

## Recommender system

نام و نام خانوادگی: 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 می کنیم کاربرانی که به مواردی که امتیاز داده اند، چه امتیازی لحاظ کرده اند و آن کاربرانی که شبیه یکدیگر امتیاز داده اند امتیاز مشابه ای دارند.

۱. ابتدا کتابخانه های مورد نیاز از اسپارک را صدا می کنیم
۲. یک session اسپارک درست می کنیم
۳. سپس فایل های csv را توسط اسپارک مطالعه می کنیم
۴. مشاهده می کنیم ماکسیمم rate موجود ۵ است.
۵. سپس برای دستوره های کویری با استفاده از registerTempTable دیتاست لود شده را به عنوان table به spark اضافه می کنیم.
۶. سپس دو جدول مورد نیاز را join می کنیم
۷. پس از آن داده های مورد نیاز را به دو دسته train و test تقسیم می کنیم تا ALS را آموزش دهیم.
۸. ALS یک روش است که در collaborative filtering استفاده می شود. در collaborative filtering هدف پر کردن جاهای خالی ماتریس user و game است. برای اینکار از یک سری فاکتور استفاده می شود و این فاکتورها توسط الگوریتم ALS قابل یادگیری هستند.
۹. پیاده سازی als چند پارامتر ورودی دارد که توضیحات هر یک را مشاهده می کنیم.

- *numBlocks* is the number of blocks the users and items will be partitioned into in order to parallelize computation (defaults to ۱۰).
- *rank* is the number of latent factors in the model (defaults to ۱۰).
- *maxIter* is the maximum number of iterations to run (defaults to ۱۰).
- *regParam* specifies the regularization parameter in ALS (defaults to ۱.۰).
- *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 ۱.۰).
- *nonnegative* specifies whether or not to use nonnegative constraints for least squares (defaults to false).

۱۰. روشی که از user-item استفاده می کند explicit است.
۱۱. Cold start strategy برابر با drop قرار داده شده تا زمانی که یک item درون مدل قرار ندارد مقدار NaN جایگذاری نشود. و این مفید است زمانی که می خواهیم prediction جدید داشته باشیم.
۱۲. تنها معیار ارزیابی که برای ارزیابی ALS استفاده می شود RMSE است.
۱۳. مشاهده می کنیم RMSE مدل ما ۰.۹ که خیلی کم است.
۱۴. سپس داده های یوزر را بر اساس بیشترین شباهت مرتب سازی می کنیم.