

2020-2021 Güz Yarıyılı

Algoritma Analizi

Dönem Projesi

Ders Yürütücüleri

Doç. Dr. M. Elif KARSLIGİL

Dr. Öğr. Üyesi M. Amaç GÜVENSAN

Konu : Kitap Öneri Sistemi

Günümüzde Youtube, Netflix, Amazon, Pinterest gibi internet ortamında milyonlarca kullanıcısı olan pek çok firma makine öğrenmesi tabanlı tavsiye sistemleri ile kullanıcılara kişiselleştirilmiş öneriler sunmaktadır.

Bu çalışmada işbirlikçi filtre (collaborative filtering) yöntemi ile bir kişinin önceki seçimlerine bakarak yeni kitap öneren bir sistem tasarlanacak ve gerçekleştirilecektir.

Sistem 3 alt bölümden oluşmaktadır :

1. **Bir okuyucunun diğer okuyuculara benzerliğinin hesaplanması :** Bir okuyucunun diğer okuyuculara benzerliği *pearson katsayısı(pearson coefficient)* kullanılarak hesaplanacaktır. Hesaplama yapılırken iki kişinin de okuduğu kitaplar için işlem yapılacaktır. Pearson katsayısı kullanılarak a ve b kişilerinin benzerliği aşağıdaki şekilde hesaplanır :

$$sim(a, b) = \frac{\sum_{p \in P} (r_{a,p} - \bar{r}_a)(r_{b,p} - \bar{r}_b)}{\sqrt{\sum_{p \in P} (r_{a,p} - \bar{r}_a)^2} \sqrt{\sum_{p \in P} (r_{b,p} - \bar{r}_b)^2}}$$

P : hem a hem b tarafından okunmuş kitaplar

$r_{a,p}$: a'nın p kitabı için verdiği puan

$r_{b,p}$: b'nin p kitabı için verdiği puan

\bar{r}_a : a'nın okuduğu kitapların puan ortalaması

\bar{r}_b : b'nin okuduğu kitapların puan ortalaması

Benzerlik değeri -1, +1 arasındadır.

2. **Bir okuyucunun en benzer olduğu k kişinin belirlenmesi :** Verilen bir k değeri için bir önceki hesaplamanın sonucuna göre, okuyucuya en benzer **k** okuyucuyu bulunuz. En benzer k kişiyi bulmak için bütün elemanları sıralamaktan daha efektif bir çözüm planlayınız.
3. **Okuyucuya yeni kitap önerme :** Okuyucuya kitap önermek için en benzer k kişi için aşağıdaki eşitlik ile hesaplama yapınız . Bu eşitlikte **N** değeri kullanıcının en çok benzediği **k** kişiyi göstermektedir Hesaplanan $pred(a,p)$ değeri kişinin okumadığı bir kitaba vereceği puanın tahmini değeridir.

$$pred(a, p) = \bar{r}_a + \frac{\sum_{b \in N} sim(a, b) * (r_{b,p} - \bar{r}_b)}{\sum_{b \in N} sim(a, b)}$$

Buna göre bir okuyucunun okumadığı bütün kitaplar için verebileceği puan tahmin edilerek en yüksek puanı alan kitabı öneriniz.

! Güncelleme: RecommendationDataSet için okunmasında kolaylık olması açısından CSV formatı kullanılacaktır. Dosya ismi **RecomendationDataSet.csv** olacak şekildedir. Dosya ismini de değiştirmeyiniz. Ayrıca kullanıcıdan alınması gereken girdileri lütfen kodunuza hard-coded koymayınız, konsolda kullanıcıdan isteyiniz. Örnek bir çalışma senaryosu aşağıdaki gibi tasarlanabilir.

RecomendationDataSet.csv verileri için bir örnek senaryo:

Girdiler:

Kitap önerisi yapılacak kullanıcı adı: **NU1**

Benzer kullanıcı sayısı(k): **3**

Çıktılar:

NU1 kullanıcıasına en yakın kullanıcılar(k=3) ve hesaplanan pearson benzerlikleri sırasıyla, <kullanıcı adı, pearsons coefficient>:

1. 'U5', 0.966
2. 'U9', 0.909
3. 'U19', 0.857

NU1 kullanıcıasında okunmamış olan kitaplar için hesaplanan tahmini beğenme değerleri:

1. 'THE DA VINCI CODE', 3.4021
2. 'RUNNY BABBIT', 1.7139

Sonuç olarak önerilen kitap: THE DA VINCI CODE

!! Güncelleme: Bir önceki çözümde r_a r_b hesaplanırken okunmayan kitaplar da dikkate alınmıştı. Bu değerlerde sadece okunan kitapların ortalaması dikkate alındığında aşağıdaki sonuçlar elde edilmektedir.

Girdiler:

Kitap önerisi yapılacak kullanıcı adı: **NU1**

Benzer kullanıcı sayısı(k): **3**

Çıktılar:

NU1 kullanıcıasına en yakın kullanıcılar(k=3) ve hesaplanan pearson benzerlikleri sırasıyla, <kullanıcı adı, pearsons coefficient>:

1. 'U16', 0.945
2. 'U5', 0.866
3. 'U9', 0.849

NU1 kullanıcıasında okunmamış olan kitaplar için hesaplanan tahmini beğenme değerleri:

1. 'THE DA VINCI CODE', 3.5379
2. 'RUNNY BABBIT', 2.9163

Sonuç olarak önerilen kitap: THE DA VINCI CODE

Testler : Tasarladığınız tavsiye sisteminin başarısını ölçmek için recommend.xls* dosyasındaki örnekleri kullanınız. Her satır bir okuyucunun 8 kitap için verdiği puanları göstermektedir. U1-U20 okuyucularının verdikleri puanları kullanarak NU1-NU5 okuyucularına kitap öneriniz. NU1-NU5 okuyucularının her biri için ayrı ayrı aşağıdaki işlemleri yapınız :

1. U1-U20 arasındaki okuyucular arasından en çok benzediği 3 ($k=3$) okuyucuyu tespit ediniz.
2. En çok benzeyen 3 okuyucunun NU okuyucusunun okumadığı kitaplara verdiği puanlara bakarak puan tahmini yapınız.
3. Bütün NU kullanıcıları için en çok benzediği 3'er kullanıcıyı ve en çok önerilen kitabı tablo halinde veriniz.

*Boş hücreler okunmayan kitapları gösterir. "0" puan olarak değerlendirilmelidir.

Ödev Teslimi ile ilgili önemli bilgiler:

Aşağıda verilen bütün bilgileri içeren tek bir doküman hazırlayarak **20.01.2021 saat 23:59'a** kadar online.yildiz.edu.tr adresi üzerinden **OgrenciNumarasi.rar** dosyasını yükleyiniz.

Teslim Edilecekler:

1. Algoritmanızın **C** dilinde programını hazırlayarak dokümana da ekleyiniz.
2. Algoritmanız için farklı durumları gösteren ekran çıktılarını dokümanda paylaşınız.
3. Teslim Edilecekler
 - a. HW#_OgrenciNumarasi.rar (Örn: 15011001.rar)
 - i. OgrenciNumarasi_QuestionNo.pdf (Örn: 15011001_1.pdf)
 - ii. OgrenciNumarasi_QuestionNo_Part.c (Örn: 15011001_1.c)

Değerlendirme

Algoritma Tasarımı ve Programın Çalışması: (%80)

1. Ödev, istenilen işlerin tamamını yerine getirmelidir.
2. Gereksiz kontrollerden ve işlemlerden arınmış bir tasarım yapılmalıdır.
3. Program hatasız çalışmalıdır.
4. Programın çalışması sırasında, konuyu bilmeyen kişilerin rahatlıkla anlayabilmesi için, giriş ve çıkışlarda mesajlarla bilgi verilmelidir.

Rapor Dokümantasyonu: (%20)

1. Raporun kapak sayfasında, dersin adı, öğrencinin ad, soyad ve numarası, ödev konusu bilgileri yer almalıdır.
2. Kaynak kodda değişken deklarasyonu yapılırken her değişken tek satırda tanımlanmalı, tanımın yanına değişkenin ne için kullanılacağı açıklama olarak yazılmalıdır.
3. Değişken isimleri anlamlı olmalıdır.
4. Her fonksiyonun yaptığı iş, parametreleri ve dönüş değeri açıklanmalıdır.

5. Gerekli yerlerde açıklama satırları ile kodda yapılan işlemler açıklanmalıdır.
6. Gereksiz kod tekrarı olmamalıdır.
7. Kaynak kodun formatı düzgün, okunabilir ve takip edilebilir olmalıdır.

ÖNEMLİ NOT: Ödevinizi lütfen zamanında sisteme yükleyiniz. Sistem aksaklıklarını da gözönünde bulundurarak yükleme işlemini son dakikalara bırakmayınız. E-mail ile gönderilen **ÖDEVLER** kesinlikle **DEĞERLENDİRİLMEMEYECEKTİR**: