به نام خدا

دانشگاه صنعتی امیرکبیر

دانشکده‌ی مهندسی کامپیوتر

تمرین امتیازی مبانی رایانش ابری

دکتر سید احمد جوادی

طراح سوال:

احمدی

دی ۱۴۰۱

Recommender system

یکی از رویکردهای collabrative filtering انتخاب اجناس بر اساس user-user است که با استفاده از آن می‌توانیم سیستم های توصیه‌گر بسازیم. مجموعه داده مورد استفاده در این بخش در دو فایل game.csv و ratings.csv می باشد. در فایل اول مشخصات بازی ها و در فایل دوم امتیاز کاربرها با جزییات آیدی بازی، آیدی کاربر و امتیاز داده شده است.

با استفاده از رویکرد collabrative filtering و بر اساس روش user-user سیستم توصیه گری بسازید که با دادن آی دی از میان بازی هایی که به آنها نمره نداده اند، تعداد ۵ بازی با بیشترین شباهت را پیشنهاد کند. در حل این مساله از معیار Cosine similarity استفاده کنید.

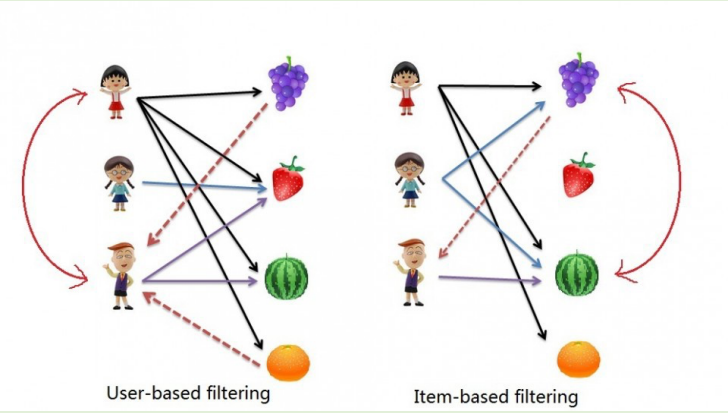
- در خروجی نام بازی ها را به همراه امتیاز شباهت آنها (به ترتیب نزولی امتیاز شباهت) بیاورید.
- درصورتی که دوبرنامه امتیاز یکسان داشتند، برنامه با ایندکس کمتر را انتخاب کنید.

در سال ۲۰۰۷ شرکت Netflix یک مسابقه برای پژوهش روی سیستم های توصیه گر و با استفاده از مجموعه داده ای از این شرکت ترتیب داد. در سال ۲۰۰۹ جایزه یک میلیون دلاری به یکی از پژوهشگران رسید. یکی از الگوریتم هایی که آن ها در روش خود استفاده کرده بودند همین روش ماتریس عامل بندی بود و این نشانه قدرت این روش است.

در مسائلی که ما اطلاعات خوبی از کاربر (مثل سن، جنسیت، شغل و ...) داریم اما در مورد آیتم ها دچار کمبود اطلاعات هستیم یا به دست آوردن ویژگی های آیتم ها دشوار است سیستم های توصیه گر فیلتر مشارکتی بسیار مناسب اند. دراین روش براساس شباهت رفتاری و عملکردی کاربرانی که در گذشته الگوی رفتاری مشابهی با کاربر فعلی داشته اند، پیشنهادها ارائه می شود. شاید تعریف آن کمی پیچیده باشد ولی به طور ساده روش فیلتر مشارکتی بر این فرض استوار است که کاربرانی که یک سری نظرهای مشابه درباره یک آیتم دارند، درباره آیتم های دیگر هم نظرهای مشابه دارند. فیلتر مشارکتی خود به دو دسته تقسیم می شود:

مبتنی بر کاربر(User-based) ،

مبتنی بر آیتم.(Item-based)



در روش مبتنی بر کاربر میزان شباهت سلیقه ی کاربران مشخص می شود و با توجه به آن، مقدار علاقه مندی کاربر به یک آیتم که تا به حال ندیده است مشخص می شود. در این سیستم از تشکیل بردار کاربر، که شامل تمام آیتم هایی است که کاربر امتیاز دهی کرده است به همراه امتیاز داده شده به آن ها، شروع می کنیم. سپس میزان تشابه کاربرها در یک ماتریس n*n که n تعداد کاربران است محاسبه می شود. برای محاسبه ی میزان تشابه می توان از تشابه کسینوسی استفاده کرد. حال ماتریس پیشنهادها می تواند تشکیل شود. امتیاز داده شده به آیتم ها در میزان شباهت کاربری که آن آیتم را امتیاز دهی کرده است با کاربر فعال ضرب می شود و به عنوان امتیاز پیشبینی شده در ماتریس پیشنهادها گذاشته می شود. این کار برای تمام آیتم هایی که کاربر ندیده است انجام می شود و سپس امتیاز ها مرتب می شوند و پیشنهادهای با بیشترین امتیاز، به کاربر هدف توصیه می شوند.کاربرها تما یل دارند محصولاتی را بخرند که کاربرهای با سلیقه مشابه آنها خریده‌اند.

برای مثال در جدول زیر کاربر U۱ احتمالا تمایل دارد تا محصول I۲ را بخرد، زیرا کاربرهای U۱ و U۴ محصول I۱را می‌پسندند و کاربر U۴ امتیاز بالایی به محصول I۲ داده است.

$I_۴$	$I_۳$	$I_۲$	$I_۱$	
۵	۵	؟	۴	$U_۱$
	۱	۲	۴	$U_۲$
۴	۲		۳	$U_۳$
		۴	۴	$U_۴$
۵	۳	۱	۲	$U_۵$

منبع[۲۱]

[1] https://www.geeksforgeeks.org/user-based-collaborative-filtering/
[2] https://sokanacademy.com/blog/recommender-system-collaborative-filtering-cf

توضیحات مهم:

- تمامی مستندات خود شامل گزارش و کدهای خود را در یک فایل فشرده با فرمت zip ذخیره کرده و با عنوان #StudentId_HW۴.zip بارگذاری نمایید .
- تمرین بدون گزارش و درک عمیق نسبت به اهداف تمرین، فاقد ارزش می‌باشد و **نمره‌ای به آن تعلق نمی‌یابد**.
- مطابق قوانین دانشگاه هرگونه کپی برداری **ممنوع** می‌باشد و برای شناسایی تقلب در این تمرین از Apache MOSS برای تشخیص مشابهت در کدها استفاده خواهد شد، در صورت مشاهده تشابه، نمره هر دو طرف صفر در نظر گرفته می‌شود.
- در هر مرحله، نتایج خود را تحلیل کنید.
- کدهای خود را برای خوانایی بیشتر کامنت گذاری کنید.
- در تمامی سوال‌ها تنها مجاز به استفاده از کتابخانه‌های numpy، apache spark و pandas می‌باشید. توجه داشته باشید کد می‌بایست توسط شما پیاده سازی شده و از استفاده از کدهای آماده اجتناب کنید.