



Recommendation Systems (Tavsiye Sistemleri)

- 1) Association Rule Learning
- 2) Content Based Recommendation
- 3) Item Based Collaborative Filtering
- 4) User Based Collaborative Filtering
- 5) Model Based Matrix Factorization

* Tavsiye Sistemleri: Kullanıcılara bazı teknikleri kullanarak ürün ya da hizmet önermek, tavsiye etmektir. 1990'dan beri vardır. e-ticaret, sosyal medya kanalları vb.

* İçerik bazlı olduğu için filtreleme yapmak gerektir.

* Ürün, sayfa, film, video içerik önerileri gibi ürün satma, içerik pazarlama gibi bir çok uygulamada kendine yer bulmaktadır.

• Simple Recommender Systems (Basit Tavsiye Sistemleri)

→ İş bilgisi ya da basit tekniklerle yapılan genel öneriler. (Bu kısma değinilmeyecek)

→ Kategorinin en yüksek puanları, trend olanlar, etkenler.

• Association Rule Learning (Birlikte Alınması Gerekeni)

→ Sepet Analizi

→ Birlikte alınması öğrenilen kurallara göre ürün önerileri

• Content Based Filtering (İçerik Temelli Filtreleme)

→ Ürün benzerliğine göre öneriler yapılan özellik temelli yöntemler

• Collaborative Filtering (İş Birlikli Filtreleme)

→ Topluğun kullanıma ya da ürün bazında ortak konuları yansıtan yöntemler

→ User-Based

→ Item-Based

→ Model-Based (Matrix Factorization)



Association Rule Learning (Birlikte Alınan Kurallı Öğrenimi)

Veri içerisindeki benzerlikleri (pattern, ilişki, yapı) bulmak için kullanılan kural tabanlı bir makine öğrenmesi tekniğidir.
Ekmek olan sütte oldu gibi.

Apriori Algoritması : Sepet analizi yöntemidir. Ürün benzerliklerini ortaya çıkarmak için kullanılır.

$$\text{Support}(X, Y) = \text{Freq}(X, Y) / N$$

(X ve Y'nin birlikte görüme olasılığı)

$$\text{Confidence}(X, Y) = \text{Freq}(X, Y) / \text{Freq}(X)$$

(X satın alındığında Y'nin satın alınması olasılığı)

$$\text{Lift} = \text{Support}(X, Y) / (\text{Support}(X) * \text{Support}(Y))$$

(X satın alındığında Y'nin satın alınma olasılığı lift kat kadar artar)

X'in tek başına
görülme olasılığı

Y'nin tek
başına görüme
olasılığı

<u>Örnek</u>	<u>İşlem</u>	<u>Örnekler</u>	
1		Ekmek, Süt, Kola, Peynir	$\text{Support}(\text{Ekmek}) = \frac{4}{8}$
2		Ekmek, Süt, Kola	
3		Ekmek, Süt	$\text{Support}(\text{Süt}) = \frac{6}{8}$
4		Ekmek, Soğan	
5		Un, Süt, Kola, Peynir	$\text{Support}(\text{Ekmek, Süt}) = \frac{3}{8}$
6		Un, Süt, Kola	
7		Un, Süt	$\text{Confidence}(\text{Ekmek, Süt}) = \frac{\text{Freq}(\text{Ekmek, Süt})}{\text{Freq}(\text{Ekmek})} = \frac{3}{4}$
8		Un, Soğan	

$$\text{Lift}(\text{Ekmek, Süt}) = \frac{\text{Support}(\text{Ekmek, Süt})}{\text{Support}(\text{Ekmek}) * \text{Support}(\text{Süt})} = \frac{\frac{3}{8}}{\frac{4}{8} * \frac{6}{8}} = 1 \rightarrow 1'den daha yüksek olması beklenir.$$



Subject :

Unstack()

Date :/...../.....

→ Erkek satın alındığında s+t satın alma olasılığı 1 kat artmak kadar denir. 1 olması kayda değer olduğu anlamına gelir. Fakat pratikte 1'den daha yüksek sayıda olmasını bekleriz. Yani demek alındığında s+t satın alma olasılığı 5 kat arttığı gibi mesela.

Leverage → kaldıraç etkisi

supportu yüksek olan değerlere öncelik verir bu yüzden yarıdır (biraz)

Lift → yansız bir metriktir.

conviction → y \bar{X} olmanın x \bar{X} olmasının beklenen değeri frekansıdır. ya da x \bar{X} olmadan y \bar{X} BD ya da frekansı.

Content Based Filtering (İçerik Temelli Filtreleme) 12.11.22

İçerik özelliklerinin benzerlikleri üzerinden tavsiyeler getirilir.

İçerik özellikleri = İçeriğin meta bilgisi demek.

İçerik = film, kitap, e-ticaret şirketlerinden alınan her türlü malzeme,

İçerik : açıklama

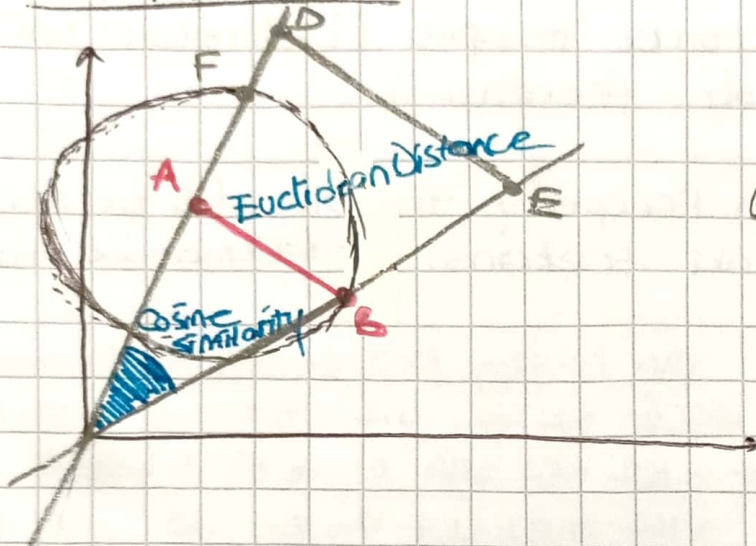
Euclidean Distance (Eklid uzaklığı)

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

Cosine Similarity (Kosinüs Benzerliği)

$$\text{Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \cdot \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}}$$

Geometrie G8stern.



Cosine similarity \rightarrow a cosol

Allele better have gotten

Metinleri matematiksel ifadelere sokmak Btlyovm.
Nasil yapabiliriz?

Adım 1: Eşsiz tüm terimleri sütunlara, b^k 'in de k maddesi satırlara yerleştir.

This is the first document

This document is the second document

And this is the third one

Is this the first document

this is the first document ^{second addition} one

1	1	1	1	1					
1	1	1		2	1				
1	1	1					1	1	1
1	1	1	1	1					

Adım 2 = Terimlerin Dokümanlarda Geçme frekansını
hesablamaya yerleştir.

1.0000 1.0000 1.0000 1.5108 1.2231 1.9163 1.9163 1.9163 1.9163

IDF button dökümana islem yaptır.
TF cümneye



Subject : İkinci cümle

Date :/...../.....

Adım 4 : TF * IDF 'i hesapla

this is the first document second and third one

0,2	0,2	0,2	0,3	0,24463	0	0	0	0
0,167	0,167	0,167	0	0,40771	0,3194	0	0	0
0,167	0,167	0,167	0	0	0	0,32	0,32	0,32
0,2	0,2	0,2	0,3	0,24463	0	0	0	0

Adım 5 : L2 Normalizasyon Yap

Satırların kareleri toplamının karekökünü al, ila ila satırları
film kareleri bulduğun değere böl.

this is the first document second and third one

									Karelerin toplamı karekökü
0,384	0,384	0,384	0,58	0,46979	0	0	0	0	0,5207
0,281	0,281	0,281	0	0,68762	0,5386	0	0	0	0,5929
0,267	0,267	0,267	0	0	0	0,51	0,51	0,51	0,6239
0,384	0,384	0,384	0,58	0,46979	0	0	0	0	0,5207

this is the first document second and third one

$$\sqrt{(0,2)^2 + (0,2)^2 + (0,2)^2 + (0,3)^2 + (0,24463)^2 + 0^2 + 0^2 + 0^2 + 0^2} = 0,5207$$

$$\frac{0,2}{0,5207} = 0,384 \quad \checkmark$$

{ Manhattan
Pearson
cosine similarity



Subject :

Date :/...../.....

Collaborative Filtering (İş Birlikli Filtreleme)

- Item-Based Collaborative Filtering) memory Based
- User Based
- Model Based

Item-Based Collaborative Filtering : Item benzerliği üzerinden öneriler yapılır. Beğenilme yapısına benzer bir film öneririz.

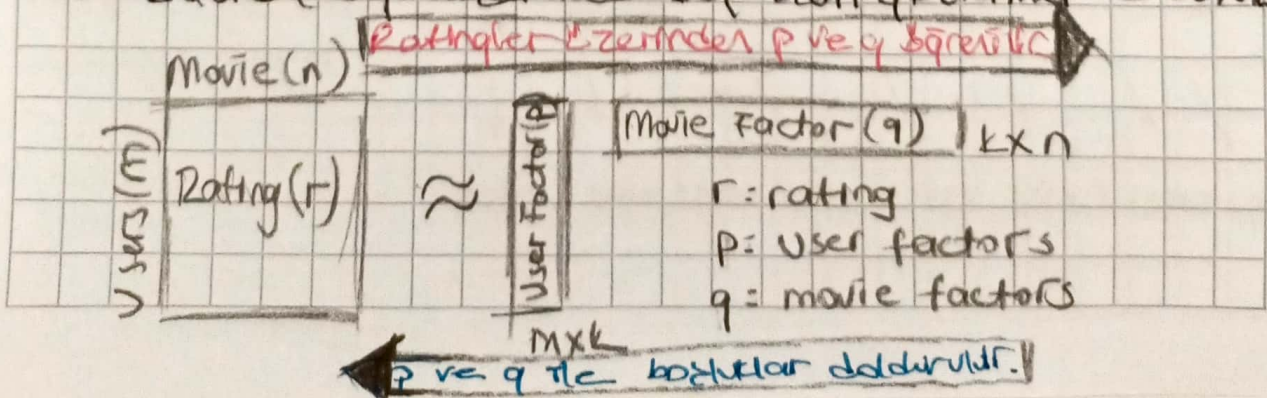
İki film beğenilmesi arasındaki benzerlik ilişkisi bizim film öneririz.

User-Based Collaborative Filtering : kullanıcı benzerlikleri üzerinden öneriler yapılır.

Matrix Factorization (Model Based)

Böylelikle doldurmak için user'lar ve movie'ler için var olduğu varsayılan latent featureların açıkları var olan veri üzerinden bulunur ve bu açıklar ile var olmayan gözlemler için tahmin yapılır.

- User - Item matrisini 2 tane daha az boyutlu matrise ayırır.
- 2 matristen User-Item matrisine ^{gizli} gizli latent factor'lar ile gerçekleştirildiği varsayımında bulunur.
- Daha olan gözlemler üzerinden latent factorlerin ağırlıkları bulunur.
- Bulunan ağırlıklar ile boş olan gözlemler doldurulur.



Subject :

Date :/...../.....

Matrix Factorization

■ Rating matrixin iki factor matrixin carpimi (dot product) ile olustugu varsayilir.

■ Factor matrixler? user latent factors, movie latent factors

■ Latent factors? Latent features? Gizli faktörler ya da deęişkenler. Örneğin filmi beğenmesincedeki eden bazı sebepler. Yönetmeni sever, oyuncuyu sever, film türünü sever, IMDB puanı yüksekliği etki edebilir vs.

■ kullanıcıların ve filmlerin latent feature'lar için skorlara sahip olduğu düşünülür.

■ Bu öğütükler (skorlar) önce var olan veri üzerinden bulunur ve daha sonra BLANK (boş) hümler bu öğütüklerle göre doldunılır.

→ bazı gizli faktörler var ve bunların öğütükleri var. Nasıl bulacağız?

(p) User factors: User \times Latent factor

(q) Movie factors: Latent factor \times movie

Date: / /

	M_1	M_2	M_3	M_4
U1	1	2		4
U2	5		3	3
U3	3	5	5	
U4		2	1	4

~

	F_1	F_2
U1	P_{11}	P_{12}
U2	P_{21}	P_{22}
U3	P_{31}	P_{32}
U4	P_{41}	P_{42}

 $m \times k$

	M_1	M_2	M_3	M_4
F1	q_{11}	q_{12}	q_{13}	q_{14}
F2	q_{21}	q_{22}	q_{23}	q_{24}

 $k \times n$

$$r_{11} = P_{11} * q_{11} + P_{12} * q_{21}$$

$$r_{21} = P_{21} * q_{11} + P_{22} * q_{21}$$

$$r_{33} = P_{31} * q_{13} + P_{32} * q_{23}$$

$$1 = P_{11} * q_{11} + P_{12} * q_{21}$$

$$5 = P_{21} * q_{11} + P_{22} * q_{21}$$

$$5 = P_{31} * q_{13} + P_{32} * q_{23}$$

Peki p ve q değerleri nasıl bulunacak?

→ Var olan değerler üzerinden iteratif şekilde tüm p ve q 'lar bulunur ve sonra kullanılır.

→ Başlangıçta rastgele p ve q değerleri ile rating matrisindeki değerler tahmin edilmeye çalışılır.

→ Her iterasyonda hatalı tahminler düzeltilenerek rating matrisindeki değere yaklaşılmaya çalışılır.

→ Örneğin bir iterasyonda 5'e 3 denildiyse sonrakinde 4 sonrakinde 5 denir.

→ Böylece belirli bir iterasyon sonunda p ve q matrisleri doldurulmuş olur.

→ Var olan p ve q 'lara göre boş gözlemler için tahmin yapılır.

Subject:

Filmlerde oğrulukların olduğunu varsayalım, onları kullanarak boş qızemleri doldurmak istiyoruz.

$$\begin{aligned} r_{13} &= p_{11} \times q_{13} + p_{12} \times q_{23} \\ &= 0.8 \times 0.1 + 0.5 \times 4.2 \\ &= \underline{\underline{2.18}} \end{aligned}$$

ortık kullanıcı 113 odu filmi izlemesi bize oğruluklardan o filmi kaç puan verebileceğini bulduk.

RMSE, MSE

Hata kareleri minimize etmeye çalışır p ve q değerlerine erişmeye çalışırız. Bunun için oğrulukları ihtiyacımız var. Bunu bulmanın bir yolu da Gradient Descent kullanmaktır.

Matematiksel olarak karşılığı :

Rating associated with user u , item i $\rightarrow r_{ui}$

Regularization parameter $\rightarrow \lambda$ Overfit'i önlemek için düzeltme yapar

$$\min_{q^*, p^*} \sum_{(u,i) \in K} (r_{ui} - q_i^T p_u)^2 + \lambda (\|q_i\|^2 + \|p_u\|^2)$$

Find the minimum matrix q, p

Loop over pair of users and items

User vector u

Item vector i

Magnitude of user vector p

Magnitude of item vector q

Gradient Descent (Gradyon İnz)

* Gradient Descent fonksiyon minimizeasyonu için kullanılan bir optimizasyon yöntemidir.

* Gradyonun negatif olarak tanımlanan 'en dik in' yönünde iteratif olarak parametre değerini güncelleyerek ilgili fonksiyonun minimum değerini verecek parametreleri bulur.

→ Bir fonksiyonun bir noktadaki +3'ü o fonksiyonun maksimum artış yönünü verir. Aynı yönün negatif olan yöne doğru gittiğimizde, ilgili fonksiyon minimize edecek şekilde parametre değerlerini güncelliyor oluruz.

yakınlaşana kadar tekrar et
Repeat until convergence {

$$\theta_j \leftarrow \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

hız θ 'nin t'rewine göre

Hatamızı minimize etmeye çalışıyoruz.

p ve q iteratif bir şekilde optimize etmeye çalışıyoruz.

$$p_{ki}^{t+1} = p_{ki}^t + 2 \underbrace{\gamma}_{\text{error}} (r_{ij} - p_i^t q_j^t) q_{kj}^t$$

$$q_{kj}^{t+1} = q_{kj}^t + 2 \underbrace{\gamma}_{\text{learning rate}} (r_{ij} - p_i^t q_j^t) p_{ki}^t$$

formül iteratif olarak
değerlik
gösterilebilir.