بهترین روش برای توسعه‌ی یک Recommender System، ابتدا توسعه‌ی متدهای تست و ارزیابی متریک هاست. در این گزارش، چگونگی ارزیابی سیستم پیشنهاد دهنده و توسعه‌ی تست‌ها را با هم مرورمی‌کنیم.

شماره‌ی ۱ گزارش پیشرفت پروژه

پالایش مشارکتی سیستم پیشنهاد دهنده | ۱۴ فروردین ۱۴۰۰

فائزه سادات سعیدی نژاد

یکی از مهمترین مراحل ساخت یک سیستم پیشنهاد دهنده (و احتمالا هر سیستم و برنامه‌ی دیگری) نوشتن تست‌های آن بر اساس متریک‌های ارزیابی‌اش است. در ادامه متریک‌های ارزیای سیستم پیشنهاد دهنده را نام برده و توضیح مختصری می‌دهم.

۱. پوشش‌دهی[۱]: درصد پیشنهادهای امکان‌پذیر که سیستم پیشنهاد دهنده می‌تواند ارائه دهد.
۲. گوناگونی[۲]: مقدار گستردگی سیستم پیشنهاد دهنده برای پیشنهاد دادن آیتم‌ها گوناگون، البته باید توجه داشته باشیم که آیتم‌ها را خیلی رندوم هم نباید پیشنهاد دهد.
۳. تازگی[۳]: تازگی و محبوبیت آیتم‌های پیشنهادی توسط سیستم پیشنهاد دهنده.
۴. اعتماد کاربر[۴]: این متریک که ارتباط زیادی با محبوبیت و تازگی دارد، به این معناست که آیتم‌هایی باید به کاربر پیشنهاد داده شود که به چشمش آشنا است و خیلی آیتم‌هایی نیست که کاربر تا به حال ندیده است، برای جلب کردن اعتماد کاربر. توجه داشته باشید که اگر novelty درست پیاده‌سازی شده باشد، بین آیتم‌هایی که به چشم کاربر آشنا هستند و آیتم‌هایی که تازگی دارند باید مصالحه[۵] وجود داشته باشد.
۵. ریزش[۶]: هرچند وقت یک‌بار آیتم‌های پیشنهادی تغییر می‌کنند؛ آیا با هربار اکشن کاربر (امتیاز دادن، خرید کردن و...) سریعا آیتم‌های پیشنهادی تغییر می‌کنند؟
۶. واکنشی[۷]: چقدر سریع آیتم‌های پیشنهادی می‌توانند بر اساس تغییر رفتار کاربر، تغییر کنند.

روش‌های تست سیستم پیشنهاد دهنده:

یکی از روش‌های تست که زیاد کاربردی ندارد این که با استفاده از نظرسنجی، مستقیما از کاربر بپرسیم که آیتم‌های پیشنهاد داده شده را دوست دارد یا خیر (Perceived Quality). روش رایج‌تر A/B Testing است. برای استفاده از این روش باید Top-N recommender را آماده کرد که به دو روش می‌توان این کار را انجام داد (فهمیدن علاقه‌مندی‌های کاربر): صریح[۸] و ضمنی[۹] (در مقاله‌ی اصلی در رابطه با این موضوع بیشتر توضیح خواهم داد).

از نمونه تست‌های دیگر می‌توان به متریک دقت[۱۰] اشاره کرد که به دو صورت می‌توان آن را اندازه‌گیری کرد:

۱. خطای میانگین مطلق[۱۱]: همانطور که از اسمش مشخص است، برای اندازه‌گیری‌اش، میانگین تفاضل نمره‌ای که سیستم پیشنهاد دهنده فکر می‌کند کاربر به یک آیتمی می‌دهد از نمره‌ای است که در حقیقت کاربر به آن آیتم داده است. فرمولش به صورت زیر است:

---

[1] Coverage
[2] Diversity
[3] Novelty
[4] User trust
[5] Trade-off
[6] Churn
[7] Responsiveness
[8] Explicit
[9] Implicit
[10] Accuracy metric
[11] Mean absolute error (MAE)

$$\frac{\sum_{i=1}^{n}|y_i - x_i|}{n}$$

مثال:

predicted rating	actual rating	error
5	3	2
4	1	3
5	4	1
1	1	0

$$MAE = (2+3+1+0)/4 = 1.5$$

۲. خطای جذر میانگین مربعات[۱۲]: تفاوت میان مقدار پیش‌بینی شده توسط سیستم پیشنهاد دهنده و مقدار واقعی می‌باشد. RMSD یک ابزار خوبی است برای مقایسه خطاهای پیش‌بینی توسط یک مجموعه داده‌است و برای مقایسه چند مجموعه داده کاربرد ندارد. فرمولش بصورت زیر است:

$$\sqrt{\frac{\sum_{i=1}^{n}(y_i - x_i)^2}{n}}$$

مثال:

predicted rating	actual rating	$error^2$
5	3	4
4	1	9
5	4	1
1	1	0

$$RMSE= \sqrt{(4 + 9 + 1 + 0)/4} = 1.87$$

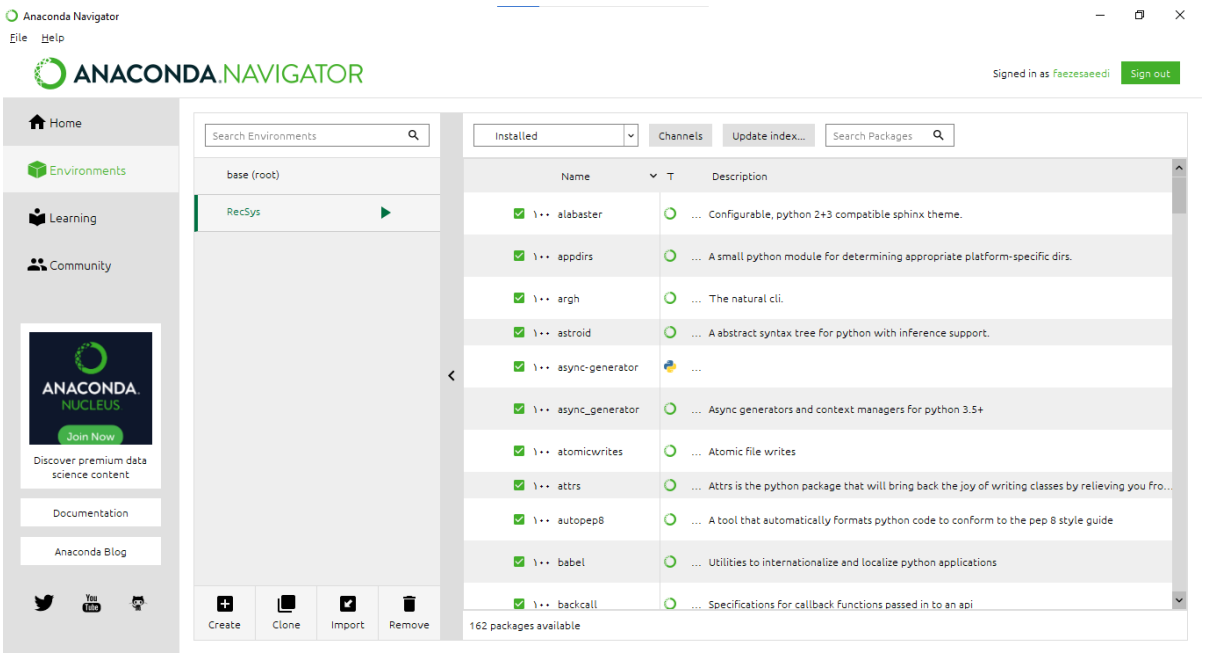
توجه کنید که هردو خطا هستند، پس هرچه مقدارشان کمتر باشد، بهتر است.

از روش‌های دیگر می‌توان به Hit rate اشاره کرد که به چندگونه می‌توان آن را انجام داد، اعتبار سنجی یک طرفه[۱۳]، average reciprocal hit rate، cumulative hit rate، rating hit rate.

[12] Root mean square error (RMSE)
[13] Leave-one-out cross validation

قسمت عملی:

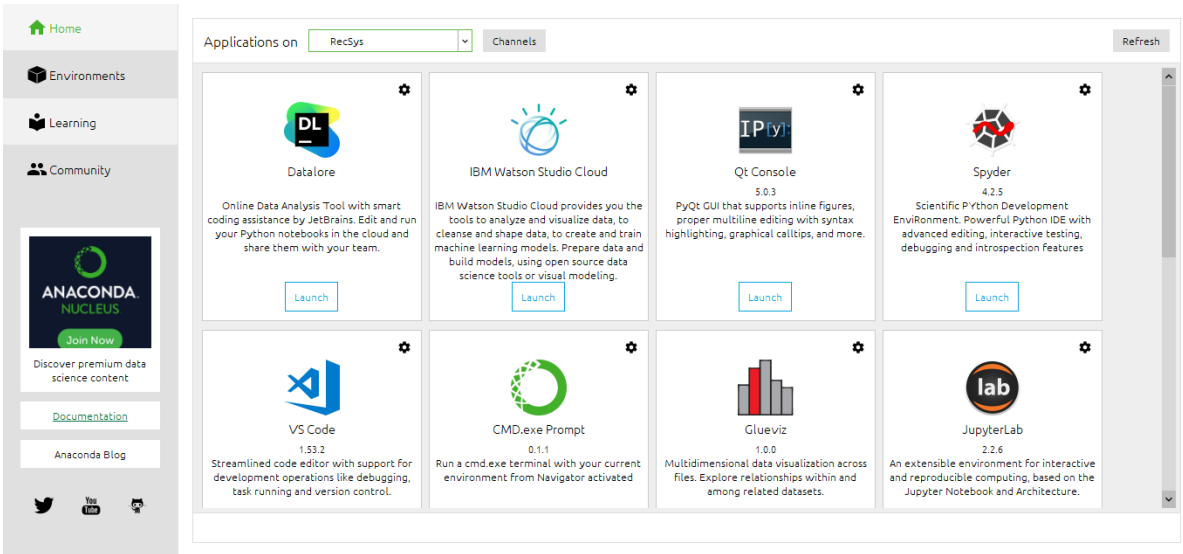
برای نوشتن تست‌ها از نرم‌افزار Spyder در Anaconda استفاده می‌کنیم. ابتدا Anaconda را باز می‌کنیم، سپس یک environment می‌سازیم به نام RecSys:



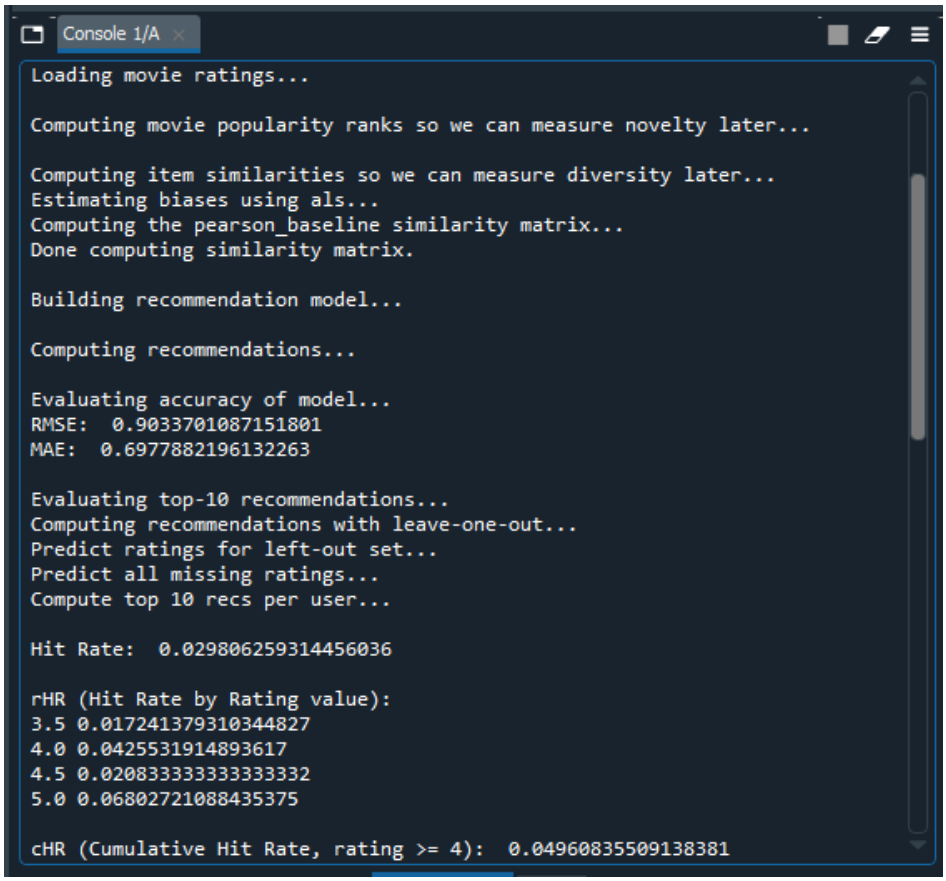
حال در ترمینال این محیط، کتابخانه‌ی Surprise را که برای استفاده از تست‌های از قبل نوشته شده‌ی سیستم پیشنهاد دهنده است، با استفاده از دستور زیر نصب می‌کنیم:

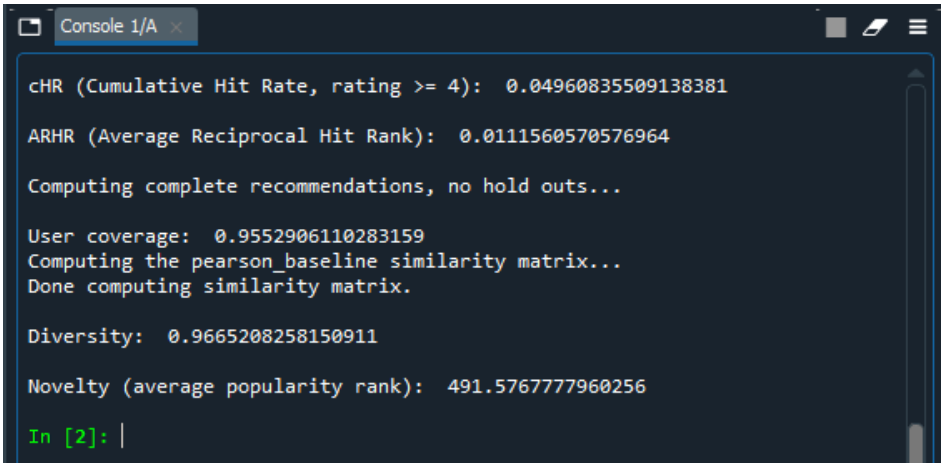
```
Conda install –c conda-forge scikit-surprise
```

سپس به صفحه‌ی Home رفته و Application on را برروی RecSys قرار داده و برنامه‌ی Spyder را باز می‌کنیم.



بعد از باز شدن Spyder کدهای Evaluating را که باز کرده و فایل TestMetrics.py را اجرا می‌کنیم، در خروجی نتیجه‌ی ارزیابی متریک‌های ذکر شده را می‌بینیم.





در این پروژه، از داده‌ی سایت Movie lens استفاده شده. کد تست‌ها و فایل دیتاست در ادامه‌ی فایل‌های این گزارش آمده است.