

# سیستم‌های پیشنهادگر

امین نظری

# فهرست مطالب

## • بخش اول مقدمه بر سیستم‌های پیشنهادگر

- معرفی سیستم‌های پیشنهادگر
- کاربردهای سیستم‌های پیشنهادگر
- مزایای و ضرورت استفاده از سیستم‌های پیشنهادگر

## • بخش دوم انواع سیستم‌های پیشنهادگر

- سیستم توصیه گر مبتنی بر محتوا
- سیستم توصیه گر فیلترینگ مشارکتی
- سیستم توصیه گر مبتنی بر دانش
- سیستم توصیه گر متن آگاه
- سیستم‌های توصیه گر مبتنی بر شبکه‌های اجتماعی
- و ...

# فهرست مطالب

- بخش سوم پیاده سازی یک سیستم پیشنهادگر

- سیستم توصیه گر فیلترینگ مشارکتی

- بخش چهارم چالشها و حوزه های پژوهشی و کاربردی

- شروع سرد

- کمبود داده ها

# معرفی سیستمهای پیشنهادگر

- افزایش روز افزون تعداد صفحات اینترنتی و تعداد کاربران
- مشکل سرریز داده‌ها
- سیستم توصیه‌گر، سیستمی هوشمند است که با تحلیل اطلاعات و رفتار کاربران، پیشنهادات مناسبی را برای هر شخص ارائه می‌کند
- حدس زدن شیوه تفکر کاربر
- صرفه جویی در وقت و انرژی به جای جستجوی کورکورانه

# مزیت سیستمهای پیشنهادگر

۱. کاربر را در اخذ تصمیم یاری کند
۲. موجب افزایش آگاهی کاربر، در زمینه موردعلاقه وی می شود
۳. معرفی محصولات جدید
۴. افزایش تعداد فروش محصولات
۵. شخصی سازی جستجو
۶. افزایش رضایت مشتری

# کاربردها سیستمهای پیشنهادگر

- تجارت الکترونیک : برای توصیه محصولات و خدمات مختلف.
- کتابخانه دیجیتال : پیدا کردن کتاب، مقاله و ...
- کاربردهای پزشکی : انتخاب پزشک متناسب با شرایط (مکان، نوع بیماری، زمان و ...)، انتخاب دارو
- گردشگری: توصیه به گردشگران برای مسافرت و بازدید از مکان‌ها مورد علاقه
- راهنمای کاربران برای استفاده از سرویس‌های موجود بر روی اینترنت: توصیه به کاربران برای استفاده از منابع مورد نیاز آن‌ها مانند سایتهایی برای آپلود عکس و ...
- توصیه اخبار، کتاب، صفحات وب، تماشای فیلم و شنیدن موزیک.

# کاربردها سیستمهای پیشنهادگر

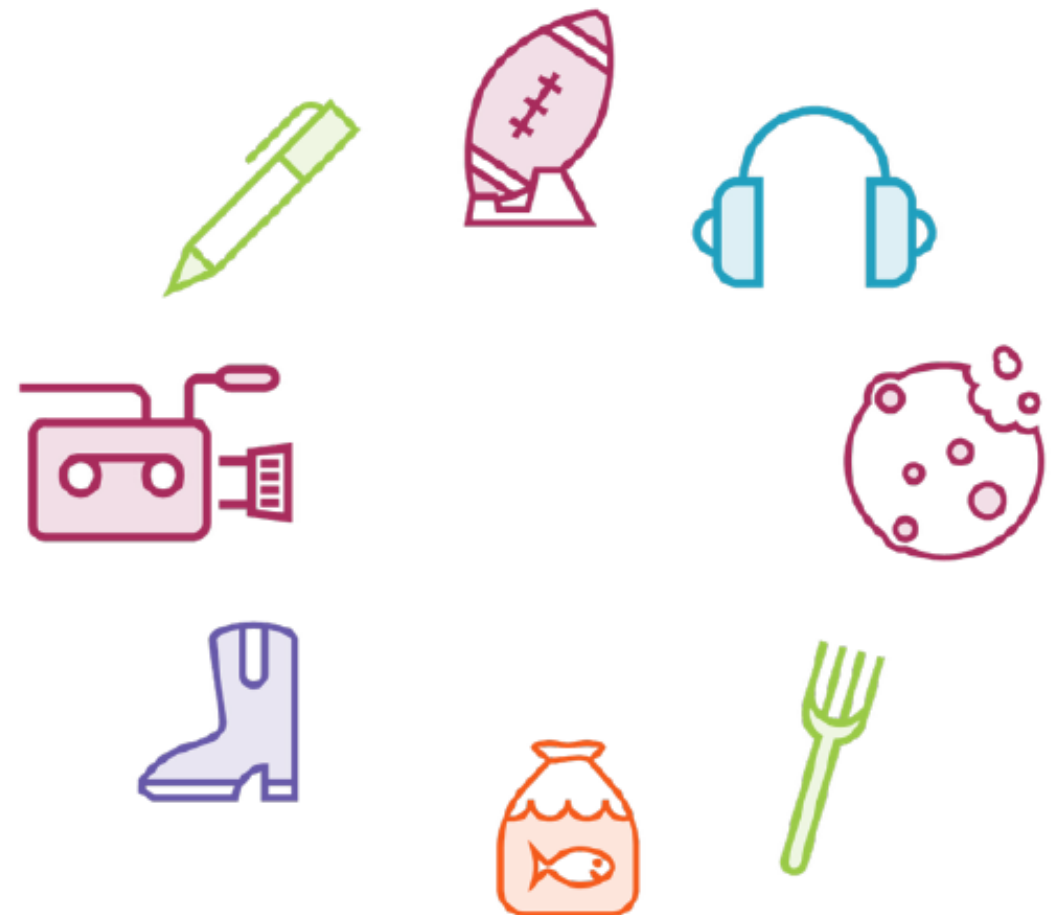
سیستم	محصول هدف
Amazon.com	Books and other products
Netflix	DVDs, Streaming Video
Jester	Jokes
GroupLens	News
MovieLens	Movies
last.fm	Music
Google News	News
Google Search	Advertisements
Facebook	Friends, Advertisements
Pandora	Music
YouTube	Online videos
Tripadvisor	Travel products
IMDb	Movies

# انواع سیستم‌های پیشنهادگر

- سیستم توصیه گر مبتنی بر محتوا
- سیستم توصیه گر فیلترینگ مشارکتی
- سیستم توصیه گر مبتنی بر دانش
- سیستم توصیه گر متن آگاه
- سیستم‌های توصیه گر مبتنی بر شبکه های اجتماعی
- و ...



# محصولات و کاربران مختلف



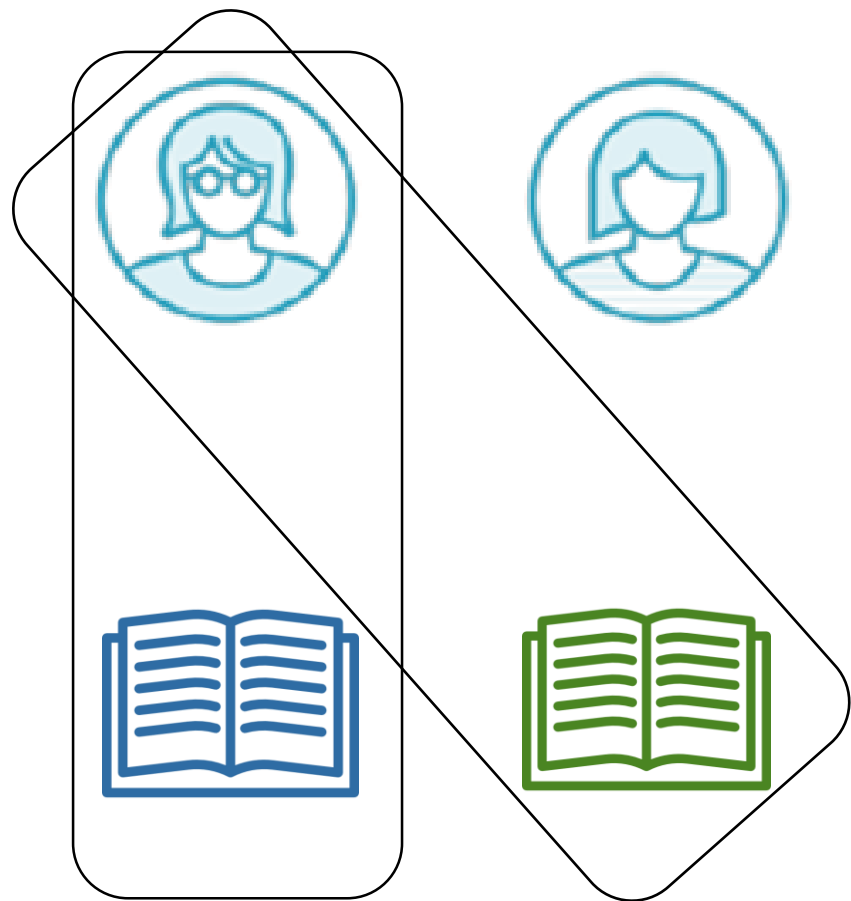
# انواع ارتباطات بین داده ها



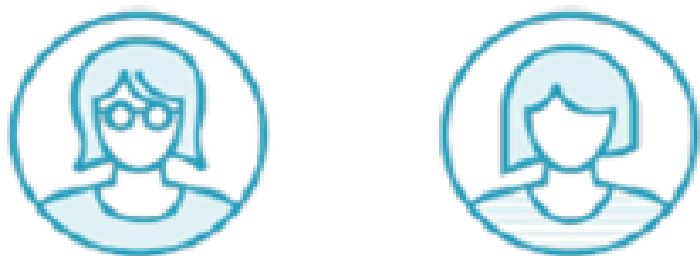
# انواع ارتباطات بین داده ها

- هر کاربر کالاهای خاصی را ترجیح میدهد

- ارتباطات بین کاربران و محصولات



# انواع ارتباطات بین داده ها

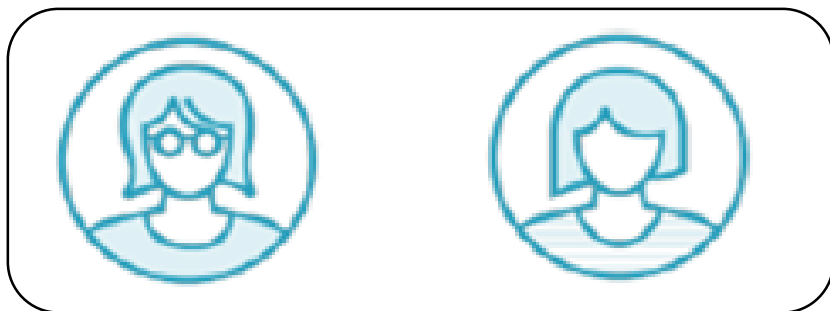


- برخی از محصولات ذاتا مشابه هم هستند
  - کتابهای با ژانر یکسان
  - اخبار با یک موضوع خاص

- ارتباطات بین محصولات و محصولات



# انواع ارتباطات بین داده ها

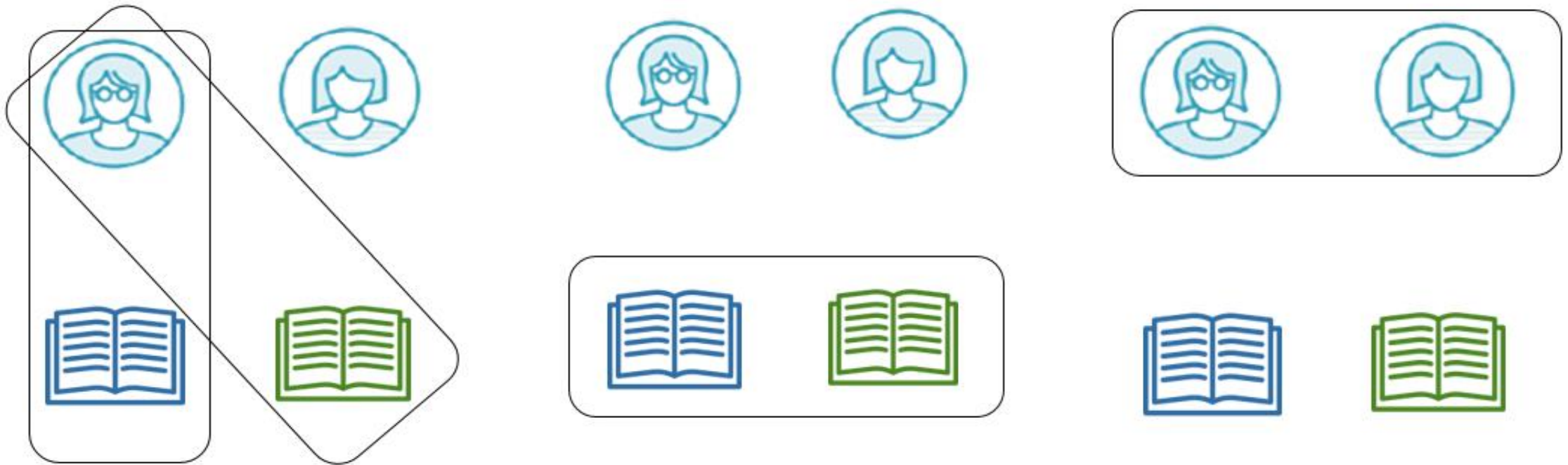


- برخی از کاربران رفتارهای مشابه بههم دارند
- کتابهای یکسان را دوست دارند
- دوستان مشترک دارند

- ارتباطات بین کاربران و کاربران



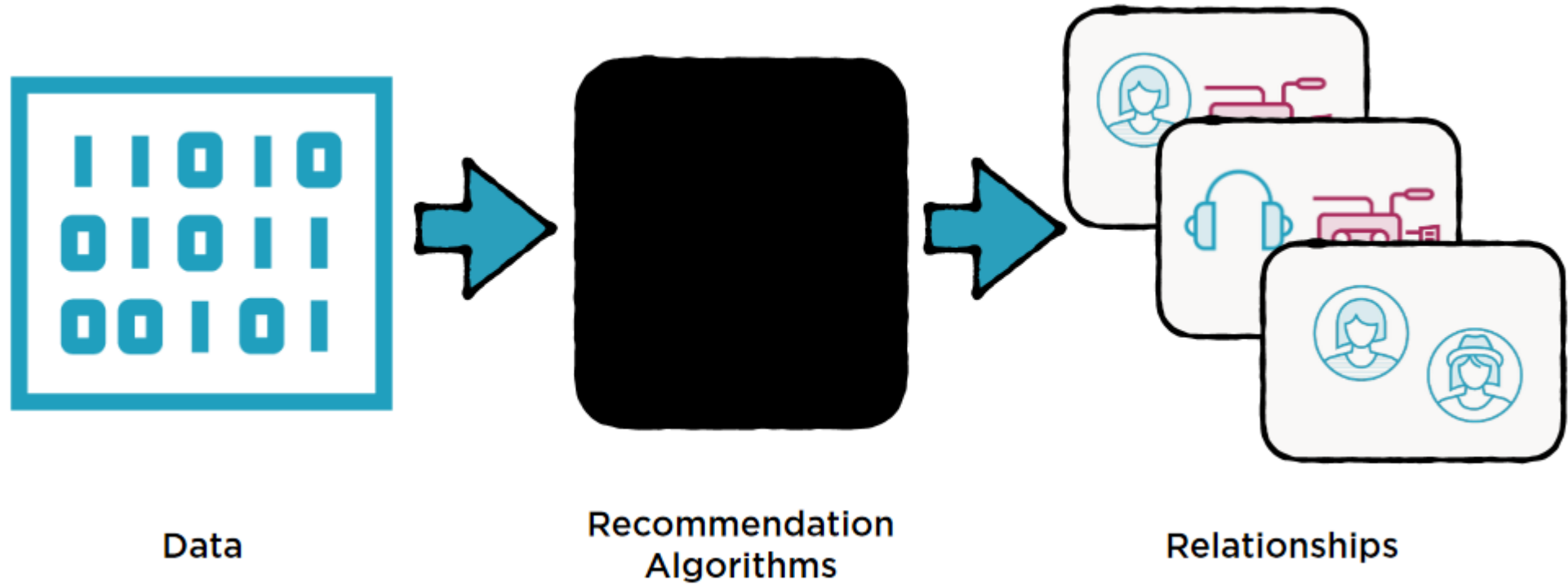
# انواع ارتباطات بین داده ها



# انواع داده ها در سیستمهای پیشنهادگر

- داده های رفتار کاربر
  - رتبه بندی، کلیک، خرید
- اطلاعات کاربر
  - سن، تحصیلات، درآمد، محل
- داده های ویژگی محصول
  - ژانر (کتاب)، بازیگران (فیلم)، غذاها (غذا)

# فرآیند کلی سیستمهای پیشنهادگر





# الگوریتم پیشنهادگر

Option 1



Find products with  
"similar" attributes

Option 2



Find products liked  
by "similar" users

Option 3



Find "complementary"  
products

# مبتمی بر محتوا

- هدف:

یافتن آیت‌های محبوب با آیت‌های کاربران مشابه

- مراحل:

۱. شناسایی آیت‌های مشابه با بررسی ویژگی‌های آیت‌ها را و گروه‌بندی آنها
۲. رتبه دهی به آیت‌هایی که توسط کاربر امتیازی نگرفته‌اند با استفاده از امتیازهای  $k$  نزدیک همسایه
۳.  $N$  آیت که بیشترین امتیاز را گرفته‌اند به کاربر پیشنهاد می‌شوند

مبتنی بر محتوا

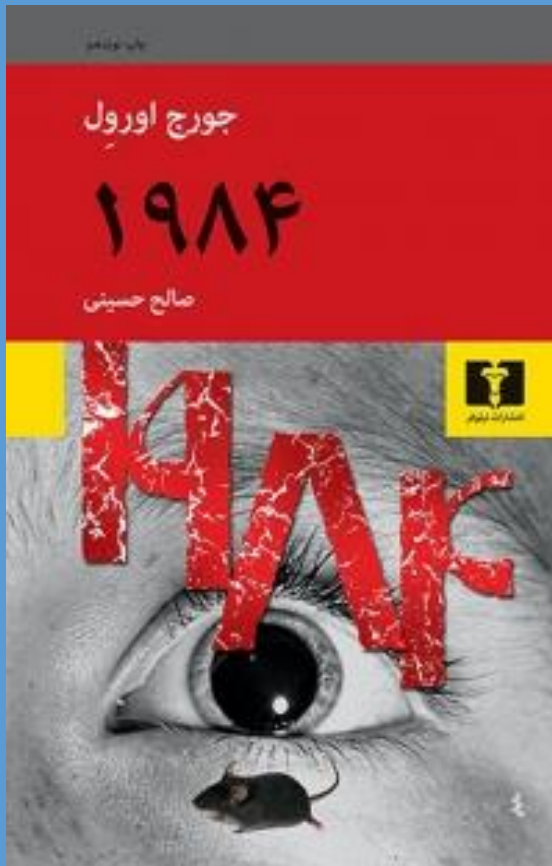


# مبتنی بر محتوا



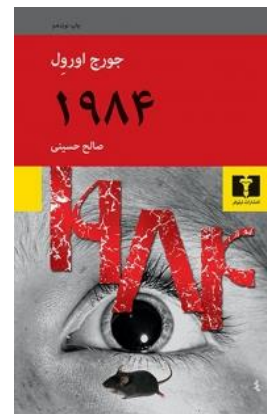
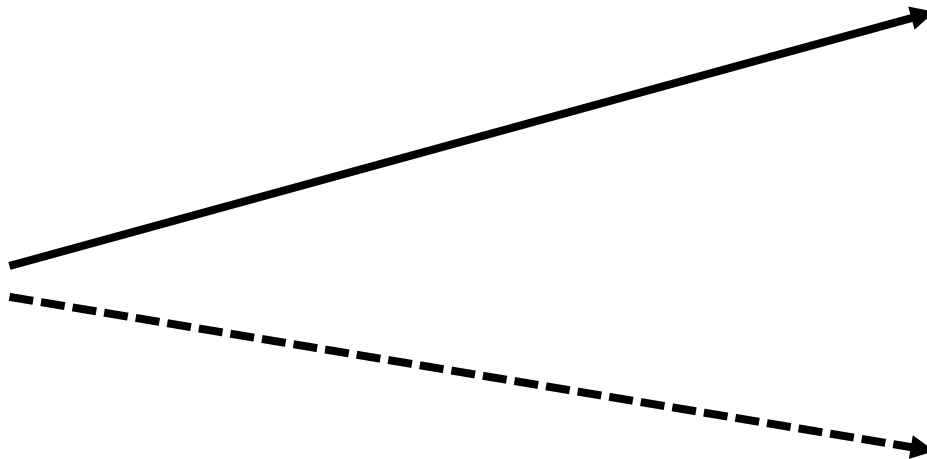
ادبیات داستانی - رمان	دسته بندی اصلی	
ادبیات انتقادی - اجتماعی - سیاسی	دسته بندی فرعی	
جورج اورول	نویسنده	
صالح حسینی - معصومه نبی زادہ	مترجم	
دوستان	انتشارات	

# مبتنی بر محتوا



ادبیات داستانی - رمان	دسته بندی اصلی	
ادبیات انتقادی - اجتماعی - سیاسی	دسته بندی فرعی	
جورج اورول	نویسنده	
صالح حسینی	مترجم	
نیلوفر	انتشارات	

# مبتنی بر محتوا



# فیلترینگ مشارکتی مبتنی بر کاربر

- هدف:

یافتن آیتمهای محبوب با شناسایی کاربران مشابه

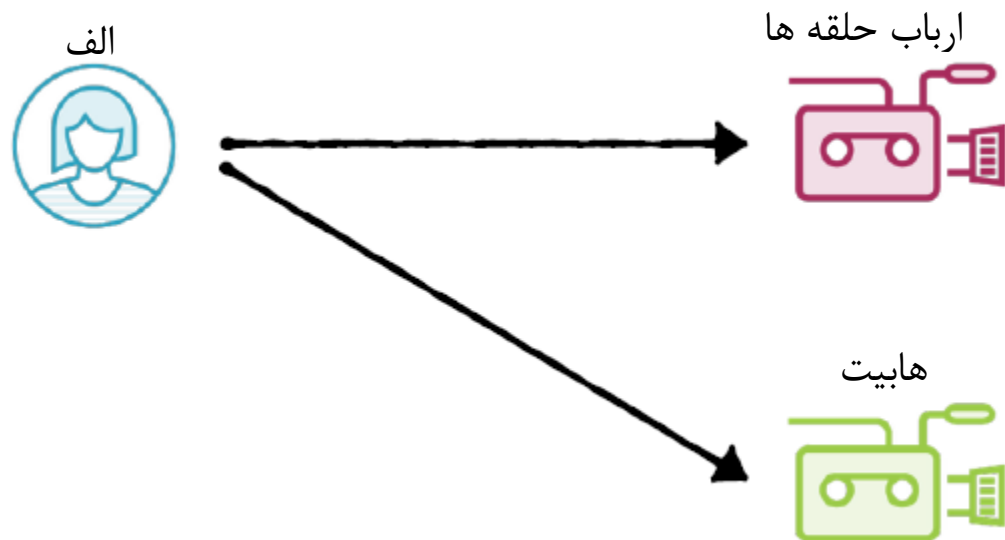
- مراحل:

۱. شناسایی کاربران مشابه از نظر علایق را و گروه‌بندی آنها
۲. رتبه دهی به آیتمهایی که توسط کاربر امتیازی نگرفته‌اند با استفاده از امتیازهای  $k$  نزدیک همسایه
۳.  $N$  آیتم که بیشترین امتیاز را گرفته‌اند به کاربر پیشنهاد می‌شوند

- کاربرد:

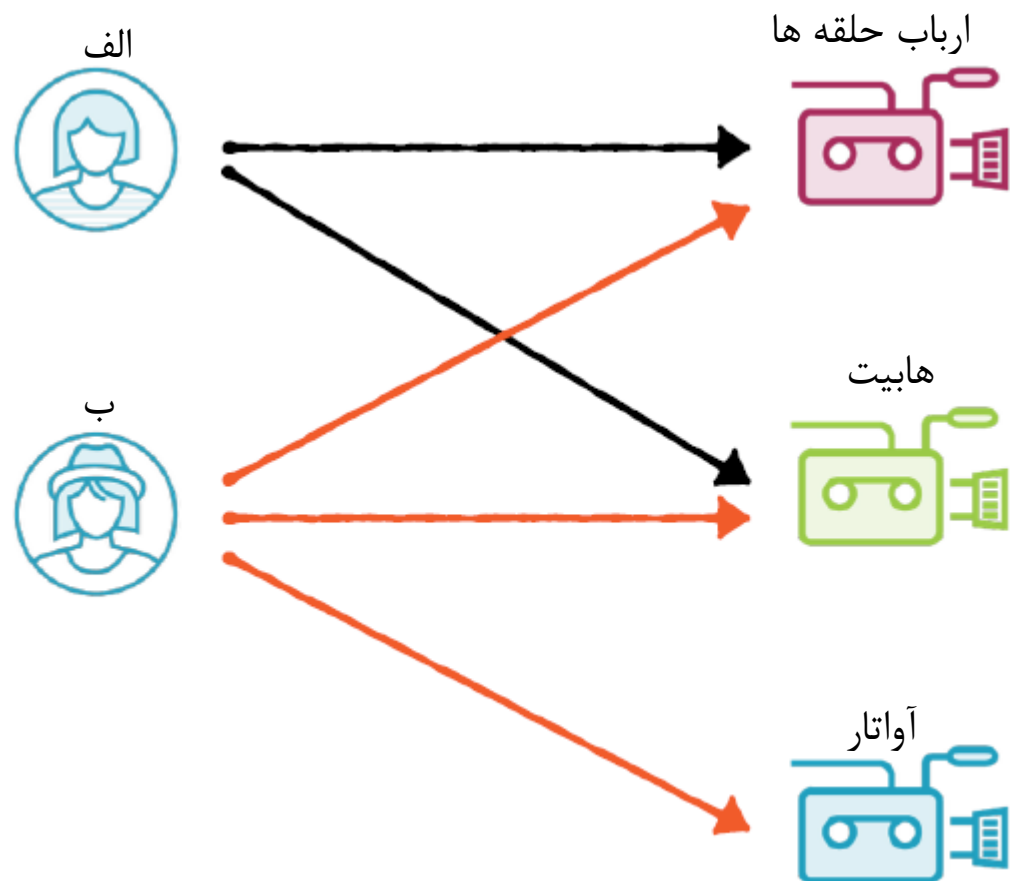
فیلتر مشترک به صورت گسترده در تجارت الکترونیک کاربرد دارد

# فیلترینگ مشارکتی مبتنی بر کاربر

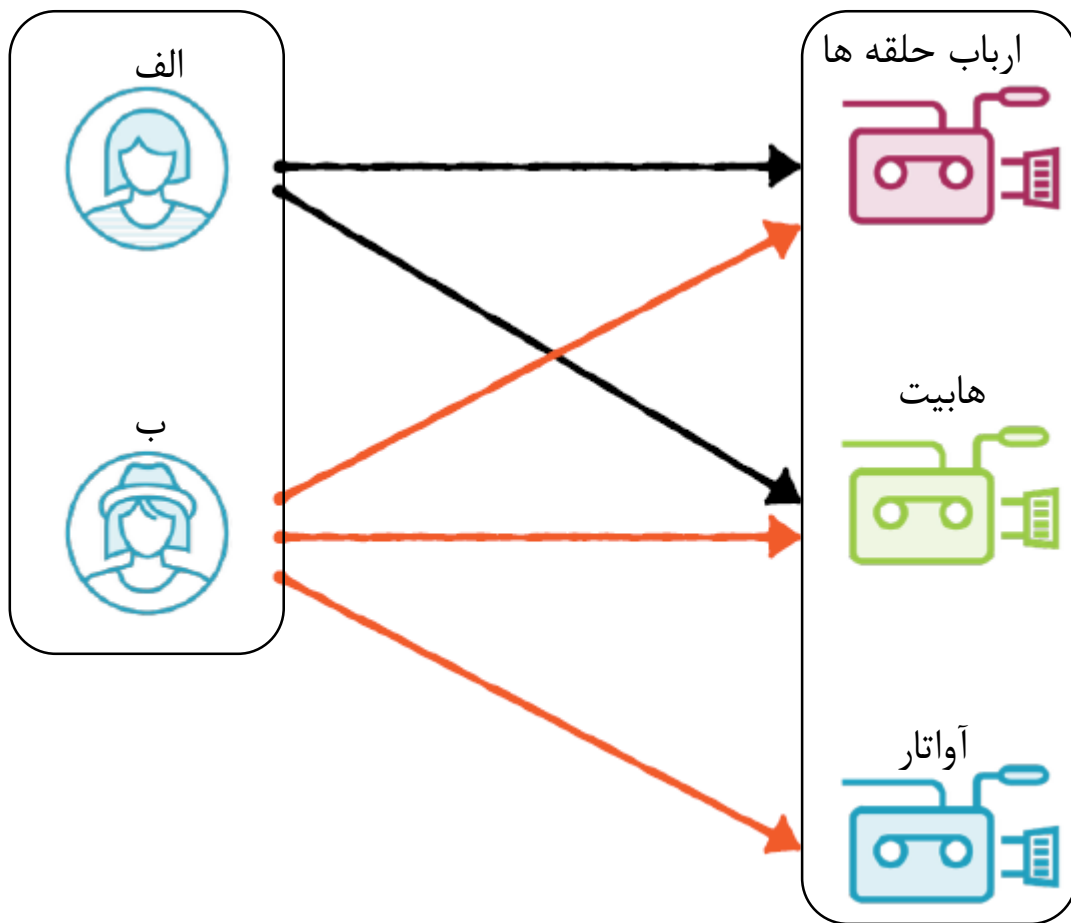




# فیلترینگ مشارکتی مبتنی بر کاربر



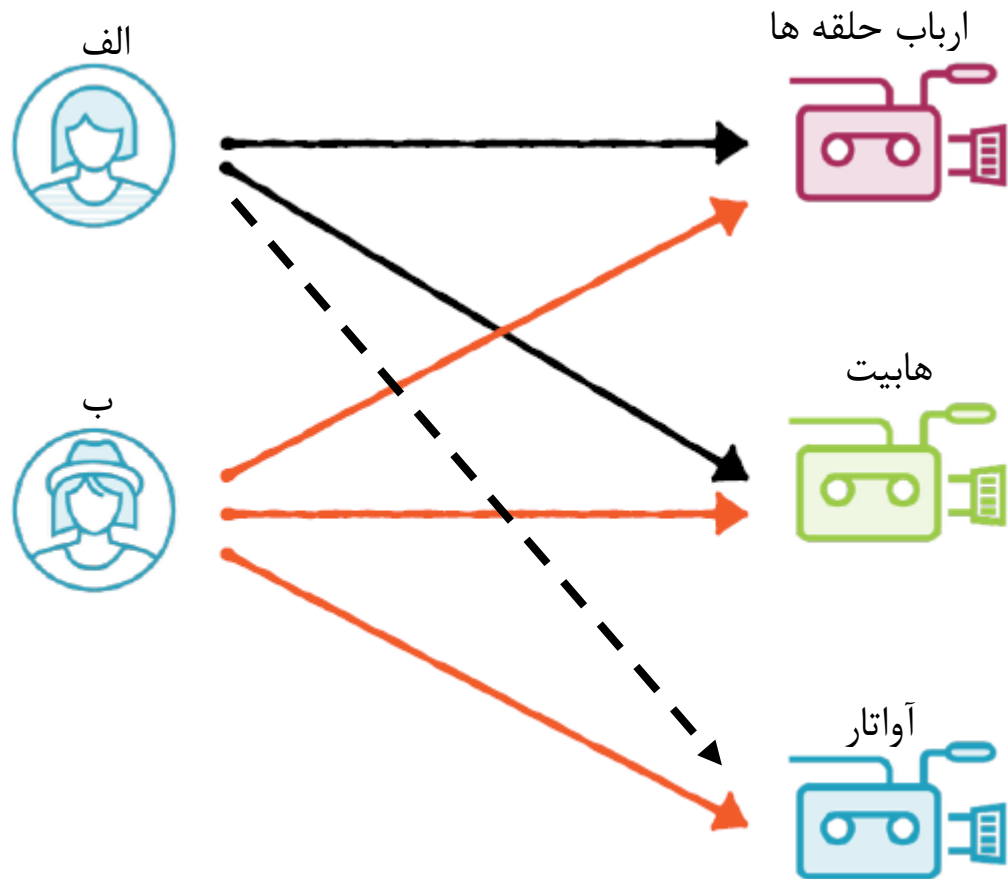
# فیلترینگ مشارکتی



۱. کاربر B مشابه با کاربر A است
۲. فیلم هایی که کاربر B دوست دارد با فیلم هایی که کاربر A دوست دارد مشابه هستند

# فیلترینگ مشارکتی

- فیلم آواتار می تواند به کاربر الف پیشنهاد شود



# داده ها در فیلترینگ مشارکتی مبتنی بر کاربر

	I1	I2	I3	I4	I5
U1	3	4	-	-	4
U2	3	5	3	4	5
U3	4	2	-	5	4
U4	3	-	4	5	2
U5	1	-	4	2	1
U6	3	4	-	2	5

# داده ها در فیلترینگ مشارکتی مبتنی بر کاربر

	I1	I2	I3	I4	I5
U1	3	4	-	-	4
U2	3	5	3	4	5
U3	4	2	-	5	4
U4	3	-	4	5	2
U5	1	-	4	2	1
U6	3	4	-	2	5

# داده ها در فیلترینگ مشارکتی مبتنی بر کاربر

	I1	I2	I3	I4	I5
U1	3	4	-	-	4
U2	3	5	3	4	5
U3	4	2	-	5	4
U4	3	-	4	5	2
U5	1	-	4	2	1
U6	3	4	-	2	5

# داده ها در فیلترینگ مشارکتی مبتنی بر کاربر

	I1	I2	I3	I4	I5
U1	3	4	-	-	4
U2	3	5	3	4	5
U3	4	2	-	5	4
U4	3	-	4	5	2
U5	1	-	4	2	1
U6	3	4	-	2	5

- میزان شباهت هر کاربر با کاربر ۱ محاسبه می شود
- K کاربر که بیشترین شباهت را با کاربر ۱ دارند انتخاب می شوند

# داده ها در فیلترینگ مشارکتی مبتنی بر کاربر

	I1	I2	I3	I4	I5
U1	3	4	-	-	4
U2	3	5	3	4	5
U3	4	2	-	5	4
U4	3	-	4	5	2
U5	1	-	4	2	1
U6	3	4	-	2	5

- میانگین امتیازات کاربران به آیتم مورد نظر به عنوان مقدار امتیاز پیش بینی در نظر گرفته می شود. (میانگین وزنی براساس میزان شباهت)
- از بین آیتمهایی که مقدار امتیاز احتمالی برای آنها پیش بینی شده است. ۱۰ آیتمی که بیشترین امتیاز را دارند به کاربر پیش بینی می شود.



# معیارهای سنجش شباهت بین کاربران

• فاصله اقلیدسی

U1	3	5	3	4	5
----	---	---	---	---	---

U2	1	5	4	2	1
----	---	---	---	---	---

•  $\sqrt{(3 - 1)^2 + (5 - 5)^2 + (3 - 4)^2 + (4 - 2)^2 + (5 - 1)^2}$

# معیارهای سنجش شباهت بین کاربران

U1	3	5	3	4	5
----	---	---	---	---	---

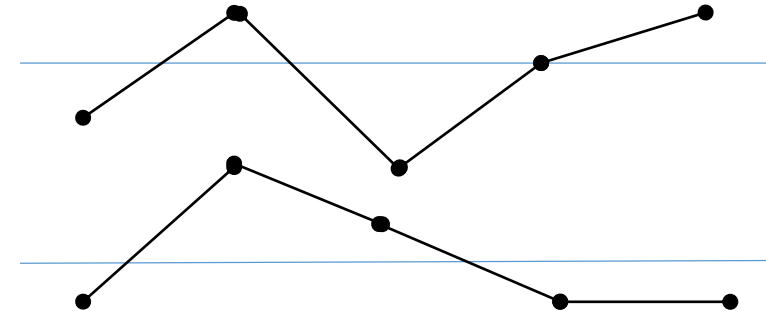
U2	1	5	4	2	1
----	---	---	---	---	---

- فاصله همبستگی
- ضریب همبستگی پیرسون

$$\text{Corr}(x, y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2} \sqrt{\sum (y_i - \bar{y})^2}}$$

# معیارهای سنجش شباهت بین کاربران

U1	3	5	2	5	5	$\bar{x}$	4
U2	2	5	4	2	2	$\bar{y}$	3



$$\frac{(-1 * -1) + (1 * 2) + (-2 * 1) + (1 * -1) + (1 * -1)}{\sqrt{(3 - 4)^2 + (5 - 4)^2 + (2 - 4)^2 + (5 - 4)^2 + (5 - 4)^2} + \sqrt{(2 - 3)^2 + (5 - 3)^2 + (4 - 3)^2 + (2 - 3)^2 + (2 - 3)^2}}$$

# معیارهای سنجش شباهت بین کاربران

• شباهت کسینوسی

U1	3	5	2	5	5	$\bar{x}$	4
U2	2	5	4	2	2	$\bar{y}$	3

$$\text{Cosine Similarity : } \text{Sim}(u_i, u_k) = \frac{r_i \cdot r_k}{|r_i||r_k|} = \frac{\sum_{j=1}^m r_{ij}r_{kj}}{\sqrt{\sum_{j=1}^m r_{ij}^2 \sum_{j=1}^m r_{kj}^2}}$$

# معیارهای سنجش شباهت بین کاربران

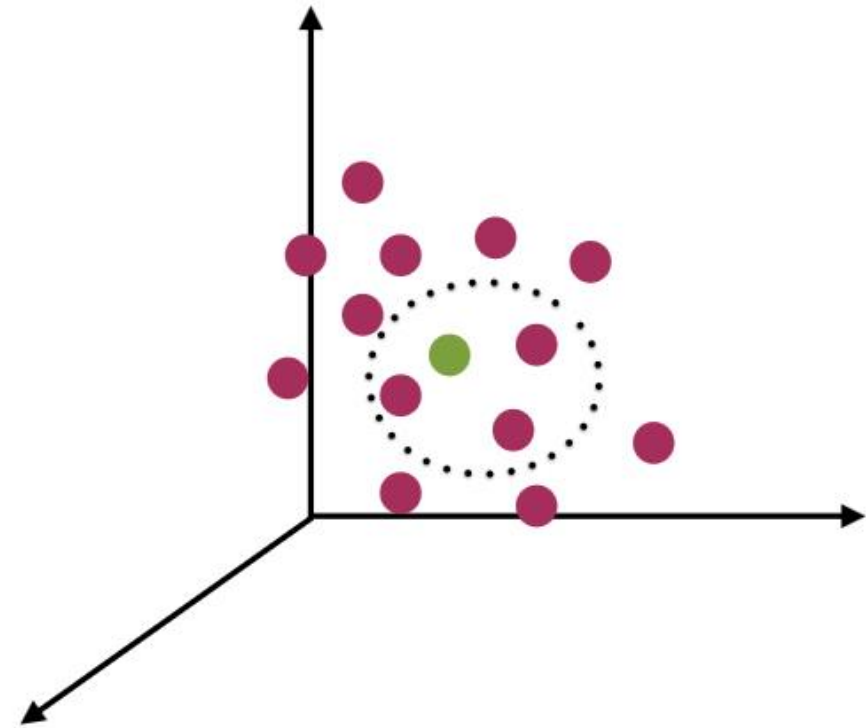
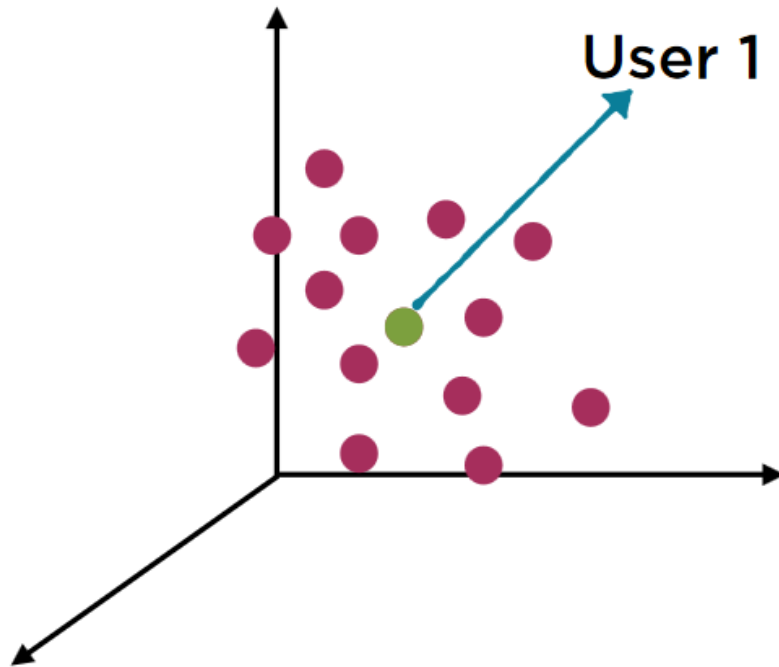
• فاصله همینگ

U1	3	5	2	5	5	$\bar{x}$	4
U2	2	5	4	2	2	$\bar{y}$	3

پیش بینی در فیلترینگ مشارکتی مبتنی بر کاربر

$$p_{a,i} = \bar{r}_a + \frac{\sum_{u \in K} (r_{u,i} - \bar{r}_u) \times w_{a,u}}{\sum_{u \in K} w_{a,u}}$$

## مدل نزدیکترین همسایه



# سیستم پیشنهادگر مبتنی بر دانش

- هدف:

کشف قوانین انجمنی و ارائه پیشنهادات براساس آن

- مراحل:

۱. شناسایی کاربران مشابه از نظر علایق را و گروه‌بندی آنها

۲. رتبه دهی به آیتم‌هایی که توسط کاربر امتیازی نگرفته‌اند با استفاده از امتیازهای  $k$  نزدیک همسایه















۳.  $N$  آیتم که بیشترین امتیاز را گرفته‌اند به کاربر پیشنهاد می‌شوند

- کاربرد:

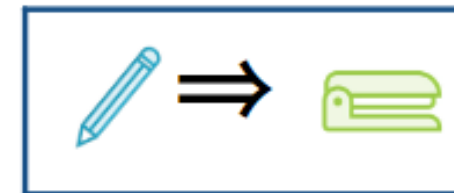
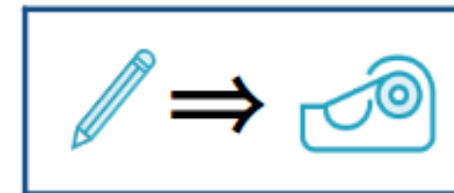
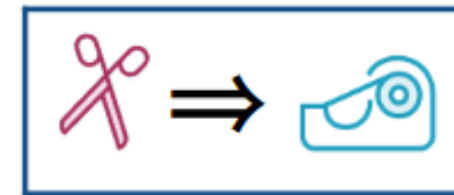
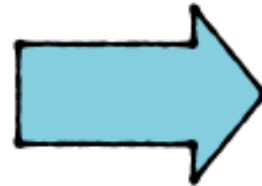
فیلتر مشترک به صورت گسترده در تجارت الکترونیک کاربرد دارد



# کشف قوانین انجمنی

1	  
2	  
3	   
4	 
5	 

Transactions



Rules

# کشف قوانین انجمنی

## Measuring Rule Strength



What **proportion** of all transactions contain both items?



$$P(\text{headphones}, \text{smartphone}) = 1\%$$

Support

# کشف قوانین انجمنی

## Measuring Rule Strength



Out of all transactions with  how many include  ?

$$P(\text{headphones} / \text{smartphone}) = 5\%$$

$$\text{Confidence} \quad P(\text{headphones} / \text{smartphone}) = \frac{P(\text{headphones}, \text{smartphone})}{P(\text{smartphone})}$$

# چالش های اصلی سیستم توصیه گر

- مشکل شروع سرد
  - داده جدید
  - کاربر جدید
  - خلوت بودن داده ها
- قابلیت مقیاس پذیری
- ارزش زمان
- رابط کاربر
- حریم خصوصی