سوال 1)  
سیستم های توصیه گر،الگوریتم هایی هستند که سعی میکنند با استفاده از دیتاهای موجود ایتم ها را به کاربران پیشنهاد بدهند.برای مثال این دیتاها میتواند شامل سوابق خرید یک کاربر در فروشگاه،بازدید های کابر از کالا و یا فیلم های دیده شده توسط کاربر و امتیازی که کاربربه آنها داده است، باشد.

این الگوریتم ها با مشاهده این دیتاها و یادگیری الگوهای موجود در دیتا،سعی میکنند براساس سلایق و اولویت های کاربران،ایتم ها را(ایتم شامل هرچیزی میتواند باشد مثلا فیلم،کالا،موسیقی،خدمات و...) به آنها پیشنهاد بدهند.

دو دسته اصلی سیستم های توصیه گر عبارتند از:

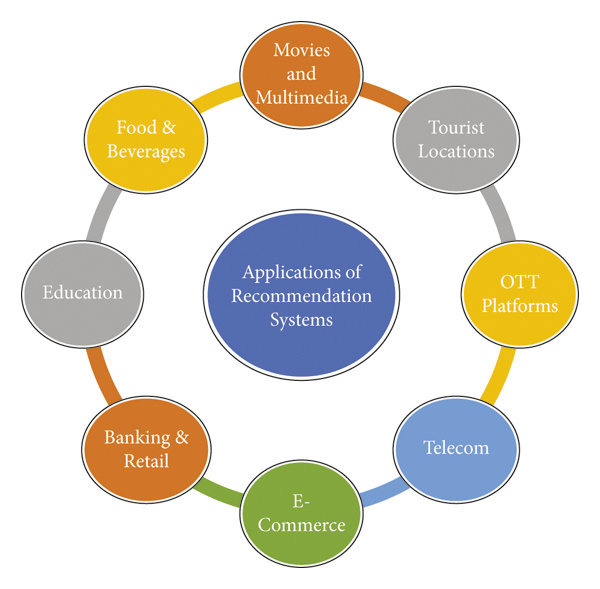
content-based و collaborative filtering

سوال 2)

برخی ازکاربرد های سیستم های توصیه گر عبارتند از:

- پیشنهاد فیلم و سریال به کاربران که نمونه آن را در پلتفرم هایی مثل نتفلیکس،آمازون پرایم و.. میتوان مشاهده کرد.

- پیشنهاد موزیک وپادکست به کاربران که برای مثال در اسپاتیفای میتوان نمونه آن را دید.

- پیشنهاد کالاها در فروشگاه های انلاین مانند امازون

- یافتن افراد مشابه و پشنهاد آنها به کاربران درشبکه های اجتماعی مانند اینستاگرام

- و....

سوال3:

چالش ها:

- یکی از چالش ها cold start است.به این معنا که وقتی یک کاربر جدید وارد سیستم میشود(مثلا وقتی تازه اکانت خود را در نتفلیکس یا اسپاتیفای ساخته اید)،دیتای مناسبی وجود ندارد تا بتوان براساس آن ایتم ها را به کاربر پشنهاد داد.البته راه حل هایی برای این چالش وجود دارد،مانند کاری که اسپاتیفای در شروع میکند.ابتدا سوالات مختلفی را از کاربر میپرسد تا به یک شناخت نسبی برسد.

-چالش دیگر سنجش عملکرد سیستم توصیه گر است.معیارهای مختلفی برای سنجش وجود دارد که بایستی با توجه به شرایط انتخاب شود.

-چالش دیگری که با آن مواجه هستیم تنک (sparse) بودن ماتریس utility میباشد.چون بطور مثال همه کاربران به فیلم هایی که دیده اند امتیاز نمیدهند(یا مسلما همه کاربران همه فیلم ها را ندیده اند) و این موضوع باعث تنک شدن ماتریس خواهد شد.

- چالش مهم دیگراین است که سلایق کاربر ممکن است درطول زمان تغییر کند ویا حتی بعضی سلایق موقتی باشند.(برای مثال به سفر رفتید و موزیک های مناسب سفر را در اسپاتیفای پلی میکنید،قاعدتا بعد از اتمام سفر نیازی به توصیه شدن این سبک موسیقی نمیباشد.)

سوال7:

a)

در روش content-based شباهت ایتم ها را با توجه به ویژگی آنها میسنجیم.برای مثال اگر دو ایتم Aو B مشابه باشند و یک کاربر ایتم A را دوست داشته باشد(مثلا ایتم یک فیلم است و کاربر به فیلم A امتیاز بالا داده)،پس احتمالا ایتم B را نیز دوست دارد.بنابراین ایتم B میتواند یک پیشنهاد به کاربر باشد.

b)

1-کارگردان 2-مجموعه بازیگران 3-ژانر 4-سال ساخت 5-امیتازفیلم(مثلا امیتازIMDB)

6-کشورسازنده 7-شرکت سازنده 8-میزان فروش 9- نویسنده 10-بودجه ساخت و میزان سود دهی

c)

Movie1= [ 0 1 1 0 1 1 0 1 3a]  
 Movie2= [ 1 1 0 1 0 1 1 0 4a]

در اینجا یکی از ویژگی ها عددی است(امیتاز) و ویژگی دیگر به صورت صفرو یکی(boolean). برای محاسبه کوسینوس زاویه میان دو بردار،فیچرهای عددی را میتوان اسکیل کرد.ضریب a را به عنوان فاکتوراسیکل در نظر میگیریم و براساس این پارامتر کوسینوس زاویه بین دو بردار با بدست میاوریم.

Cos = (2 + 12a^2)/sqrt(25 + 125a^2 + 144a^4)

-مثلا به ازای a = 1کسینوس برابر 0.816 و به ازای a=2 برابر با 0.940 خواهد بود.بنابراین بسته به این که چه مقداری برای پارامترa انتخاب کنیم مقدار کسینوس متفاوت خواهد بود.

d)

از بین فیلم هایی که کاربر مورد نظر دیده است،در 20 درصد انها جولیا رابرتز بازی کرده است.بنابراین مولفه مربوط به این بازیگر در پروفایل کاربر برابر با 0.2 خواهد بود.

e)

ابتدا امتیازی که هر کاربر داده است را نرمالایز میکنیم(میانگین امتیازات داده شده توسط هرکاربر را از امتیازی که کاربر به هر فیلم داده است را کم میکنیم)،و سپس میانگین میگیریم.عدد بدست امده مولفه بازیگر مورد نظر در پروفایل کاربر خواهد بود.

U: ( ( 3 – 3 ) + (4 – 3) +( 5 – 3)) / 3 =1

V: ( ( 2 – 4 ) + (3 – 4) +( 5 – 4)) / 3 =-2/3 = -0.67

-بنابراین مولفه مربوط به بازیگر جولیا رابرتز در پروفایل کاربران U و V به ترتیب برابر با 1 و0.67- میباشد.

f)

cos (m1, m2):

cos (m1, m3):

cos (m2, m3):

g)

a=b=1

cos (m1, m2): 1 🡪angel=0

cos (m1, m3): 1🡪 angel=0

cos (m2, m3): 0.9996🡪 angel=1.62

h)

a=0.01, b=0.5

cos (m1, m2):0.9888🡪 angel=8.5

cos (m1, m3): 0.9795🡪 angel=11.62

cos (m2, m3):0.9487🡪 angel=18.43

i)

mean of f2=416.67

mean of f3=4.3

بنابراین ضریب a و b بدین صورت خواهند بود:

a = 1/mean of f2 = 0.002

b = 1/mean of f3 = 0.232

cos (m1, m2):0.9930🡪 angel=6.78

cos (m1, m3): 0.9921🡪 angel=7.2

cos (m2, m3):0.9769🡪 angel=12.33

j)

mean of f1= 2.8

mean of f2=416.67

mean of f3=4.3

بردار های نرمالایز شده:

|  |  |  |  |
| --- | --- | --- | --- |
| m3 | m2 | m1 |  |
| 0.07 | -0.25 | 0.18 | f1 |
| 183.33 | -166.67 | -16.67 | f2 |
| 0.7 | -1.3 | 0.7 | f3 |

k)

یک راه برای اسکیل کردن بردار های نرمالایز شده،تقسیم بر انحراف معیار است.با اینکار انحراف معیار هر مولفه برابر یک خواهد بود.و چون نرمالایز شده هستند،میانگین نییز برابر صفر میباشد.)یک راه دیگر برای اسکیل کردن تقسیم بر ماکسیمم است)

داده های قسمت قبل پس از نرمالایز و اسکیل:

|  |  |  |  |
| --- | --- | --- | --- |
| m3 | m2 | m1 |  |
| 0.383 | -1.370 | 0.986 | f1 |
| 1.278 | -1.162 | -0.116 | f2 |
| 0.707 | -1.414 | 0.707 | f3 |

هر چه زاویه بین دو بردار کمترباشد فاصله کوسینوسی دو بردار کمتر بوده و به معنای شباهت بیشتر دو بردار(دو ایتم مدنظر) خواهد بود.

زاویه بین بردارها پس از نرمالایز کردن:

m1,m2: حدود صفر درجه

m1,m3:حدود 180 درجه

m2,m3:حدود 180 درجه

l)

mean of rates =(4+2+5 )/3=3.67

m1= 4-3.67=0.33

m2=2-3.67=-1.67

m3=5-3.67=1.33

m)

ساخت پروفایل کاربر:

f1 = 4\*2.98+2\*2.55+5\*2.87=31.37

f2 =4\*400+2\*250+5\*600=5100

f3 =4\*5+2\*3+5\*5=51

user profile:[31.37,5100 ,51]

همچنین میتوان امتیازات کاربر را اسکیل کرد مثلا بر ماکسیمم که 5 است تقسیم کنیم.که در این صورت داریم:

user profile:[6.274,1020 ,10.2]

سوال9:

بطور کلی،کاهش بعد روشی است که در آن سایز دیتا را کاهش میدهیم و در عین حال اطلاعات اصلی را حفظ میکنیم.

روش های مختلفی برای اینکار وجود دارد که عبارتند از:

SVD

Bayesian Clustering

Probabilistic Latent Semantic Analysis (p-LSA)

Latent Dirichlet Allocation (LDA)

UV-Decomposition

همانطور که میدانیم ماتریسUtility اسپارس است،با استفاده از تجزیه U-V تقریبی از ماتریس Utility میزنیم(با توجه به درایه های موجود در ماتریس) و سپس با محاسبه ضرب ماتریسیUV به تخمینی از درایه های خالی در ماتریس Utility میرسیم.