
5. سپس داده‌های یوزر را بر اساس بیشترین شباهت مرتب‌سازی می‌کنیم.