
GAZİ ÜNİVERSİTESİ
Gazi
1926

İŞBİRLİKÇİ FİLTRELEME TABANLI TAVSİYE SİSTEMLERİNDE SEYREKLİK PROBLEMİ İÇİN YENİ BİR MODEL

Mohammed Jameel MOHAMMED

YÜKSEK LİSANS TEZİ
BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

GAZİ ÜNİVERSİTESİ
BİLİŞİM ENSTİTÜSÜ

MART 2024

MOHAMMED JAMEEL MOHAMMED tarafından hazırlanan "İŞBİRLİKÇİ FİLTRELEME TABANLI TAVSİYE SİSTEMLERİNDE SEYREKLİK PROBLEMİ İÇİN YENİ BİR MODEL" adlı tez çalışması aşağıdaki jüri tarafından OY BİRLİĞİ / ~~OY ÇOKLUĞU~~ ile Gazi Üniversitesi Bilgisayar Bilimleri Ana Bilim DalındaYÜKSEK LİSANS TEZİ olarak kabul edilmiştir.

Danışman: Prof. Dr. Murat YÜCEL

Elektrik- Elektronik Mühendisliği AnaBilim Dalı, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum. ……..……..

İkinci Danışman: Dr. Öğr. Üyesi. Fahrettin HORASAN

Bilgisayar Mühendisliği Ana Bilim Dalı, Kırıkkale Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum. ...……

Başkan: Prof. Dr. Fatih Vehbi ÇELEBİ

Bilgisayar Mühendisliği, Ankara Yıldırım Beyazıt Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum. ...……

Üye: Prof. Dr. Yusuf SÖNMEZ

Bilgisayar Mühendisliği, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum. ...……

Üye: Doç. Dr. Murat DÖRTERLER

Bilgisayar Mühendisliği, Gazi Üniversitesi

Bu tezin, kapsam ve kalite olarak Yüksek Lisans Tezi olduğunu onaylıyorum/onaylamıyorum. ...……

Tez Savunma Tarihi: 27/03/2024

Jüri tarafından kabul edilen bu tezin Doktora Tezi olması için gerekli şartları yerine getirdiğini onaylıyorum.

……..……

Prof. Dr. Aslıhan TÜFEKCİ

Bilişim Enstitüsü Müdürü

ETİK BEYAN

Gazi Üniversitesi Bilişim Enstitüsü Tez Yazım Kurallarına uygun olarak hazırladığım bu tez çalışmasında;

- Tez içinde sunduğum verileri, bilgileri ve dokümanları akademik ve etik kurallar çerçevesinde elde ettiğimi,
- Tüm bilgi, belge, değerlendirme ve sonuçları bilimsel etik ve ahlak kurallarına uygun olarak sunduğumu,
- Tez çalışmasında yararlandığım eserlerin tümüne uygun atıfta bulunarak kaynak gösterdiğimi,
- Kullanılan verilerde herhangi bir değişiklik yapmadığımı,
- Bu tezde sunduğum çalışmanın özgün olduğunu,

bildirir, aksi bir durumda aleyhime doğabilecek tüm hak kayıplarını kabullendiğimi beyan ederim.

Mohammed Jameel MOHAMMED

27/03/2024

İŞBİRLİKÇİ FİLTRELEME TABANLI TAVSİYE SİSTEMLERİNDE SEYREKLİK PROBLEMİ İÇİN YENİ BİR MODEL

(Yüksek Lisans Tezi)

Mohammed Jameel MOHAMMED

GAZİ ÜNİVERSİTESİ

BİLİŞİM ENSTİTÜSÜ

Mart 2024

ÖZET

Web'de bulunan bilgilerin günden güne artması ve kullanımının yaygınlaşması, kullanıcıları hangi içerikleri tercih edecekleri konusunda karar veremez hale getirmektedir. Bunun yanı sıra artan içeriklerin tamamının analizi ve işlenmesi için zorluklar oluşmaktadır. Öneri sistemleri bu bağlamda, kullanıcılara tercihlerine ve ilgi alanlarına göre uygun içerik önerileri sunmaktadır. Bununla birlikte, birçok alanda ve uygulama alanındaki başarılarına rağmen, yine de çözülmeyi bekleyen bazı sorunlar ve sınırlamalar mevcuttur. Seyreklik problemi, özellikle işbirlikçi filtreleme tabanlı tavsiye sistemlerinde sıkça karşılaşılan bir durumdur. Bu, sınırlı veriye dayanarak kullanıcı tercihlerini ve benzerlikleri belirlemeyi zorlaştırır. Bu çalışmada işbirlikçi önerisistemlerinde karşılaşılan seyreklik problemine odaklanılmıştır. Ölçeklenebilirlik sorunu ise, büyük ölçekli platformlarda milyonlarca kullanıcı ve ürün bulunduğunda ortaya çıkar. Bu durum, büyük veri kümelerini etkili bir şekilde işlemeyi ve analiz etmeyi zorlaştırır. Her iki zorluk da tavsiye sistemlerinin etkinliğini ve performansını olumsuz yönde etkileyebilir. Bu nedenle, bu sorunlara yönelik etkili çözümler geliştirmek, daha iyi tavsiyeler sunmak için kritik öneme sahiptir. Ayrıca bu problemle birlikte ölçeklenebilirlik probleminin de üstesinden gelinmiştir. Literatürde kullanılan mevcut veri setleri ile yapılan testlereden alınan sonuçlara göre önerilen yöntemin oldukça verimli ve maliyetinin düşük olduğu görülmektedir.

BilimKodu	:	92431
Anahtar Kelimeler	:	Öneri sistemleri, seyreklik problemi, işbirlikçi filtreleme, SVD
SayfaAdedi	:	49
Danışman	:	Prof. Dr. Murat YÜCEL
İkinci Danışman	:	Dr. Öğr. Üyesi Fahrettin HORASAN

A NEW MODEL FOR THE SPARSITY PROBLEM IN COLLABORATIVE FILTERING BASED RECOMMENDATION SYSTEMS

(M. Sc. Thesis)

Mohammed Jameel MOHAMMED

GAZİ UNIVERSITY

INSTITUTE OF INFORMATICS

March 2024

ABSTRACT

The increasing availability and widespread use of information on the Web makes users unable to decide which content they prefer. In addition, there are difficulties in analyzing and processing all of the increasing content. In this context, recommender systems offer users appropriate content recommendations based on their preferences and interests. However, despite their success in many fields and applications, some issues and limitations still need to be addressed. In this paper, the focus is on the sparsity problem in collaborative recommender systems. The sparsity problem is a common occurrence, particularly in collaborative filtering-based recommendation systems, making it challenging to determine user preferences and similarities based on limited data. This study focuses on addressing the sparsity problem encountered in collaborative recommendation systems. On the other hand, the scalability issue arises when there are millions of users and items on large-scale platforms, making it difficult to process and analyze large datasets effectively. Both challenges can negatively impact the effectiveness and performance of recommendation systems. Therefore, developing effective solutions to these problems is critically important for providing better recommendations. Furthermore, the scalability problem is tackled along with this problem. According to the results obtained from tests with existing dataset used in the literature, the proposed method is highly efficient and cost-effective.

ScienceCode	:	92431
KeyWords	:	Recommender systems, sparsity problem, collaborative filtering, SVD
PageNumber	:	49
Supervisor	:	Prof. Dr. Murat YUCEL
Co-Supervisor	:	Dr. Instructor Member. Fahrettin HORASAN

TEŞEKKÜR

Yüksek lisans ders ve tez dönemlerimde tecrübesini, bilgisini ve ilgisini esirgemeyen araştırmalarımda, tezimde ve planlanmasında beni yönlendiren, teşvik eden değerli danışmanım sayın Prof. Dr. Murat YÜCEL'e (Gazi Üniversitesi) teşekkürlerimi sunarım, yine tezimin hazırlanması sürecinde desteklerini her zaman hissettiğim ikinci danışmanım Dr.Öğr. Üyesi Fahrettin HORASAN'a (Kırıkkale Üniversitesi) teşekkür ederim.

Ayrıca, eğitim hayatım boyunca her zaman yanımda olan ve desteklerini esirgemeyen aileme ve arkadaşlarım (Akerke, Luiza, Aisylu, Wisam, Kemal) teşekkür ederim.

İÇİNDEKİLER

Sayfa

ÖZET iv

ABSTRACT v

TEŞEKKÜR vi

İÇİNDEKİLER vii

ŞEKİLLERİN LİSTESİ ix

TABLOLARIN LİSTESİ x

SİMGELER VE KISALTMALAR xi

1. GİRİŞ 1

2. ÖNERİ SİSTEMLERİ VE FAYDALARI 5

2.1.1. Tarihsel Bağlamda Öneri Sistemleri 5

2.1. Yaygın Kullanılan Uygulamalar 6

2.1.1. Netflix 6

2.1.2. Amazon 7

2.1.3. YouTube 7

2.1.4. Spotify 8

2.1.5. LinkedIn 8

2.2. Faydaları 8

2.2.1. Gelir Artışı 8

2.2.2. Az Bilinen Ürünleri Tanıtma 9

2.2.3. Kullanıcı Memnuniyeti 9

2.2.4. Müşteri Bağlılığı 9

2.2.5. Detaylı Raporlama 9

Sayfa

3. İLİŞKİLİ ÇALIŞMALAR 11
3.1. Hibrit Filtreleme 11
3.2. Yeni Benzerlik Ölçütlerinin Geliştirilmesi 12
.3.3 Kendi Kendini Düzenleyen Haritaları Kullanan ÖS 13
.3.4 Film Öneri Sistemleri 13
4. METOTLAR 15
4.1. İşbirlikçi Filtreleme 16
4.2. Benzerlik Metriği 17
4.3. Tahmin metriği 18
4.4. SVD Kullanarak Ortak Filtreleme 19
4.5. İçerik Tabanlı Filtreleme 19
4.5. Hibrit filtreleme 21
5. PRATİK UYGULAMA 23
5.1. Ön İşleme Adımı 23
5.2. Önerilen Yöntemin Etkinliği 24
5.3. Deneysel Çalışma 24
.5.3.1 Tekil değer ayrıştırması SVD 24
5.3.2. Büyük matris 25
5.3.3. Veri seti 26
5.3.4. Sistemin uygulanması 27
6. DENEYSEL SONUÇLAR 31
7. SONUÇ VE ÖNERİLER 39
KAYNAKLAR 41
ÖZGEÇMİŞ 49

ŞEKİLLERİN LİSTESİ

Şekil **Sayfa**

Şekil 2.1.Netflix, öneri sistemi 6

Şekil 2.2.Amazon, öneri sistemi 7

Şekil 4.1.İF Kullanıcı öğelendirmeleri ve aralarındaki korelasyon örneği 18

Şekil 4.2. SVD matrisinin U,S ve V^T'ye ayrıştırılması. 20

Şekil 4.3.Hem İF hem de İTF kullanan çeşitli hibrit filtreleme modelleri 22

Şekil 5.1.Önerilen modelin akış şeması 23

Şekil 5.2.Matris Σ 25

Şekil 6.1.α ve λ parametrelerine göre RMSE ve MAE sonuçları (Movielens 100k)..... 32

Şekil 6.2.α ve λ parametrelerine göre RMSE ve MAE sonuçları (Movielens 1m) 32

Şekil 6.3.α ve λ parametrelerine göre RMSE ve MAE sonuçları(Movielens 100k)..... 33

Şekil6.4. k parametresine göre MAE sonuçları(Movielens 100k)..... 33

Şekil 6.5. k parametresine göre RMSE sonuçları(Movielens 1M) 34

Şekil 6.6. k parametresine göre MAE sonuçları(Movielens 1M) 35

TABLOLARIN LİSTESİ

Tablo **Sayfa**

Tablo4.1.İF kullanıcı tabanlı yaklaşımı kullanan öğelendirme tahmini örneği 17

Tablo6.1.Farklı α=0.01 ve λ=0.1 durumlarında farklı k değerlerine göre elde edilen sonuçlar (Movielens 100k) 35

Tablo6.2.Farklı α=0.01 ve λ=0.1 durumlarında farklı k değerlerine göre elde edilen sonuçlar(Movielens 1M) 36

Tablo6.3.Performans Karşılaştırması 36

SİMGELER VE KISALTMALAR

Bu çalışmada kullanılmış simgeler ve kısaltmalar, açıklamaları ile birlikte aşağıda sunulmuştur.

Simgeler	**Açıklamalar**
α	kullanıcıların ve grupların özelliklerini ağırlıklandırması
$\boldsymbol{\mu_{u,i}}$	u ve i için bir düzeltme terimi
$\boldsymbol{p_{u,i}}$	belirlenen tahmin değere
$\boldsymbol{\omega_{i,j}}$	i ve j arasındaki ağırlık
$\boldsymbol{G_{u,i}}$	u ve i için benzerlik ağı
i	item maddesi
j	matristeki item satırları
Jaccard($\boldsymbol{r_u}$,$\boldsymbol{r_j}$)	benzerlik ölçümü, iki kümenin elemanlarının sayısını ifade eder.
MSD($\boldsymbol{r_u}$,$\boldsymbol{r_j}$)	Ortalama Kare Farkı (Mean Squared Difference) ölçümü
$\boldsymbol{newsimilarity}$	yeni benzerlik ölçümü , veri noktası arasındaki benzerlik
$\boldsymbol{r_u}$, $\boldsymbol{r_j}$	iki tarafından paylaşılan öğeleri içeren kullanıcı öğelendirme kümesi
$\boldsymbol{sim_{user}(u,j)}$	kullanıcılar arasındaki geleneksel benzerlikler
$\boldsymbol{sim_{group}(u,j)}$	kullanıcı gruplarını temel alan benzerlikler
$\boldsymbol{sim(u,j)}$	İki kullanıcı arasındaki benzerlik
TFi,j	I ve j'deki frekansı
u	Aktif kullanıcı
$\#\boldsymbol{r_{u,j}}$	iki kullanıcı tarafından paylaşılan öğelendirmelerin önemi

Kısaltmalar	**Açıklamalar**
IDF	Ters Belge Frekansı
İF	İşbirlikçi Filtreleme
İTF	İçerik Temelli Filtreleme (Content-Based Filtering)
MAE	Ortalama Mutlak Hata
ÖS	Öneri sisteme

Kısaltmalar	Açıklamalar
PVR	Personal Video Ranker
SOM	Kendi Kendini Düzenleyen Haritalar (Self-Organizing Maps)
RMSE	Ters Belge Frekansı
SVD	Tekil Değer Ayrışımı (Singular Value Decomposition)
TF	Terim Frekansı

1. GİRİŞ

Öneri sistemleri, kullanıcılara belirli içerikleri veya ürünleri tavsiye etmek için kullanılan bilgi işleme sistemleridir. Bu sistemler, kullanıcıların geçmiş tercihlerini, beğenilerini veya davranışlarını analiz ederek, kişiselleştirilmiş öneriler sunarlar. Öneri sistemleri genellikle çevrimiçi platformlarda, özellikle e-ticaret siteleri, medya platformları ve sosyal ağlar gibi alanlarda kullanılır. Temel amacı, kullanıcı deneyimini artırmak, satışları teşvik etmek ve platformlara olan kullanıcı bağlılığını güçlendirmektir. Bu sistemler genellikle veri madenciliği, yapay zeka ve makine öğrenimi tekniklerini kullanarak çalışır ve büyük miktarda veri analizi yaparak etkili öneriler sunarlar.

Günümüzde, öneri sistemleri alanında pek çok araştırma yapılmaktadır. Öneri sistemleri (ÖS), önerinin nasıl oluşturulduğuna göre dört türe ayrılabilir: İşbirlikçi filtreleme (İF), demografik filtreleme, içerik tabanlı filtreleme ve hibrit sistemler (Bobadilla, Ortega, Hernando, & Gutiérrez, 2013:109,132). İçerik tabanlı ÖS, öneriyi tahmin etmek ve oluşturmak için belirli öğelerin özelliklerini analiz ederek bir kullanıcı profili oluşturur (Alhijawi, Obeid, Awajan, Tedmori, 2018, Chen, Chen, Wang, 2015). Bu nedenle, önerilen öğeler genellikle kullanıcının geçmişte beğendiği öğelere benzer.

Demografik tabanlı yöntem, ortak demografik profillere sahip kullanıcıların (örneğin yaş, cinsiyet ve ülke gibi) aynı ilgi alanlarına sahip olacağını varsayar. Bu nedenle, bu öneri yöntemi farklı demografik durumlar için farklı öğeler tahmin eder. İF yöntemi, en popüler öneri yöntemlerinden biri haline gelmiş ve doğruluk açısından önemli başarılar elde etmiştir (J. Bobadilla, F. 2010:520,528). İF yöntemlerinin prensibi, öneriyi tahmin etmek için tarihsel kullanıcı öğelendirme bilgilerini analiz etmeye dayanır. Başka bir deyişle, (T. Hofmann,Jan. 2004) İF, öğelerin özelliklerinden bağımsız olarak, benzer kullanıcıların tercih ettiği öğeleri bir kullanıcıya önerir (X. Guan, C.-T. Li and Y. Guan, 2016). İF, öneriyi oluşturmak için iki farklı mekanizma uygular: bellek tabanlı İF ve model tabanlı İF (Alhijawi & Kilani, 2020). Bellek tabanlı İF, K-benzer kullanıcıları aktif kullanıcıya (AK) yerleştirmek için tarihsel öğelendirme verilerini kullanır (Alhijawi, 2019a, Chen, Chen, Wang, 2015). K-benzer kullanıcılar hakkındaki tarihsel öğelendirme verileri, öneriyi tahmin etmek için kullanılır. Model tabanlı İF, tarihsel öğelendirme verilerini eğitmek için Bayesian ağları, bulanık algoritmalar, kümeleme modelleri ve Genetik Algoritma (GA) gibi bir model kullanır (Alhijawi, 2019b, Bobadilla, Ortega, Hernando, Gutiérrez, 2013:27,34). Herhangi iki veya

daha fazla öneri tekniğini birleştirmek, bir hibrit ÖS'ye yol açar (Alhijawi, Obeid, Awajan, Tedmori, 2018).

Öneri sistemlerinin internet sitelerinde yaygın şekilde kullanılması ve bu kullanımın giderek artması, bu sistemlerin ne denli önemli olduğunu göstermektedir. İşte bu uygulamalara netflix, amazon, youtube, spotify, linkedIn gibi platformlar örnek olarak verilebilir (Netflix, 2009). Netflix, önerileri sunmak için tek bir algoritma kullanmak yerine, farklı kullanım durumlarına uygun olarak çeşitli algoritmaları bir araya getirmektedir (Smith, B., & Linden, G. 2017:12,18). 1998'de Amazon.com, öneri sistemini oluşturmak için öğe tabanlı bir işbirliği filtreleme algoritmasını uyguladı. Algoritmada önceki kullanıcı davranışlarına, bağlamlara ve öğelendirmelere dayanarak farklı bir müşteri deneyimi sunmayı amaçlamıştır(J. Bobadilla, F. Serradilla, and A. Hernando,2009:261,265). Bu sistem, tüm satın alma süreci boyunca entegre bir şekilde kullanılır. Kullanıcının daha önce gördüğü öğeler temel alınarak hızla öneriler sunar. Görüntülenen ürünle birlikte satın alınan ürünler için önerilerde bulunur. Ayrıca, daha önce keşfedilmemiş ilgi alanlarını açığa çıkarmak için ürünlerle ilişkilendirilebilecek diğer ürünler için öneriler sunar (J. Bobadilla, F. Ortega, A. Hernando, and A. Guti´errez, 2013:27,34). Sistem ayrıca, geçmiş müşteri satın almalarından elde edilen önerileri içeren bir bölüm de içerir. Dünyanın en çok ziyaret alan web sayfalarından biri olan YouTube'un öneri sistemi, en iyi önerileri sunanlardan biri olarak bilinir ve mevcut kullanıcı etkinliklerine dayanarak kişiselleştirilmiş öneriler sunmaya odaklanır. Ayrıca, sistemin ikinci bir amacı da sunduğu kapsamlı içerik seçimini öneriler aracılığıyla kullanıcılara tanıtmaktır. Müzik dinleme platformu olan Spotify, kullanıcıların dinleme alışkanlıklarınıanaliz eder ve benzer müzik türlerinde yeni şarkılar ve sanatçılar önerir. Kullanıcıların müzik keşfini kolaylaştırır. İş odaklı sosyal medya platformu olan LinkedIn, kullanıcıların iş deneyimleri, bağlantılar ve ilgi alanları gibi verilere dayalı olarak iş bağlantıları ve iş ilanları önerir. Kullanıcıların kariyerlerini geliştirmelerine yardımcı olur (J. Bobadilla, F. Serradilla, and J. Bernal, 2010:520,528). Bahsedilen bu popüler uygulamalar, kullanıcıların platformlar üzerinde daha fazla zaman geçirmesini sağlamak ve kullanıcı deneyimini artırmak amacıyla öneri sistemlerini etkili bir şekilde kullanır (Szymanski BK,2013:138,43). Bu da gösteriyor ki, öneri sistemleri günümüzde dijital platformların vazgeçilmez bir parçası haline gelmiştir.

Öneri sistemleri, müşterilere sadece ilgi çekici ürünler önermekle kalmaz, aynı zamanda e-ticaret gelirlerini de artırabilir. Kullanıcının talep ve tercihlerini analiz ederek, kişiselleştirilmiş ürün önerileri yapabilir ve çapraz satışı teşvik edebilir. Kullanıcının

ihtiyaçlarına uygun tamamlayıcı ürünler için öneriler sunmak, dönüşüm oranlarını artırmanın önemli bir yoludur. Öneri sistemleri, kullanıcının ilgi alanlarına göre çeşitli ürünleri tanıtabilir. Bu yöntem, içerik üreticilerin ve dijital ortam yöneticilerinin özellikle az bilinen ürünleri için hedef kitleye reklam yapmasına olanak tanımaktadır. Öneri sistemleri, kullanıcılara aradıkları ürünleri bulmalarında yardımcı olduğu için büyük bir kullanıcı memnuniyeti sağlar (Y. Karabulut, J. Mitchell, P. Herrmann, and C. Jensen, 2008:135,150). Kullanıcılar, sistem tarafından sunulan kişiselleştirilmiş deneyimlere bağlılık geliştirdikçe, müşteri bağlılığı artar. Mutlu müşteriler, sık sık ziyaret etme eğiliminde olduklarından içeriklerin hedef kitleye ulaşmasını ve gelir kazanımını sağlar. Öneri sistemleri ayrıca detaylı raporlar oluşturarak ve kullanıcı davranışını analiz ederek işletmelere değerli bilgiler sağlar. Farklı reklam kampanyalarının veya diğer stratejilerin etkilerini gözlemlemek, hizmet sağlayıcılara stratejilerini optimize etme imkânı sunar(M. Lee, P. Choi, and Y. Woo, 2002:531,534). Kullanıcıların ilgi alanları hakkında bilgi sahibi olmak, işletmelere kullanıcıları daha etkili bir şekilde çekmek için kullanılabilecek değerli bir yetenektir.

Bu çalışmada, rating matrisinin seyrekliğini azaltarak daha etkili tahminler yapma amaçlanmıştır. Aynı zamanda, özellik sayısını azaltarak ölçeklenebilirlik problemine çözüm getirme hedeflenmiştir. Bu doğrultuda, rating matrisindeki boş alanlar, rasgele seçilen kullanıcı ve öğe çiftleri temel alınarak doldurulmuştur. Seçilen kullanıcı-öğe sayısı, matrisin toplam rating sayısına eşit olacak şekilde belirlenmiştir. Bu yöntemle rating matrisinin yoğunluğu iki katına çıkarılmış ve diğer bir deyişle matrisin seyrekliği azaltılmıştır. Oluşturulan modelde, rasgele seçilen kullanıcı-içerik çiftlerine yönelik tahminler, Tekil Değer Ayrışımı (Singular Value Decomposition – SVD) yöntemi kullanılarak gerçekleştirilmiştir. Matris yoğunlaştırma işlemi tamamlandıktan sonra elde edilen yeni matris, ikinci bir SVD tekniğiyle eğitilmiştir. Daha sonra bu yeni veriyle oluşturulan model test edilmiştir.

Bu tez dört bölüme ayrılmıştır. Birinci bölümde, öneri sistemleri ve faydaları hakkında bilgiler sunulmuştur. Bu kısımda, genel olarak öneri sisteminden, tarihinden ve bu konuyla ilgili yapılan araştırmalardan bahsedilmektedir. Ayrıca, öneri sistemini kullanan önemli web siteleri ve bu sitelerde kullanılan algoritmalar hakkında bilgiler yer almaktadır. Bunun yanı sıra, öneri sisteminin gelirler, ürün tanıtımı ve kullanıcı memnuniyeti açısından sağladığı faydalara da değinilmektedir. İkinci bölümde, ilişkisel çalışmalar ele alınmıştır. Bu bölümde, araştırmamızla ilgili olan çalışmalar ve modeller, bu modellerin eksiklikleri ve bu sorunların nasıl çözülebileceği aktarılmıştır. Üçüncü bölüm olan "Metotlar" kısmında, tez çalışmasında

oluşturulan model için kullanılan algoritmalar açıklanmaktadır. Temel algoritma olarak İşbirlikçi Filtreleme'de kullanılan SVD kullanılmıştır. Dördüncü bölümde, "Deneysel Sonuçlar" kısmında, araştırmanın deney ortamı ve elde edilen sonuçlar açıklanmıştır.

2. ÖNERİ SİSTEMLERİ VE FAYDALARI

Öneri sistemleri hakkında birçok araştırma yapılarak yeni stratejiler belirlemek ve bu tür sistemleri geliştirmek amaçlanmıştır (J. Bobadilla ve diğerleri, 2013), (N. Manouselis ve diğerleri, 2012). Bu bölümde, 2002'den günümüze kadar yapılan çalışmaların özetleri sunulmuştur, (G. Adomavicius ve A. Tuzhilin, 2005: 734,749).

2.1.1. Tarihsel Bağlamda Öneri Sistemleri

Öneri sistemleri (ÖS), bilgisayar bilimleri ve yapay zeka alanında önemli bir gelişme kaydeden, kullanıcılara özelleştirilmiş içerik ve ürün önerileri sunan yazılımlardır. Bu sistemler, kullanıcıların geçmiş tercihlerini ve davranışlarını analiz ederek, gelecekte ilgilerini çekebilecek öğeleri tahmin etmeye yönelik çalışır (P. C. Vaz, D. M. de Matos, B. Martins ve P. Calado, 2012:1203,5324). Öneri sistemlerinin tarihçesi oldukça köklüdür. 1992 yılından bu yana, ÖS'lerin geliştirilmesi ve optimize edilmesi için birçok önemli çalışma yapılmıştır.

Erken dönem ÖS araştırmaları, farklı filtreleme tekniklerinin nasıl birleştirileceği ve hibrit filtreleme ile ÖS oluşturmanın yöntemleri üzerinde odaklanmıştır (Goldberg ve diğerleri, 1992). Bu dönemde, bilgiye dayalı filtreleme tekniklerinin önemi vurgulanmıştır (Burke, 2002). Ancak, 2004 ve 2005 yıllarında, genişleyen internet hizmetleri ile birlikte ÖS araştırmaları yeni bir boyut kazanmıştır. Bu dönemde, ÖS'leri çalıştırmak için gereken bellek miktarı ve önerilerin hesaplanma süresi gibi hesaplama maliyeti sorunları gündeme gelmiştir (Adomavicius ve Tuzhilin, 2005).

Öneri sistemlerinin popülerleşmesiyle birlikte, kullanıcıların farklı ihtiyaçlarını karşılayabilmesi için özelleştirmenin önemi ortaya çıkmıştır. Örneğin, bir müzik platformunda kullanıcılara yapılan öneriler ile bir e-ticaret platformunda yapılan öneriler farklı olabilir. Bu dönemde, öneri sistemlerinin doğruluğu ve şeffaflığı da büyük bir önem kazanmıştır (Adomavicius ve Tuzhilin, 2005; Herlocker ve diğerleri, 2004).

Sonuç olarak, öneri sistemleri günümüzde çeşitli platformlarda (alışveriş siteleri, müzik ve video platformları, sosyal medya vb.) yaygın olarak kullanılmaktadır. Bu sistemler, kullanıcı deneyimini artırmak, satışları artırmak ve içeriği daha etkili bir şekilde sunmak için önemli bir rol oynamaktadır. Geçmişten günümüze, öneri sistemlerinin evrimi, teknolojik ilerlemeler ve kullanıcı taleplerine uyum sağlamak adına sürekli olarak şekillenmiştir (R.

Correa ve B. Pinheiro, 2010:31,36), bu da öneri sistemleri alanındaki araştırmaların ve uygulamaların geniş bir yelpazede gerçekleşmesine olanak tanımıştır. Bu alandaki gelecekteki gelişmeler, kullanıcı deneyimini daha da kişiselleştirmeye ve öneri sistemlerinin daha akıllı, hızlı ve doğru hale gelmesine odaklanacaktır.

2.1. Yaygın Kullanılan Uygulamalar

Öneri sistemlerinin internet sitelerinde yaygın şekilde kullanılması ve bu kullanımın giderek artması, bu sistemlerin ne denli önemli olduğunu göstermektedir. Bu uygulamalardan bazılarına ve açıklamalarına dair örnekler aşağıda sunulmuştur.

2.1.1. Netflix

Önde gelen bir öneri sistemi örneğidir. Ancak, Netflix'in önerileri sunmak için tek bir algoritma kullanmak yerine, farklı kullanım durumlarına uygun olarak çeşitli algoritmaları bir araya getirir. Kullanıcıların bir program aramak için harcadığı sürenin, hizmeti bırakma olasılığını artırdığı gerçeğini göz önüne alarak, Netflix, kullanıcıların dikkatini çekecek önerilere odaklanmak için şov listelerini ön plana çıkarmıştır. Ayrıca, kullanıcının her önerinin mantığını anlamasını sağlamak da önemlidir (O. Nalmpantis and C. Tjortjis, 2017).



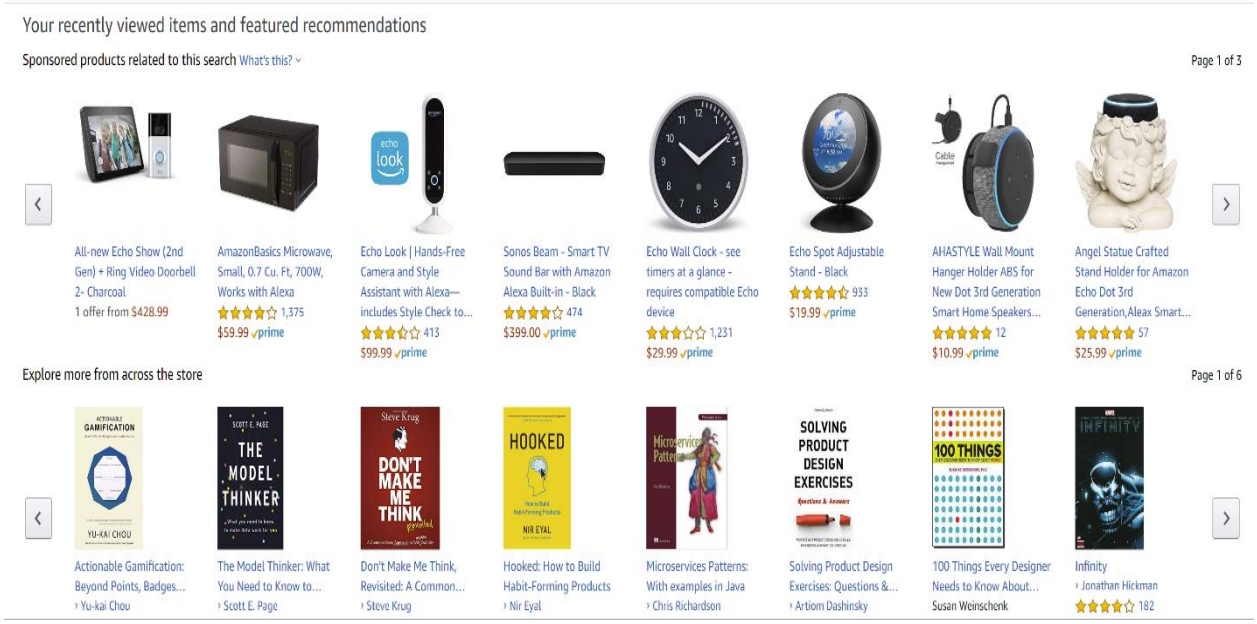
Şekil 2.1 Netflix, öneri sistemi

Sistem, çeşitli algoritmaları içerir, örneğin, kişiselleştirme sağlamak ve belirli bir türdeki filmlerin sıralamasını belirlemek için kullanılan bir yöntem olan "Personal Video Ranker" (PVR) bu algoritmaların bir örneğidir. (S.-M. Choi, S.-K. Ko, and Y.-S. Han, 2012:8079 ,8085) Netflix, hangi tür olursa olsun, kullanıcının tercihlerine uygun en iyi önerileri belirlemek için "Top-N Video Ranker" adı verilen bir algoritmayı kullanır. Ayrıca, "İzlediğiniz İçin" bölümü, videodan videoya benzerliklere dayalı olarak şovlar önerir (J.

Bennett , 2007:35). Netflix, Trend Olan ve İzlemeye Devam Edilen satırlarındaki sıralamaları belirlemek için çeşitli algoritmaları kullanarak, kullanıcılara daha iyi öneriler sunmayı hedefler.

2.1.2. Amazon

1998'de Amazon.com, öneri sistemini oluşturmak için öğe tabanlı bir işbirliği filtreleme algoritmasını uyguladı. Şekil 2'te görülen web sayfasının önizlemesinde, önceki kullanıcı davranışlarına, bağlamlara ve öğelendirmelere dayanarak farklı bir müşteri deneyimi sunmayı amaçlamıştır (G. Linden , 2003). Bu sistem, tüm satın alma süreci boyunca çeşitli şekillerde entegre edilir ve kullanılır. Kullanıcının daha önce gördüğü öğeler temel alınarak hızla öneriler sunar. Görüntülenen ürünle birlikte satın alınan ürünler için önerilerde bulunur(B. Smith and G. Linden, 2017). Ayrıca, daha önce keşfedilmemiş ilgi alanlarını açığa çıkarmak için ürünlerle ilişkilendirilebilecek diğer ürünler için öneriler sunar. Sistem ayrıca, geçmiş müşteri satın almalarından elde edilen önerileri içeren bir bölüm de içerir (G. Linden, B. Smith and J. York, 2003:76,80).



Şekil 2.2 Amazon, öneri sistemi

2.1.3. YouTube

kendi öneri sistemini geliştirmiştir. YouTube'un öneri sistemi, en iyi önerileri sunanlardan biri olarak bilinir ve mevcut kullanıcı etkinliklerine dayanarak kişiselleştirilmiş öneriler sunmaya odaklanır. Ayrıca, sistemin ikinci bir amacı da sunduğu kapsamlı içerik seçimini öneriler aracılığıyla kullanıcılara tanıtmaktır.

Ancak, sistem çeşitli zorluklarla karşılaşmıştır; örneğin, filmlerle ilgili meta verilerin eksik olması ve kullanıcı niyetini belirlemeyi zorlaştıran kısa kullanıcı etkileşimleri gibi (J. Davidson, B. Liebald, J. Liu, P. Nandy, T. V. Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston and D. Sampath, 2010.) Videoların yüklenmesinden viral hale gelmesine kadar olan sürelerin kısa olması, önerilerin sık sık güncellenmesini gerekli kılar. Bu öneri sistemi, çeşitli kaynaklardan bilgi toplar ve videonun tüm içerik bilgilerini kullanır; açık veya örtülü tüm kullanıcı etkileşimlerini dikkate alır.

2.1.4. Spotify

Müzik dinleme platformu Spotify, kullanıcıların dinleme alışkanlıklarını analiz eder ve benzer müzik türlerinde yeni şarkılar ve sanatçılar önerir. Kullanıcıların müzik keşfini kolaylaştırır(Spotify developer).

2.1.5. LinkedIn

İş odaklı sosyal medya platformu LinkedIn, kullanıcıların iş deneyimleri, bağlantılar ve ilgi alanları gibi verilere dayalı olarak iş bağlantıları ve iş ilanları önerir. Kullanıcıların kariyerlerini geliştirmelerine yardımcı olur (S. C. Geyik, Q. Guo, B. Hu, C. Ozcaglar, K. Thakkar, X. Wu, and K. Kenthapadi. 2018).

Bu uygulamalar, kullanıcıların platformlar üzerinde daha fazla zaman geçirmesini sağlamak ve kullanıcı deneyimini artırmak amacıyla öneri sistemlerini etkili bir şekilde kullanır (S. C. Geyik and K. Kenthapadi. 2018). Bu da gösteriyor ki, öneri sistemleri günümüzde dijital platformların vazgeçilmez bir parçası haline gelmiştir.

2.2. Faydaları

Öneri sistemlerinin (ÖS), yalnızca sistem kullanıcılarına değil, aynı zamanda işletmelere de bir dizi avantaj sağladığı açıktır.

2.2.1. Gelir Artışı

Öneri sistemleri, müşterilere sadece ilgi çekici ürünler önermekle kalmaz, aynı zamanda e-ticaret gelirlerini de artırabilir.

Kullanıcının talep ve tercihlerini analiz ederek, kişiselleştirilmiş ürün önerileri yapabilir ve çapraz satışı teşvik edebilir(F. Karimova, 2016.,S. Sterlin, A. Sandhya, S. S. Merlin and B. B. Sam, 2017).

Kullanıcının ihtiyaçlarına uygun tamamlayıcı ürünler için öneriler sunmak, dönüşüm oran larını artırmanın önemli bir yoludur.

2.2.2. Az Bilinen Ürünleri Tanıtma

ÖS, kullanıcının ilgi alanlarına göre çeşitli ürünleri tanıtabilir. Bu yöntem, tedarikçilerin özellikle az bilinen ürünleri için doğru kitleye reklam yapmasına olanak tanır (Shi et al. 2014).

2.2.3. Kullanıcı Memnuniyeti

Öneri sistemleri, kullanıcılara aradıkları ürünleri bulmalarında yardımcı olduğu için büyük bir kullanıcı memnuniyeti sağlar (Zhang 2013). Kullanıcılar, doğru ve ilgili önerilerle keyifli bir alışveriş deneyimi yaşarlar, bu da gelecekte tekrar hizmeti tercih etme olasılıklarını artırır.

2.2.4. Müşteri Bağlılığı

Kullanıcılar, sistem tarafından sunulan kişiselleştirilmiş deneyimlere bağlılık geliştirdikçe, müşteri bağlılığı artar.

Sistemin, kullanıcının tercihlerine ne kadar iyi uyarlandığı, müşteri memnuniyetini artırır ve kullanıcı ile işletme arasında güçlü bir ilişki kurar (Adamopoulos&Tuzhilin 2014). Mutlu müşteriler, sık sık ziyaret etme eğiliminde olduklarından, işletmeye sadık müşteriler ve artan gelir getirir.

2.2.5. Detaylı Raporlama

Öneri sistemleri ayrıca detaylı raporlar oluşturarak ve kullanıcı davranışını analiz ederek işletmelere değerli bilgiler sağlar. Farklı reklam kampanyalarının veya diğer stratejilerin etkilerini gözlemlemek, hizmet sağlayıcılara stratejilerini optimize etme imkanı sunar (Kunaver&Požrl 2017).

Kullanıcıların ilgi alanları hakkında bilgi sahibi olmak, işletmelere kullanıcıları daha etkili bir şekilde çekmek için kullanılabilecek değerli bir yetenektir. Netflix gibi şirketler, kullanıcı zevklerini göz önünde bulundurarak etkili bir içerik önerme stratejisi geliştirmişlerdir (Castells et al. 2015:881,918). Bu, müşterilerin satın alma kararı verirken gerçek değerle ilgili bilgilere erişmelerine yardımcı olur.

3. İLİŞKİLİ ÇALIŞMALAR

Yeni stratejiler tanımlamak ve bu tür sistemleri geliştirmek için ÖS hakkında çok araştırma yapılmıştır (J. Bobadilla ve diğerleri. 2013), (N. Manouselis ve diğerleri. 2012). Bu Bölüm, 2002'den günümüze kadar incelenen ana çalışma alanlarının özetlerini sunacaktır (R. Burke. 2002: 331,370), (J. L. Herlocker ve diğerleri. 2004: 5,53).

3.1. Hibrit Filtreleme

Öneri sistemlerinde oldukça yaygın bir araştırma alanıdır, bu konuda birçok çalışma bulunmaktadır (W. Carrer-Neto ve diğerleri, 2012), (J. Bobadilla ve diğerleri, 2012), (A. A. Kardan ve M. Ebrahimi, 2013: 93,110). Bu çalışmaların çoğu, İşbirlikçi filtreleme (İF) sonuçlarını başka bir filtreleme yöntemiyle birleştirerek, her bir metodun avantajlarını kullanarak öneri sistemlerinin performansını artırmayı hedefler (J. Bobadilla ve diğerleri, 2013), (A. A. Kardan ve M. Ebrahimi, 2013: 93,110). Örneğin, (P. C. Vaz ve diğerleri, 2012) kitap önerileri için bir hibrit yaklaşım geliştirmek için şu matrisleri kullanır: her bir kullanıcının kitaplara verdiği puanlarla oluşturulan user X book ve aynı kitabın farklı kitapların ortalama puanlarıyla oluşturulan user$X$author matrisleri. Daha sonra, iki öğe tabanlı İF yaklaşımı kullanılır: (1) kitap benzerlik matrisi book Xbook ve (2) yazar benzerlik matrisi author X author. Bu iki benzerlik matrisi hesaplandıktan sonra, her birinin ağırlıklı aritmetik ortalaması alınarak birleştirilir ve hibrit bir filtreleme yöntemi elde edilir. Yazarlar, önerdikleri çözümün genel tahminleri iyileştirdiğini belirtmiş olsalar da, ağırlıklı aritmetik ortalama denkleminde kullanılan ağırlıkların daha fazla araştırmayı gerektirdiğini vurgulamışlardır.

Yuan-hong ve diğerleri (2010) ise hibrit filtreleme yöntemini kullanan bir öneri sistemi örneği sunar. Bu yaklaşımda, her kullanıcının beklentilerini içeren bir öğelendirme matrisi oluşturulur, bu matris İçerik Temelli Filtreleme (İTF)(Content-Based Filtering (İTF))ile hesaplanan öğelendirilmemiş filmler için kullanılır. Aynı zamanda, İşbirlikçi filtreleme (İF) algoritması için benzerlik hesaplamalarında yardımcı olacak olan bu matris, İF tahminlerini yapmak için normal bir öğelendirme matrisi olarak da kullanılır. Bu iki farklı filtreleme algoritmasını birleştirmek için (B. M. Kim ve diğerleri, 2006) tarafından önerilen bir yöntem kullanılabilir. Bu yöntemde, her bir benzerlik ölçüsüne bir katsayıα atanır ve bu katsayı, her bir istatistik tarafından belirlenen ortak noktalara dayalı olarak ağırlıklandırılır. Bu, kullanıcılar arasındaki geleneksel benzerlikleri kullanan $sim_{user}(u,j)$ ve kullanıcı

gruplarını temel alan benzerlikleri kullanan $sim_{\text{group}}(u,j)$ olmak üzere iki farklı benzerlik ölçüsünü birleştirmek için kullanılabilir. Bu gruplar, maddelerin açıklamalarına dayalı olarak oluşturulur ve ardından kümeleme algoritmaları uygulanarak gruplar belirlenir. Daha sonra, bu gruplar tarafından üretilen öğelendirme matrisi kullanılarak daha küçük bir öğelendirme matrisi oluşturulur (B. M. Kim ve diğerleri, 2006). Bu yaklaşım, kullanıcıların ve grupların özelliklerini ağırlıklandırmak için α terimini kullanarak $sim_{user}(u,j)$) ve $sim_{\text{group}}(u,j)$ benzerlik ölçüsünü ağırlıklandırır (Denklem 3.1).

$$sim(u,j) = (1-\alpha) \times sim_{user}(u,j) + \alpha \times sim_{group}(u,j) \quad (3.1)$$

3.2.Yeni Benzerlik Ölçütlerinin Geliştirilmesi

Bir ÖS, önerilecek ürünleri arar. Bunu yapmak için nesneleri ve kullanıcıları karşılaştırması gerekir. Yaratılan ÖS'nin (J. Kittler, ve diğerleri. 1998: 226,239)'deki Starcoll yaklaşımıyla soğuk başlatma sorununu çözmesine yardımcı olunur. Bir ÖS, kullanıcı veya nesne hakkında yeterli bilgiye sahip olmadığında, soğuk başlatma sorunu ortaya çıkar. Değerlendirilen şeyleri değiş tokuş etmeyenler de dahil olmak üzere, sağlanan yaklaşım kullanılarak tüm kullanıcılar karşılaştırılabilir. Bu, tüm öğe özniteliklerini kullanıcı ilgi alanları olarak kullanarak ve kullanıcının özelliklerini temsil eden bir kullanıcı modeli geliştirmek amacıyla ortak noktaları bulmak için bunları karşılaştırarak gerçekleştirildi (J. Kittler, ve diğerleri. 1998: 226,239).Başka bir teknik, Pearson korelasyonundan daha iyi performans gösteren bir benzerlik metriği geliştirdi (J. Bobadilla ve diğerleri. 2010). Bu yeni metrik, Ortalama Kare Farkın (Mean Squared Difference - MSD) tamamlayıcısını kullanarak İF tekniklerini geliştirmeyi başardı. (Denklem 3.2) gösterilen Jaccard benzerliği ile birleştirilmesi. 3.3, Denklem ile sonuçlanır. 3.4. r_u ve r_j'nin yalnızca her ikisi tarafından paylaşılan öğeleri içeren kullanıcı öğelendirme kümesini temsil ettiği bellidir. Her iki kullanıcı tarafından paylaşılan öğelendirmelerin önemi $\#r_{u,j}$ ile temsil edilir.

$$MSD(u,j) = \frac{|r_u - r_j|}{\#r_{u,j}} \quad (3.2)$$

$$jaccard(u,j) = \frac{|r_u \cap r_j|}{|r_u \cup r_j|} \quad (3.3)$$

$$newsimilarity(u,j) = (1 - MSD(r_u, r_j)) \times jaccard(r_u, r_j) \quad (3.4)$$

Bu metriğin en önemli faydalarından biri, daha az komşu olsa bile en çok kullanılan benzerlik ölçülerinden biri olan Pearson korelasyonundan daha iyi performans göstermesidir (J. Bobadilla ve diğerleri. 2010).

Yazarlar (J. Bobadilla ve diğerleri. 2013)'te benzerliği yazılım yerine ucuz elektrikli bileşenler kullanarak hesaplayan daha iyi bir benzerlik ölçüsü oluşturmuşlardır. Önerilerden ödün vermeden, teknikleri Pearson korelasyonundan hesaplamak için yaklaşık iki kat daha hızlı olan geliştirmeler sağladı. Zaman hesaplama maliyetinin azalması nedeniyle, teknikleri ÖS'yi daha ölçeklenebilir hale getirdi.

3.3.Kendi Kendini Düzenleyen Haritaları Kullanan ÖS

Kendi Kendini Düzenleyen Haritaları(KKDH) kullanan öneri sistemleri (ÖS), çeşitli açılardan önemli katkılarda bulunmuştur. Örneğin, (M. Lee ve diğerleri, 2002) tarafından yapılan bir çalışma, demografik verileri kullanarak hibrit bir sistem oluşturmuştur. Bu sistemde, demografik bilgiler kullanıcılar arasındaki benzerlikleri daha doğru bir şekilde hesaplamak için kullanılır. KKDH, kullanıcıları sadece demografik verilere dayanarak gruplandırır ve ardından bu gruplar kullanılarak kullanıcının İF komşulukları belirlenir. Bu yaklaşım, hesaplama verimliliğini ve ölçeklenebilirliği artırmıştır (M. Lee ve diğerleri, 2002). Büyük veri setleriyle çalışırken genellikle tercih edilen Temel Bileşen Analizi'nin ardından KKDH'un kullanıldığı bir başka yöntem, Olasılıksal KKDH (M. Mandl ve diğerleri, 2011) algoritmasıdır. Bu yaklaşım, KKDH 'da kullanılacak olan birincil bileşenlerin belirsizliğinitahmin etmek için olasılıksal bir yaklaşımı benimser. Olasılıksal KKDH tek başına kullanıldığında istenmeyen sonuçlar verebilir, ancak diğer modellerle entegre edilerek daha iyi sonuçlar elde edilebilir. Bu yöntem, örneğin Netflix Ödülü gibi büyük veri setlerinde başarılı bir şekilde uygulanmıştır (Netflix, 2009).

S. Gabrielsson ve S. Gabrielsson (2006) tarafından yapılan bir diğer çalışmada, performans komşulukları K-En Yakın Komşu (KNN) veya KKDH kullanılarak karşılaştırılmıştır. Ancak, bu çalışmada, içerik tabanlı filtreleme (İTF) kullanarak sistem iyileştirmeye yönelik yeterli çaba harcanmamıştır. Bu yaklaşımda, öğe açıklamaları için anahtar kelime tabanlı stratejiler kullanılmıştır. Bu ÖS'nin kişisel KKDH'lar yerine tüm kullanıcıların verilerini kullanarak genel bir KKDH oluşturduğu bir ayrım yapılmalıdır.

3.4.Film Öneri Sistemleri

film tabanlı veri setleri üzerinde bir dizi çalışmayı içermektedir (J. Bobadilla ve diğerleri, 2013), (P. Symeonidis ve diğerleri, 2009), (T. George ve S. Merugu, 2005), (J. J. Jung, 2012). Bu sistemler sadece film önerisi yapmakla kalmayıp, bazıları farklı alanlarda da kullanılabilir. Örneğin, (S.-M. Choi ve diğerleri, 2012) film türlerine dayalı olarak gelişmiş bir film öneri sistemi sunmaktadır. Bu sistem, filmlerin türlerine göre çözümler üreterek

tüketicilere ve kullanıcılara yönlendirmelerde bulunur. Araştırmalar, benzer türlerin miktarına ve filmlerin gösterime girdiği on yıla dayalı olarak yapıldı, bu çalışmalar decade-based tür ilişkisinin daha doğru önerilere yol açtığını gösterdi.

Ayrıca, (G. Lekakos ve P. Caravelas, 2008) tarafından geliştirilen More sistemi, film alanında özel olarak oluşturulmuş bir öneri sistemi olarak öğelendirme matrisi ve film açıklamalarını içerir. Bu sistem, İşbirlikçi filtreleme (İF) ve içerik tabanlı filtreleme (İTF) yöntemlerini birleştiren hibrit bir yaklaşım sunmayı amaçlar. Kosinüs benzerlik metriği ve K-En Yakın Komşu (KNN) tekniği gibi tipik İF yöntemlerini kullanırken, İTF oluşturmak için Naive-Bayes yöntemini benimser ve sınıf etiketlerini potansiyel öğelendirmeler olarak kullanır.

(M. Kim ve S. Park, 2013) tarafından önerilen bir diğer yaklaşım ise, çevrimiçi sosyal ağları ve grup farkındalıklı bir sosyal ağ modelini kullanarak etkileşimli bir öneri sistemi sunar. Bu metodoloji, kullanıcı tercihlerindeki hızlı değişikliklere ve sosyal çevrenin etkilerine uyum sağlayabilen dinamik bir öneri sistemi sunmayı amaçlar. Bu çözüm, grup farkındalıklı sosyal güven yönetimi ve öneri modülü olmak üzere iki ana bölüme ayrılır.

4. METOTLAR

Kullanıcıların gelenler kutusunda bulunan e-posta taşmasını gidermek için Goldberg ve diğerlerini 1992 yılında ilk kez ÖS'yi yarattı (D. Goldberg, D. Nichols, B. M. Oki ve D. Terry. 1992). Kullanıcılar makaleleri okuduktan sonra yanıtlarını göndererek filtreleme mekanizmasına katılırlar. ÖS, katılım, bilgi ve dinamik, işbirliğine dayalı iletişim sağlayan bir ağ oluşturma aracıdır (J. Bobadilla ve diğerleri. 2009). ÖS, kullanıcıların büyük miktarda veriyi yönetmesine izin verdiğinden ve onlara bir dizi özel öneri, hizmet ve materyal sunduğundan, talebi son 20 yılda arttmıştır. Sonuç olarak, bilim camiası ve işletmeler tarafından karlarını artırmalarını sağladığı için çeşitli yöntemler oluşturulmuş ve araştırılmıştır (G. Adomavicius ve A. Tuzhilin 2005: 734,749).

Bir ÖS, iki modülden oluşur: bir filtreleme yöntemi ve bir veritabanı (J. Bobadilla ve diğerleri. 2013). Kullanıcılara, mallara ve bunlara eşlik eden öğelendirmelere ilişkin bilgiler veri tabanında tutulur. Filtreleme yöntemi, genellikle iki adıma ayrılan bir algoritmadan oluşur. Ardından, insanlar veya eşyalar arasındaki karşılaştırmalarla birlikte her bir kullanıcı veya nesne için bir yakınlık sağlanır. Algoritmanın, kullanıcının henüz görüntülemediği ürünler için öğelendirmeleri tahmin ettiği ikinci aşamada, yalnızca en iyi öğelendirmeler öneri olarak kullanılır (X. Zhou ve diğerleri. 2012).

Öneri sistemleri kullandığı yöntem ve veriye göre üç grupta incelenmektedir. (J. Bobadilla ve diğerleri. 2013), (G. Adomavicius ve A. Tuzhilin 2005: 734,749), (W. Carrer-Neto ve diğerleri. 2012) (M. Dell'Amico ve L. Capra. 2008: 135,150). Bunlar;

- İşbirlikçi filtreleme: Benzer kullanıcılara ve öğelendirmelerine dayalı olarak öğeler önerir;
- İçeriğe dayalı filtreleme: Kullanıcıya, bir profilin ve bir öğenin özelliklerinin ne kadar yakından eşleştiğine bağlı olarak öneriler verilir;
- Demografik filtreleme: Kullanıcıyla ilgili bilgilere, örneğin kullanıcının cinsiyeti, yaşı, mesleği ve bulunduğu yere göre;
- Sosyal filtreleme: Kullanıcı sosyal ağına başvurarak;
- Hibrit filtreleme: Çeşitli filtreleme yöntemlerinin birleştirilmesinin nihai etkisidir.

Çalışmalarda kullanılanen popüler üç filtreleme biçimi olan işbirlikçi, içerik tabanlı ve hibrit filtreleme biçimleri olup, detayları alt bölümlerde sunulmuştur.

4.1. İşbirlikçi Filtreleme

İşbirlikçi filtreleme (İF) algoritmalarını kullanarak kullanıcı öğelendirmelerini karşılaştırmak, en benzer olan öğelendirmeleri tanımlar (J. Bobadilla ve diğerleri. 2010). Bu tür filtreleme, basit olması ve güvenilir bulgular üretmesi nedeniyle araştırmalarda çok sık kullanılmıştır. Benzerlik ölçüsünün nasıl belirleneceği ve kullanıcı öğelendirmesi almamış bir nesne için öğelendirmenin nasıl tahmin edileceği iki temel araştırma alanıdır.En çok kullanılan iki İF yaklaşımı (J. Bobadilla ve diğerleri. 2013), (G. Adomavicius ve A. Tuzhilin 2005: 734,749),(L. Candillier ve diğerleri. 2007), (P. Symeonidis ve diğerleri. 2009).

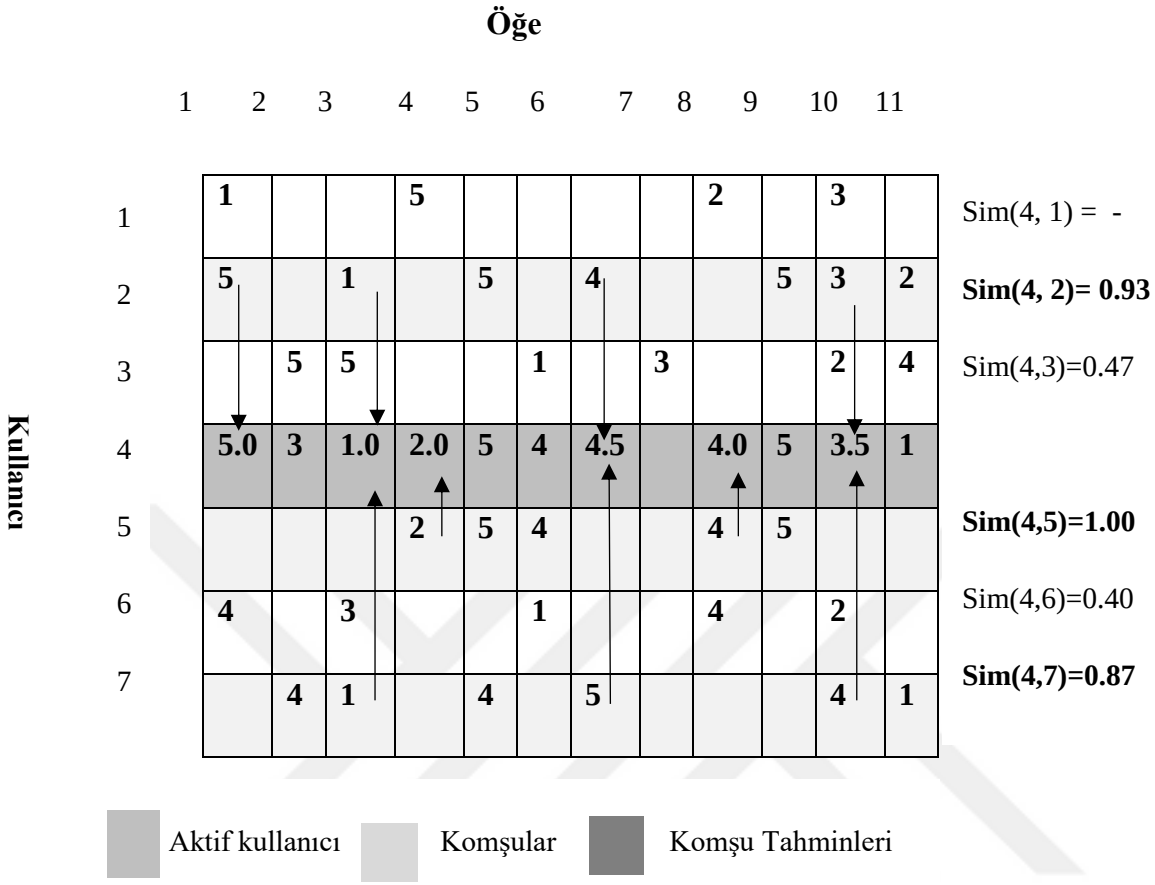
- Kullanıcı tabanlı (user-based): Kullanıcı komşularının (yani en benzer kullanıcıların) belirlenmesi ve ardından komşuların öğelendirmelerine dayalı olarak bir öğelendirmenin tahmin edilmesi;
- Öğeye dayalı (item-based): Komşu öğelerin (yani en benzer öğelerin) belirlenmesi ve ardından komşuların öğelendirmelerine dayalı olarak bir öğelendirmenin tahmin edilmesi;

Kullanıcı tabanlı bir yaklaşımla İF kullanan bir ÖS örneğini açıklayalım. İki kullanıcı arasındaki benzerlik (u ve j), öklid mesafesinin tümleyeni hesaplanarak elde edilir (Denklem 4.1). Öklid mesafesinin tümleyeni hesaplanarak elde edilir. $r_{u,i}$değişeni, u kullanıcısının i maddesine verdiği öğelendirmeye karşılık gelir. Yalnızca hem u hem de j (matristeki satırlar) tarafından paylaşılan öğelendirmeleri içerdiğinden, denklem yalnızca I kümesini kullanır.

$$sim(u,j) = 1 - \sqrt{\sum_{i \in I}(r_{u,i} - r_{j,i})^2} \tag{4.1}$$

Örnek olarak, dördüncü kullanıcı için en iyi üç öneriyi hesaplamanın bir yolunu sunumaktadır. İlk olarak, 4 ile diğer tüm kullanıcılar arasındaki $sim(4, j)$ benzerliği hesaplanacaktır.K-En Yakın Komşular (KNN) algoritması kullanılarak en benzerleri seçilir, bu örnekte bunlar yalnızca üç kullanıcıdır (2, 5 ve 7).Daha sonra, yalnızca bu komşuları kullanarak, ÖS kullanıcının öğelendirmediği tüm i öğelerinin öğelendirmesini tahmin eder. Bu örnekte tahmin denklemi, yalnızca komşu kullanıcıları göz önünde bulundurarak öğe öğelendirmesinin ortalamasıdır.Son olarak, ilk üç öğe seçilir (1, 7 ve 9) ve önerilir (J. Bobadilla ve diğerleri. 2013).

Table 4.1.İF kullanıcı tabanlı yaklaşımı kullanan öğelendirme tahmini örneği



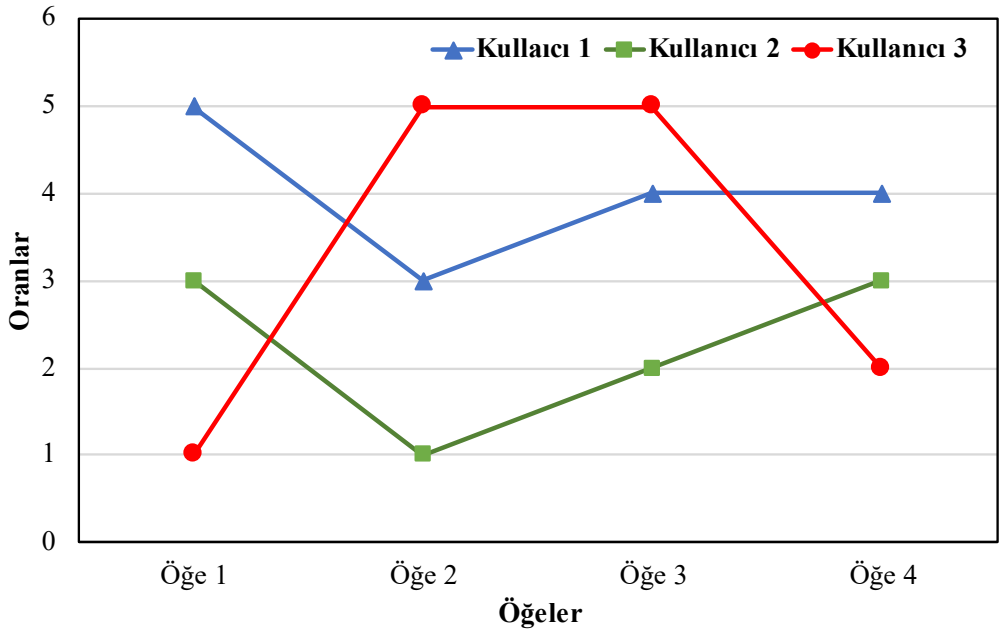
4.2. Benzerlik Metriği

Kullanıcıları veya nesneleri karşılaştırmak ve onlara benzerlik düzeyi atamak için İF algoritmalarında bir benzerlik ölçüsü kullanılır. Pearson korelasyonu İF'de sıklıkla kullanıldığından ve çalışmamızda yer alacağından, bu konuda daha derinlere ineceğiz.Pearson korelasyonu Shardanand ve Maes tarafından 1995 yılında önerilmiştir ve iyi sonuçlar verdiği için ÖS'de yaygın olarak kullanılmaktadır (J. Bobadilla ve diğerleri. 2010). Denklem uygulanarak elde edilir (Denklem 4.2), burada I, hem u hem de j kullanıcıları tarafından öğelendirilen tüm öğeleri temsil eder. u kullanıcısından gelen tüm öğelendirmelerin ortalaması $\bar{r}_u$ ile temsil edilir ve j kullanıcısının tüm öğelendirmelerinin ortalaması $\bar{r}_j$ ile temsil edilir.

$$sim(u,j) = \frac{\sum_{i\in I}(r_{u,i}-\bar{r}_u)\times(r_{j,i}-\bar{r}_j)}{\sqrt{\sum_{i\in I}(r_{u,i}-\bar{r}_u)^2}\times\sqrt{\sum_{i\in I}(r_{j,i}-\bar{r}_j)^2}} \quad (4.2)$$

Denklemin sonucu (4.2)aralığında olacaktır, burada 1'e eşit bir benzerlik negatif bir korelasyona ve +1'e eşit bir benzerlik pozitif bir korelasyona karşılık gelir . 0'a yaklaşan

sayılar, iki kullanıcı arasında doğrusal bir ilişki olmadığını gösterir. Şekil 4.1'te, ortak dört öğeye sahip üç kullanıcının bir örneğini sunuyoruz. Birinci ve ikinci kullanıcıların puanlarına göre olumlu bir çağrışım var, her şeyi tam olarak aynı şekilde öğelendirmeseler bile ikisi arasında bir bağlantının varlığı kesindir.



Şekil 4.1. Kullanıcı öğelendirmeleri ve aralarındaki korelasyon örneği

4.3. Tahmin metriği

Tahmin metriği, insanlar (veya öğeler) arasındaki benzerliklerden yararlanarak öğe için bir tahmin yapmaktan sorumludur. Tahminler birçok farklı yöntemle yapılabilir ve en popüler olanları (J. Bobadillave diğerleri. 2010)'da ele alınmıştır. Çözümümüzde tahminler üretmek için en popüler Denklemi değiştirirsek aşağıdaki gibi sonuçlanır.

Denklem ile temsil edilen düzeltilmiş ağırlıklı toplamayı (ortalamadan sapma) kabul ettik, çünkü (Denklem 4.3) geliştirilmiş bir denklemdir. Bu denklem, her bir kullanıcı değerlendirmesinin ($r_{j,i} - \bar{r}_j$) ortalamasından sapmayı dikkate almamızı sağlar, en yaygın Denklemde olduğu gibi yalnızca $r_{j,i}$öğelendirmesini dikkate almak yerine denklemde, $p_{u,i}$ belirlenen tahmin değerini ifade eder. $\mu_{u,i}$ (Denklem 2.4)'te tanımlanır ve bir normalleştirme faktörüne karşılık gelir (S. Gabrielsson ve S. Gabrielsson. 2006).

$$p_{u,i} = \bar{r}_u + \mu_{u,i} \sum_{n \epsilon G_{u,i}} sim(u,j) \times (r_{j,i} - \bar{r}_j) \quad (4.3)$$

$$\mu_{u,i} = \frac{1}{\sum_{n \epsilon G_{u,i}} sim(u,j)} \Leftrightarrow G_{u,i} \neq \emptyset \quad (4.4)$$

4.4. SVD Kullanarak Ortak Filtreleme

Tekil Değer Ayrışımı (SVD) olarak bilinen öneri yaklaşımı başlangıçta bir vaka çalışması için kullanılmıştı (B. M. Sarwar. 2000). Öğelendirme matrisi program tarafından ayrıştırılır ve sonuç olarak, bir bütün olarak ÖS'nin doğruluğunu ve verimliliğini artıran boyut sayısını azaltmayı mümkün kılar.Yöntem aşağıda anlatıldığı gibi çalışır. Öğelendirme matrisi (X), SVD kullanılarak üç matrise ayrıştırılır: U, S ve V^T(D. Kalman. 1996: 2,23). Bu ayrıştırmanın nasıl gerçekleştirileceğine dair çok sayıda araştırma var çünkü öneri sürecindeki en fazla hesaplama gerektiren görevdir (J. A. Diaz-Garcia ve F. J. Caro-Lopera. 2010), (S. Lahabar ve P. Narayanan. 2009: 1,10). SVD, k boyutunda U, S ve V^T'yi kırparak problemin boyutunda azaltmaya izin verir. Şekil 4.2'de, n öğe sayısı, m kullanıcı sayısı ve k tutulan özelliklerin sayısıdır, n ve m'den küçük veya ona eşittir.

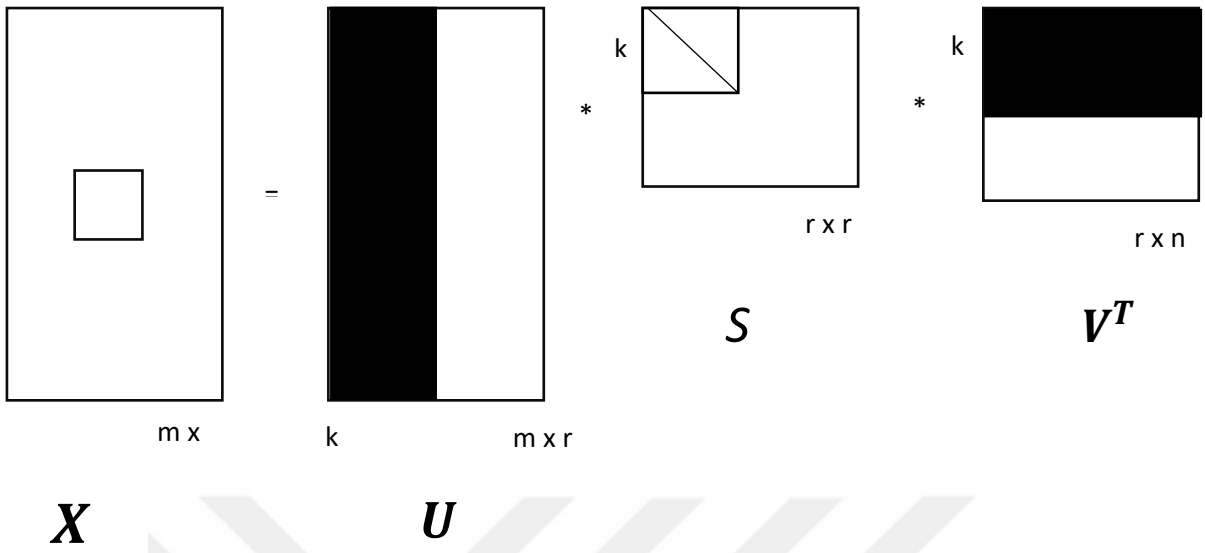
Ayrıştırılmış matrislerle ilgili olarak, U, her bir kullanıcıyı açıklayan özellikleri içerir. S, sırasıyla U ve V^T matrisi üzerindeki bir satır veya sütundaki her bir öğenin ağırlığı hakkında bilgi sağlar. V^T, öğeleri tanımlayan özelliklere sahip vektörleri içerir.Ayrıştırılmış matrisler kırpıldıktan sonra orijinal X matrisinin yeniden oluşturulamayacağını, dolayısıyla yaklaşık bir öğelendirme matrisi olan $\tilde{X}$e ile sonuçlanır.

(Denklem 4.5) ile herhangi bir kullanıcı (u) ve film (i) çifti için bir öğelendirme $(p_{u,i})$ tahmin etmek mümkündür. $\bar{r}_u$ değeri u'nun ortalama öğesini ve k ayrıştırılmış öğelendirme matrisinden seçilen özelliklerin sayısını temsil eder.

$$p_{u,i} = \bar{r}_u + U_k(u) \times \sqrt{S_k} \times V_k^T(i) \tag{4.5}$$

4.5. İçerik Tabanlı Filtreleme

Öğenin meta verileri, kullanıcı veya öğe hakkında kapsamlı bilgiler sunduğu için bazen önerileri iyileştirmek için kullanılabilir. Bu nedenle İçerik Tabanlı Filtreleme (İTF) olarak bilinen filtreleme yöntemi geliştirilmiştir. Tipik olarak, kullanıcıların halihazırda ilgi gösterdikleri nesnelerle ilgili nesneleri kullanır (X. Zhou ve diğerleri. 2012).



Şekil 4.2. SVD matrisinin U,S ve V^T'ye ayrıştırılması.

Filmler arasındaki benzerlikleri belirlemek için kullanılabilecek film açıklamalarına örnek olarak özetler, türler, yıldızlar, yönetmenler, anahtar sözcükler ve daha pek çoğu verilebilir (X. Zhou ve diğerleri. 2012).Bu açıklamaları önceden işlemek ve eşdeğer bir şeye çevirmek sıklıkla gereklidir. Metinle çalışırken, bir metin ön işleme tekniği tipik olarak, engellenecek sözcüklerin ortadan kaldırılması, kök bulma ve sözcük torbası temsilinin oluşturulması gibi bir dizi işlemi gerçekleştirmek için kullanılır (R. Correa ve B. Pinheiro. 2010:31,36).Engellenecek sözcüklerin kaldırılması, "the", "in", "on", "at", "from", "I", "gibi en yaygın ve genellikle alakalı olmayan terimleri metinden silen bir işlemdir.Terimleri daha da basitleştirmek için, benzer bir köke sahip terimlerin toplanması maksadımıza uygundur (N. Bhamidipati ve S. Pal. 2007: 350,360). Örneğin, "win", "winner" ve "wining" terimleri göz önüne alındığında, “win” formuna dönüşürlerdi. Bu nedenle, farklı terimlerin çeşitliliği ve sayısı azalacağından sistemin karmaşıklığı azaltılabilir.Durdurulacak sözcükler elenerek ve bir grup şeyin tüm metinsel tanımları çıkarılarak bir sözcük torbası temsili yapılabilir. Bunu yapmak için, her öğe, her boyutun bir sözcük olduğu bir vektör olarak temsil edilir. Metindeki her ifadenin ağırlığına sahip bir vektör, her öğeyi karakterize etmek için hesaplanabilir. Terim Frekansı-Ters Belge Frekansı(TF-IDF) olarak adlandırılan ağırlıklandırma yöntemi en sık kullanılan yöntemdir (R. Li ve X. Guo. 2010). (Denklem 4.6)'de gösterilen Terim Frekansı (TF), madde açıklamasındaki terimin alaka düzeyini belirler.Burada $f_{i,j}$, j teriminin her madde tanımı i'de bulunma sayısıdır.

$$TF_{i,j} = \frac{f_{i,j}}{\max\{f_{z,j}\}} \quad (4.6)$$

Ters Belge Frekansı(IDF), (Denklem4.7) Değerler, öğelerin koleksiyonundaki terimlerin ne kadar nadir olduğunu gösterir. N değişkeni mevcut kalem açıklamalarının sayısıdır ve n_j, j terimini içeren belgelerin toplamıdır. Bu nedenle, daha nadir terimler daha yüksek IDF ağırlığına sahiptir

$$IDF_j = log\frac{N}{n_j} \quad (4.7)$$

Son olarak, TF-IDF değeri, belgedeki terimin sıklığını (TF) dikkate alır ve onu, kalem koleksiyonundaki (IDF) terimin sıklığıyla ağırlıklandırır. it Denkleminin her bir elemanını hesaplamak için (Denklem 4.8) kullanılır.

$$\omega_{i,j} = TF_{i,j} \times IDF_i \quad (4.8)$$

Ardından, öğe açıklamasına karşılık gelen her metin, son olarak vektörle temsil edilir:

$$\vec{d_i} = \omega_{i,1}; \omega_{i,2}; \ldots; \omega_{i,k}) \quad (4.9)$$

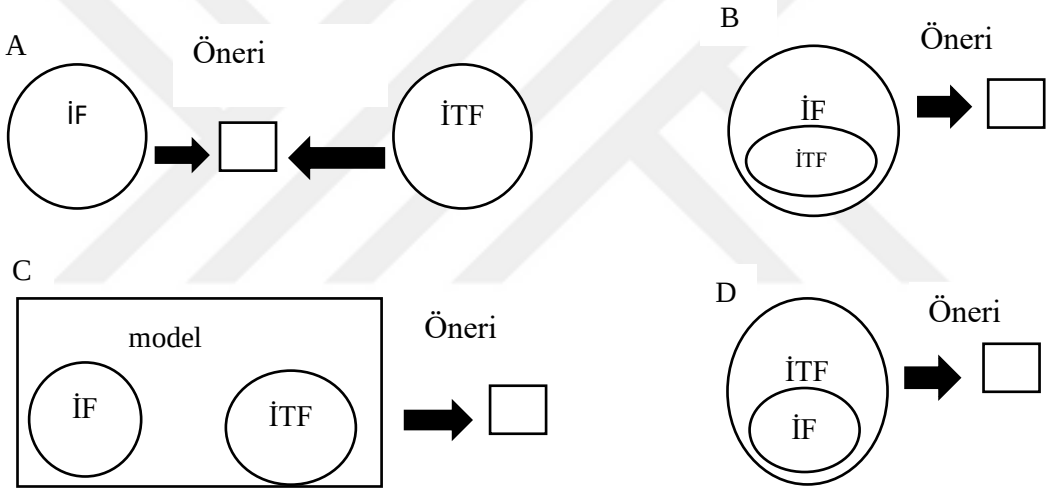
Bir kullanıcı daha sonra $\vec{d_i}$öğe açıklamalarının ortalaması ile temsil edilebilir. Örneğin, kullanıcı tarafından öğelendirilen tüm öğe açıklama vektörlerinin ortalamasını hesaplayarak. Ardından, kullanıcıları ve öğeleri karşılaştırılabilir hale getirerek, her bir kullanıcıyı temsil etmek için kullanılabilir. Her bir öğenin bir kullanıcıyla ilişkisini belirlemek için, kullanıcı ile öğeler arasında bir benzerlik metriği tanımlanmalıdır (G. Adomavicius ve A. Tuzhilin. 2005: 734,749). Bir ÖS yalnızca İTFkullanıyorsa, büyük olasılıkla aşırı uzmanlaşma sorunundan muzdarip olacaktır. Bu nedenle, genellikle İTF, İşbirliğine Dayalı Filtreleme (İF) gibi diğer filtreleme teknikleriyle birleştirilir (J. Bobadilla ve diğerleri. 2013), (W. Carrer-Neto ve diğerleri. 2012).

4.5. Hibrit filtreleme

- Hibrit filtreleme kullanan bir ÖS, daha önce özetlenen filtreleme yöntemleri birleştirilerek oluşturulabilir (P. C. Vaz ve diğerleri. 2012). Bunları birleştirmenin ÖS'nin genel standardını ve sunulan önerileri yükselteceği tahmin edilmektedir (W. Carrer-Neto ve diğerleri. 2012), (W. Carrer-Neto ve diğerleri. 2012). Herhangi bir filtreleme yöntemi birleştirilebilir olsa da, sistemimizde İF ve İTF kullanacağımız için sadece buna odaklanacağız. Şekil 4.3, iki filtreleme yöntemini birleştirmek için

çok sayıda yaklaşımı göstermektedir (G. Adomavicius and A. Tuzhilin. 2005: 734,749)

- Durum A: önerilerini bütünleştirmeden önce bir İF ve bir İTF'yi bağımsız olarak gerçekleştirmeyi içerir. Örneğinhem İF hem de İTF'nin tahmin sonuçlarının ortalamasını alırsak;
- Durum B: ÖS'ye yeni bir kullanıcı tanıtma sorununu çözmek için, kişisel bilgilerin İF'ye dahil edilmesi gibi çeşitli İTF özellikleri İF'ye aktarılır;
- Durum C: hem İF hem de İTF'den gelen özellikleri içeren genel bir model oluşturmak, ardından bulgularını bir makine öğrenme tekniğiyle birleştirmek (örneğin, Bayes-Networks kullanarak);
- Durum D: İTF'ye belirli İF özelliklerinin eklenmesini içerir. İTF algoritmasının İF'nin hesaplanan önerilerine ilişkin girdisini örnek olarak kullanmak.



Şekil 4.3. Hem İF hem de İTF kullanan çeşitli hibrit filtreleme modelleri

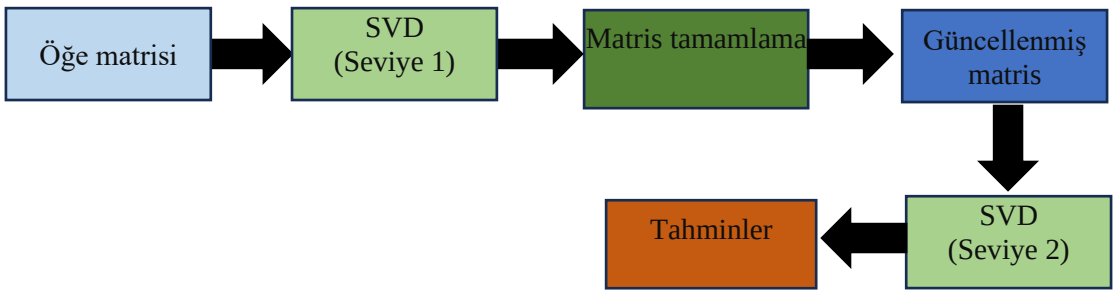
5. PRATİK UYGULAMA

Bu kısımda, geliştirilen sistemde kullanılan teknikler ve algoritmalar detaylı bir şekilde açıklanacaktır. Şekil 5.1, önerilen modelin akış şemasını göstermektedir. Seyrek matris azaltma ve modelin eğitimi için SVD (Tekil Değer Ayrışımı) yöntemi kullanılmıştır, bu nedenle bu yaklaşıma "İki Seviyeli SVD Tekniği" adı verilmiştir.

Önerilen yöntemde ilk olarak, kullanılan veri setine ait rating matrisindeki seyrekliği azaltma süreci uygulanmıştır. Bu amaçla, toplam rating sayısı kadar yeni kullanıcı-öğe çifti rasgele seçilmiştir.

Bu rasgele kullanıcıların rasgele içeriklere olan tahminleri hesaplanarak rating matrisi güncellenmiştir. Kullanıcı-öğe çifti için tahmin yapılırken, başlangıçtaki rating matrisi ile model eğitilmiştir ve Tekil Değer Ayrışımı (SVD) modeli kullanılmıştır.

Bu sayede, rating matrisinin yoğunluğu iki katına çıkarılmıştır. Yeni elde edilen matris kullanılarak SVD modeli eğitilmiş ve ardından test verileri kullanılarak tahminler gerçekleştirilmiştir.



Şekil 5.1. Önerilen modelin akış şeması

5.1. Ön İşleme Adımı

Bir kullanıcı tercihleri ağı oluşturmak (yani, öğelendirilmiş öğelerin toplanması Bu aşama, kullanıcıların belirli öğeleri ne kadar sık veya nadir kullandıklarını belirlemek ve bu bilgiyi bir ağ yapısı içinde temsil etmek için gereklidir. Bir kişi bir öğeyi öğelendirmezse, onu görmemiş kabul edilir. (U. Shardanand ve P. Maes. 1995). Bu çerçevedeki topluluklar öğelerin alt grupları ve bu çerçevedeki düğümler ise öğelerdir.

Bir ortak öğelendirme bağlantısı varsa (yani, nesneler arasında en az bir paylaşılan öğelendirme), o zaman öğeler aslında birbirleriyle etkileşime girer. Bir ilişkilendirme listesi oluşturmak amacıyla, 3'ten büyük öğelendirmelere sahip ortak oluşumlara dayalı olarak kullanıcı puanlarını çıkarmak için kullanılan yaklaşımı değiştirdik. Bu yöntem, her öğe çifti arasında bir sayı tahsis eder. Bitişiklik matrisi veya kenar listesi temelinde bir ağ oluşturulur.

5.2. Önerilen Yöntemin Etkinliği

Önerilen teknik ile İF yaklaşımındaki temel sorunlar topluluk algoritması ile birleştirilmiştir. Soğuk başlangıç ürünleri ortak öğelendirme etkileşimlerine dayalı olarak topluluklara zorla yerleştirildiğinden, soğuk başlangıç sorunuyla karşılaşılamaz. Tüm seyrek nesneleri önceden söylemek için bir öğelendirme tahmini kullanarak, seyreklik beklenebilir.

5.3. Deneysel Çalışma

5.3.1. Tekil değer ayrıştırması SVD

Veriler eğitim ve test setlerine ayırılır: R_train eğitim seti ve R_test test seti, kullanıcı-öğe matrisi R'den ayrılır. Öneri modelini eğitmek için kullanılan bilinen oranlar R_train'deyken, bilinen öğelendirmeler modelin etkililiğini değerlendirmek için R_test'te kullanılır.

Sürpriz kütüphane kullanarak eğitim setinde SVD modelini eğitilir: Tekil Değer Ayrıştırması (SVD), R_train kullanıcı-öğe matrisini üç matrise ayırır: U, Σ ve V^T(Maple .2018). SVD ayrışım denklemi aşağıdaki gibi ifade edilir:

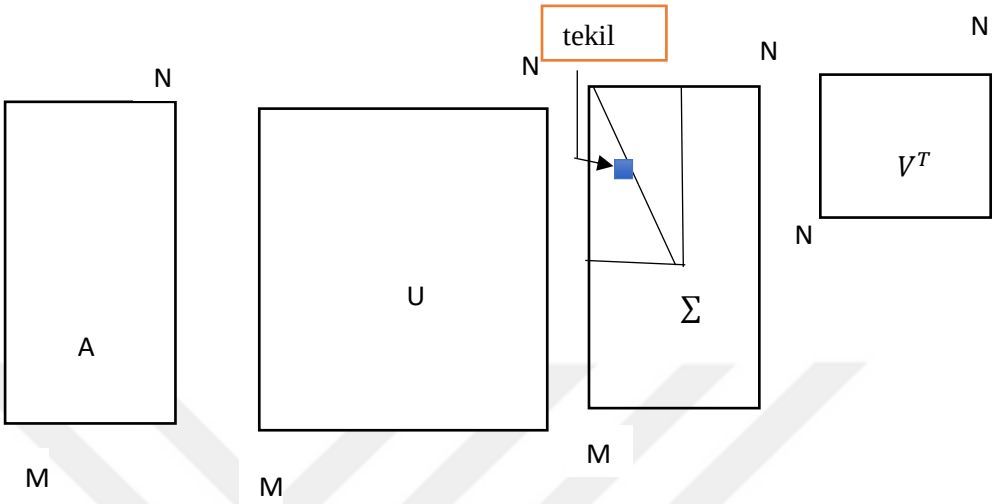
$$A \approx U \Sigma V^T \tag{5.1}$$

- R_train, eğitim user-item matrisidir.
- U, kullanıcı tercihlerini temsil eden kullanıcı matrisidir.
- Σ, tekil değerlerin köşegen matrisidir.
- V^T, madde özelliklerini temsil eden madde matrisidir.

SVD modelinin amacı, yeniden yapılandırma hatasını en aza indiren R_train kullanıcı-öğe matrisine düşük sıralı yaklaşımı bulmaktır (Gander, W., Gander, M. J., Kwok, F., Gander, W., Gander, M. J., & Kwok, F. ,2014).

Sürpriz kütüphane kullanarak birleştirilmiş veriler üzerinde SVD modelini eğitlir: SVD modeli, öncekiyle aynı SVD çarpanlarına ayırma sürecini izleyerek, birleştirilmiş kullanıcı-öğe matrisi R_updated üzerinde yeniden eğitilir (XIE, Pengpeng; WEI, Yimin; XIANG, Hua,2014).Test setinde tahminler yapın: Eğitilmiş SVD modeli, R_test test setinde tahminler yapmak için kullanılır (Akter, M. A ,2021). R_test'teki kullanıcı-film çiftleri için tahmin

edilen oranlar, yeniden eğitilmiş kullanıcı matrisi U ve madde matrisi V^T'nin karşılık gelen satırlarının iç çarpımı kullanılarak hesaplanır (Albright, Russ,2004).



Şekil 5.2. Matris Σ.

5.3.2. Büyük matris

Bu matriste, U'ların ve V'lerin yalnızca birim matris sütunları olduğu daha kapsamlı bir örnek bulunmaktadır. Bu nedenle hesaplamalar basittir ancak sütunların nasıl düzenlendiğine dikkat edilmesi gerekir (Frostig, R., Musco, C., Musco, C., & Sidford, A. , 2016) A matrisi kesinlikle üçgendir ve orantısızdır. Genel olarak sıfır özdeğere sahiptir. ATA, AAT'ye yakın değildir (Maeder, Marcel; Neuhold, Yorck-Michael, 2007).

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

özdeğerler λ = 0, 0, 0, 0 hepsi sıfır !

sadece bir özvektör (1, 0, 0, 0)

tekil değerler σ = 3, 2, 1

SVD modelinin amacı, yeniden yapılandırma hatasını en aza indiren Rmatrisine düşük sıralı yaklaşımı bulmaktır (Gander, W., Gander, M. J., Kwok, F., Gander, W., Gander, M. J., & Kwok, F.,2014). Python surprise kutuphanesi kullanarak birleştirilmiş veriler üzerinde SVD modeli eğitilir.

SVD modeli, öncekiyle aynı SVD çarpanlarına ayırma sürecini izleyerek, güncellenen kullanıcı-öğe matrisi (updated matrix) üzerinde yeniden eğitilir. Eğitilmiş SVD modelinin

performansı için önceden belirlenen test seti ile tahminler gerçekleştirilir ve performans analizleri gerçekleştirilir.

SVD yönteminde boyut indirgeme süreci yer aldığı için regularizasyon süreci yer almaktadır. Diğer bir deyişle SVD modelinde belirlenen k sayısı normaldeki matrisin boyutundan daha düşük olması sebebiyle hatalar oluşmaktadır.

Bu nedenle oluşturulan modelin eğitim süresince mevcut eğitim seti kullanılarak hata oranları aza indirgenir.

Bu maksatla öğrenme ve regularizasyon parametreleri (α ve λ) kullanılmaktadır. Bunun yanı sıra SVD modelinin boyutunu indirgememizi sağlayan k parametresinin de optimum olarak belirlenmesi gerekmektedir.

5.3.3. Veri seti

MovieLens veri kümesi, GroupLens tarafından derlenmiştir ve filmlerle ilgili kullanıcı öğelendirmelerini içermektedir. Veri kümesi, movielens.umn.edu web sitesi aracılığıyla yedi aylık bir süre boyunca toplanmıştır.

Bu öğelendirmeler, kullanıcıların filmlere verdikleri puanları 1'den 5'e kadar olan sayısal bir ölçekte içermektedir. Ayrıca, kullanıcıların demografik özellikleri, yaş, cinsiyet, meslek ve posta kodu gibi bilgiler de veri kümesinde bulunmaktadır(B. N. Miller, I. Albert, S. K. Lam, J. A. Konstan and J. Riedl, 2003).

Toplam 1682 farklı film için 943 tekil kullanıcıdan toplamda 100,000 öğelendirme bulunmaktadır.

Kullanıcılar, her biri en az 20 film için öğelendirme sağlamışlardır. Veri kümesindeki filmler, 18 farklı türe ayrılmıştır.

Veri kümesi iki ana dosyadan oluşmaktadır:

1.'u.data': Bu dosya, 100,000 öğelendirmeyi içermektedir. Her giriş, kullanıcının belirli bir filme verdiği öğelendirmeye karşılık gelmektedir. Dosyadaki sütunlar şunlardır: kullanıcı kimliği, öğe kimliği (film kimliği), öğelendirme ve zaman damgası.

'u.data' dosyasından bir örnek giriş:

Kullanıcı Kimliği: 196

Film Kimliği: 242

Öğelendirme: 3

Zaman Damgası: 881250949

2.'u.item': Bu dosya, filmlerle ilgili detaylı bilgileri içermektedir. Her satır belirli bir filme aittir ve filmin başlığı, çıkış tarihi, video çıkış tarihi, IMDb URL'si ve tür kategorizasyonları gibi verileri içermektedir.

'u.item' dosyasından bir örnek satır:

Film Kimliği: 242

Başlık: "Kolye"

Çıkış Tarihi: 1998

Türler: Drama|Romantik

MovieLens veri kümesi, öneri sistemleri alanında araştırma ve deneylerde temel bir kaynak olarak kullanılmaktadır.

Araştırmacılar ve veri bilimcileri, bu veri kümesini kullanarak çeşitli öneri algoritmalarını geliştirebilir, değerlendirebilir ve kullanıcıların tercihleri ve ilgi alanlarına yönelik öğe önerilerinin doğruluğunu ve etkisini inceleyebilirler.

5.3.4. Sistemin uygulanması

Program Python 3.6.7rc1 kullanılarak uygulandı. Algoritma, PyCharm 2022.2 Community Edition IDE kullanılarak oluşturuldu. Toplamda her biri algoritmanın genel akışına katkıda bulunan dört ayrı dosya oluşturuldu. Dosyalar boyutlarına ve oluşturulma nedenlerine göre ayrıldı.

***Dış kütüphaneler**:*

Kod, Sürpriz kütüphanesini kullanarak bir öneri sistemi oluşturma ve değerlendirmede farklı görevleri kolaylaştırmak için çeşitli harici kütüphanelerden yararlanır. Her kitaplığı ayrıntılı olarak incelenerek ve bunların kodda nasıl kullanıldığını görsterilmiştir:

Suprise kütüphanesi:

Yazılımda kullanılan kütüphaneler aşağıda verilmiştir.

- **Surprise**, öneri sistemlerini oluşturmak ve analiz etmek için özel olarak tasarlanmış bir Python kütüphanesidir.

  Tekil Değer Ayrışımı (SVD) gibi matris çarpanlara ayırma yöntemleri de dahil olmak üzere çeşitli işbirliğine dayalı filtreleme algoritmaları sağlar.

- **SVD**, öneri sistemlerinde İF için kullanılan bir matris çarpanlara ayırma algoritmasıdır. Tahminlerde bulunmak için kullanıcıların ve öğelerin gizli faktörlerini öğrenen işbirlikçi bir yaklaşımdır.

- **Dataset** ve **Reader,** modelleri eğitmek ve test etmek amacıyla verileri yüklemek ve ön işlemek için kullanılan Sürpriz'deki sınıflardır.

- **train_test_split**, verileri eğitim ve test setlerine bölmek için kullanılır.

- **GridSearchCV**, ızgara aramasını kullanarak hiperparametre ayarlaması gerçekleştirmek için kullanılır.

- **accuracy**, model değerlendirmesi için RMSE (Ortalama Kare Hatanın Kökü) ve MAE (Ortalama Mutlak Hata) gibi ölçümlerin hesaplanmasına yönelik işlevler sağlar.

NumPy Library:

- **import numpy as np**
- **NumPy** Python ile bilimsel hesaplama için temel bir pakettir. Diziler, matrisler ve matematiksel işlevler için destek sağlar.
- **"numpy"** dizileri, özellikle de kullanıcı öğesi öğelendirmelerini tutan **"MyMatrix"** dizisini oluşturmak ve değiştirmek için kullanılır.

Pandas Library:

- **import pandas as pd**
- **Pandas** veri işleme ve analiz için güçlü bir kütüphanedir. Verileri düzenlemek ve analiz etmek için DataFrame gibi veri yapıları sağlar.
- **pandas**, CSV dosyalarından öğelendirme verilerini okumak ve işlemek ve DataFrames oluşturmak için kullanılır.

Matplotlib and Seaborn Libraries:

- **import matplotlib.pyplot as plt**
- **import seaborn as sns**

- **Matplotlib,** Python'da statik, etkileşimli ve animasyonlu görselleştirmeler oluşturmaya yönelik popüler bir çizim kitaplığıdır.
- **Seaborn**, Matplotlib'i temel alan bir istatistiksel veri görselleştirme kütüphanesidir. Bilgilendirici ve ilgi çekici istatistiksel grafikler oluşturmak için üst düzey bir arayüz sağlar.

Warnings Module:

- **import warnings**
- **warnings** modülü, kod yürütme sırasında belirli uyarı mesajlarını bastırmak için kullanılır.
- **warnings**, geçerli kod yürütmeyle alakalı olmayabilecek gelecekteki belirli uyarı türlerini yok saymak için kullanılır.

Functions

- **train_and_evaluate_svd(learning_rate, lambda1, n_factoTS):** belirtilen hiper parametrelerle bir SVD modelini eğitir ve değerlendirir. Algoritma nasıl çalışıyor:
    - İşlev üç parametre alır: **learning_rate**, **lambda1**, and **n_factoTS** (number of latent factoTS).
    - öğrenme hızı ve düzenleme parametreleri dahil olmak üzere verilen hiper parametrelerle bir SVD modeli (**MY_model**) oluşturulur.
    - Model, **trenseti** verileriyle eğitilir.
    - Tahminler, test setindeki **trained** model kullanılarak oluşturulur.
    - function, Surprise'ın **accuracy** modülünü kullanarak Ortalama Karekök Hatayı (RMSE) ve Ortalama Mutlak Hatayı (MAE) hesaplar.
    - RMSE ve MAE değerleri döndürülür.
- **Main Loop for Hyperparameter Tuning:** Bu **loop performs,** bir ızgara arama yaklaşımını kullanarak hiperparametre ayarlamasını gerçekleştirir:
    - Parametre aralıklarında belirtilen **learning_rate,lambda1** ve **n_factoTS** kombinasyonları üzerinde yinelenir.
    - Her kombinasyon için, bu hiper parametrelerle bir SVD modelini eğitmek ve değerlendirmek üzere train_and_evaluate_svd işlevini çağırır.
    - Ortaya çıkan RMSE ve MAE değerleri, hiperparametrelerle birlikte resulTS_df DataFrame'e eklenir.
- **External Python Built-in Functions:**
    - **random.sample:** Rastgele modülündeki bu işlev, bir listeden belirli sayıda öğeyi rastgele örneklemek için kullanılır (rastgele kullanıcı ve film kimlikleri oluşturmak için kullanılır).

6. DENEYSEL SONUÇLAR

Bu araştırma, 12th Gen Intel(R) Core(TM) i5-1235U 1.30 GHz işlemci ve 16 GB belleğe sahip, Windows 11 işletim sistemine sahip bir bilgisayar üzerinde gerçekleştirildi. Python'un surprise kütüphanesi kullanılarak Movielens 100k ve Movielens 1M veri setleri üzerinde çalışma yapıldı. Veri setleri, %80 eğitim ve %20 test verisi olarak ayrılarak eğitim ve test işlemleri gerçekleştirildi. Ayrıca, çalışmanın performansını değerlendirmek için Mean Absolute Error (MAE) ve Root Mean Square Error (RMSE) metrikleri kullanıldı.

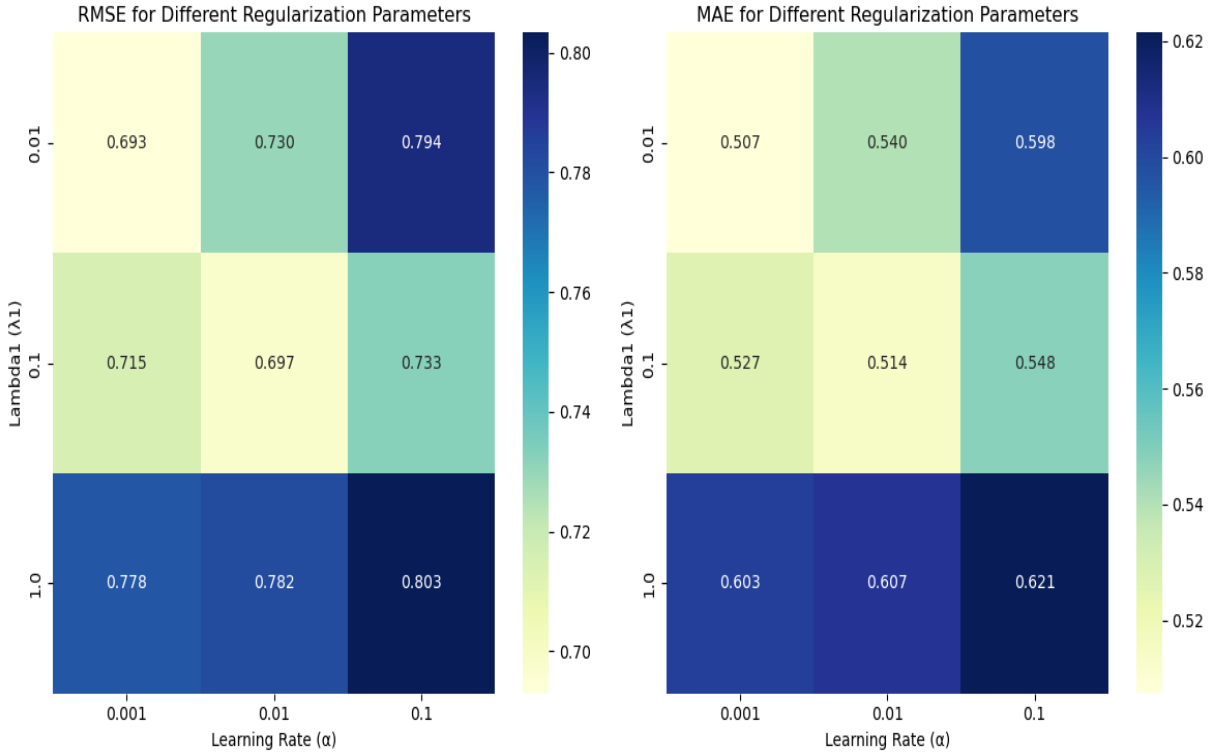
Çalışmada kullanılan veri setlerine özgü olarak, SVD modeli için en iyi öğrenme ve düzenleme parametreleri gridsearch tekniği kullanılarak belirlendi. Elde edilen sonuçlara göre, Her iki veri seti için de α ve λ parametreleri sırasıyla 0.01 ve 0.1 olarak belirlendi.Bu parametreler, veri setine en uygun öğrenme ve düzenleme sağlamaktadır.

0.01 ve 0.1 değerlerinin α ve λ parametreleri olarak seçilmesinin nedeni, gridsearch tekniği kullanılarak elde edilen en iyi performansı sağlamasıdır. Gridsearch tekniği, belirli bir parametre aralığında farklı değerleri deneyerek en iyi performansı sağlayan parametreleri belirlemek için kullanılır.

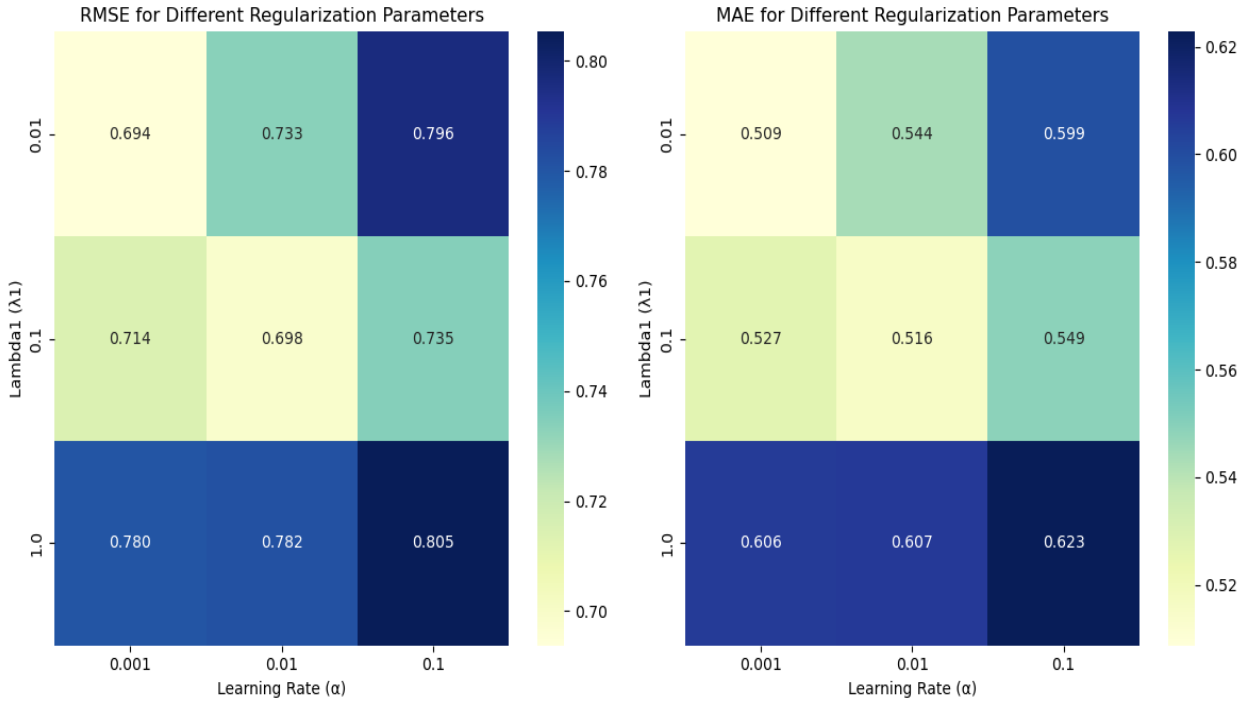
Bu durumda, α ve λ parametrelerinin Movielens 100k ve Movielens 1M veri setleri için 0.01 ve 0.1 olarak seçilmesi, bu değerlerin MAE ve RMSE metriklerinde en düşük hata oranlarını sağladığı anlamına gelir. Yani, bu parametrelerin seçilmesinin temel nedeni, modelin en iyi performansı sergilediği ve en düşük hata oranlarını elde ettiği değerleri temsil etmesidir.

Bu değerlerin seçilme nedeni, genellikle yapılan deneyler ve karşılaştırmalar sonucunda elde edilen en iyi sonuçlarla ilişkilendirilir.

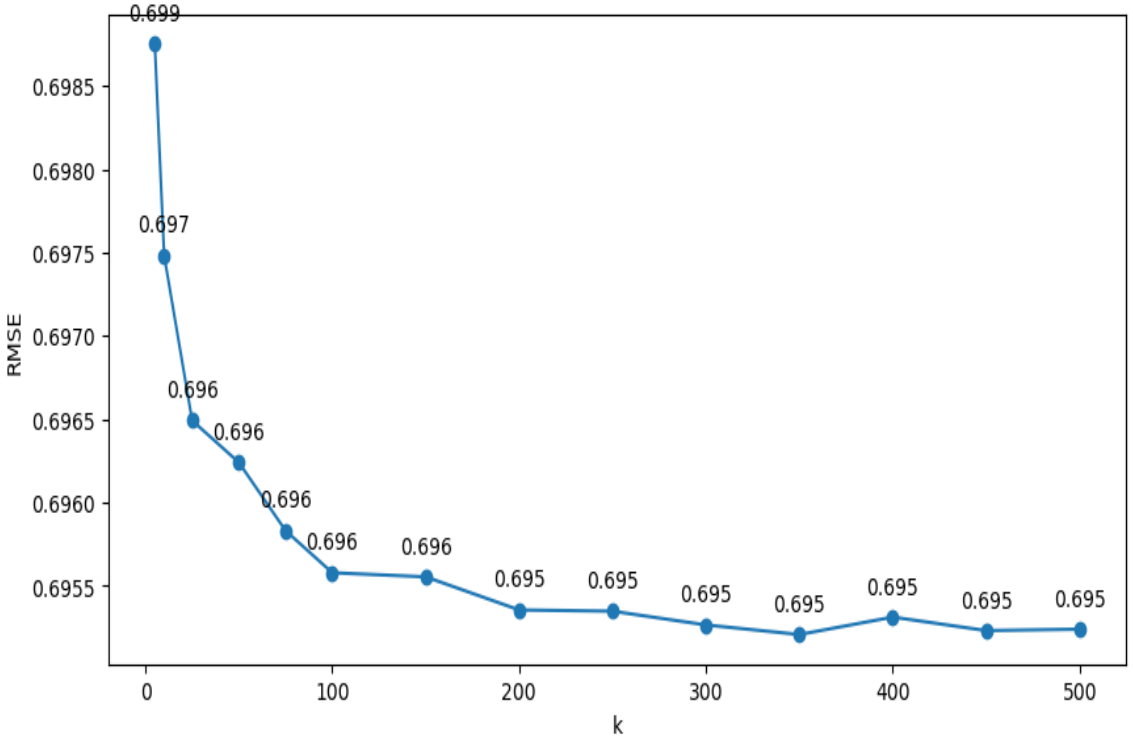
Şekil 6.1 ve Şekil 6.2 da sırasıyla Movielens 100k ve Movielens 1M için sırasıyla α ve λ parametrelerine göre RMSE ve MAE metrik sonuçları gösterilmektedir. Bu analiz, çalışmanın performansını daha ayrıntılı bir şekilde değerlendirmeye olanak tanımaktadır.



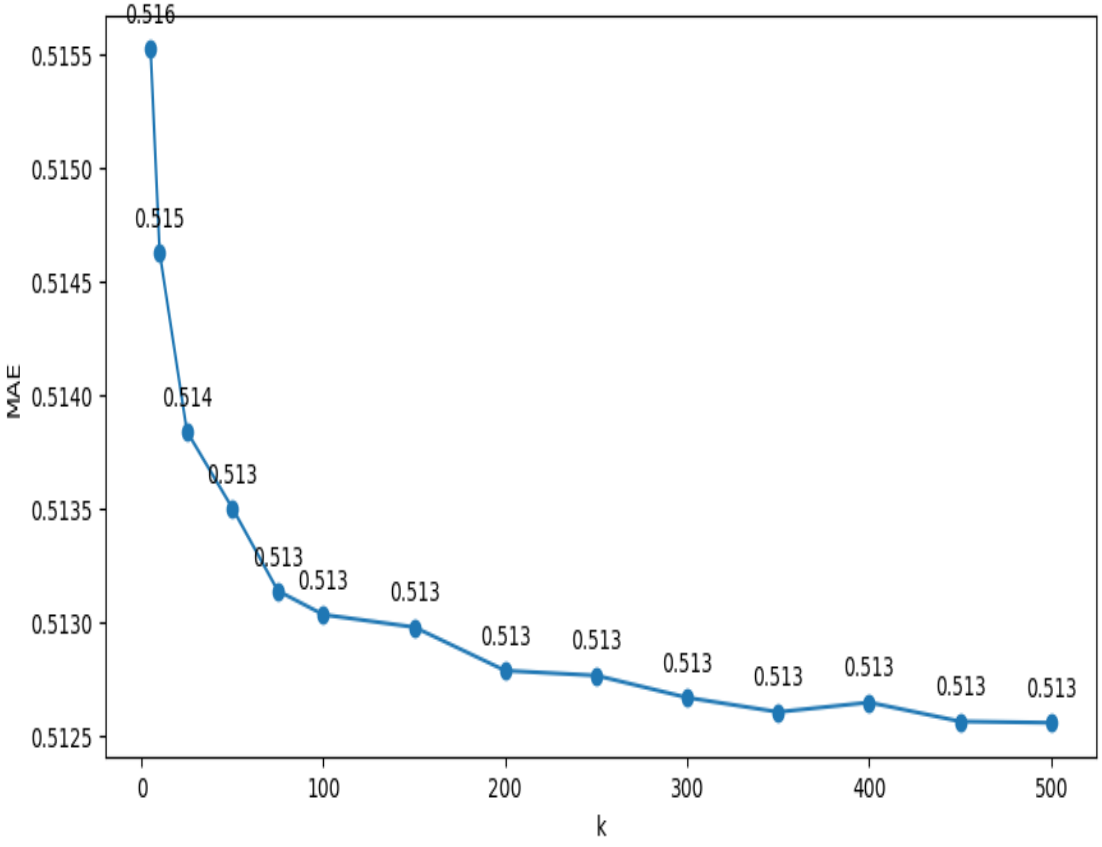
Şekil6.1.α ve λ parametrelerinegöre RMSE ve MAE sonuçları (Movielens 100k)



Şekil 6.2.α ve λ parametrelerinegöre RMSE ve MAE sonuçları (Movielens 1M)

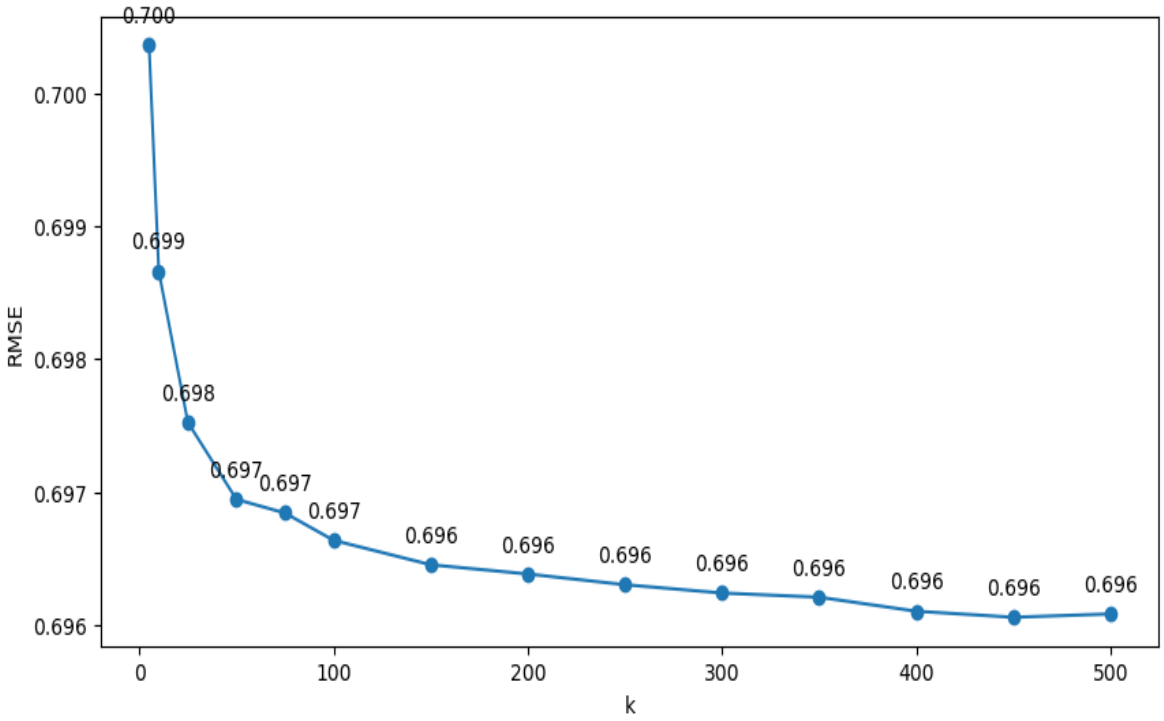


Şekil6.3.α ve λ parametrelerinegöre RMSE ve MAE sonuçları(Movielens 100k)

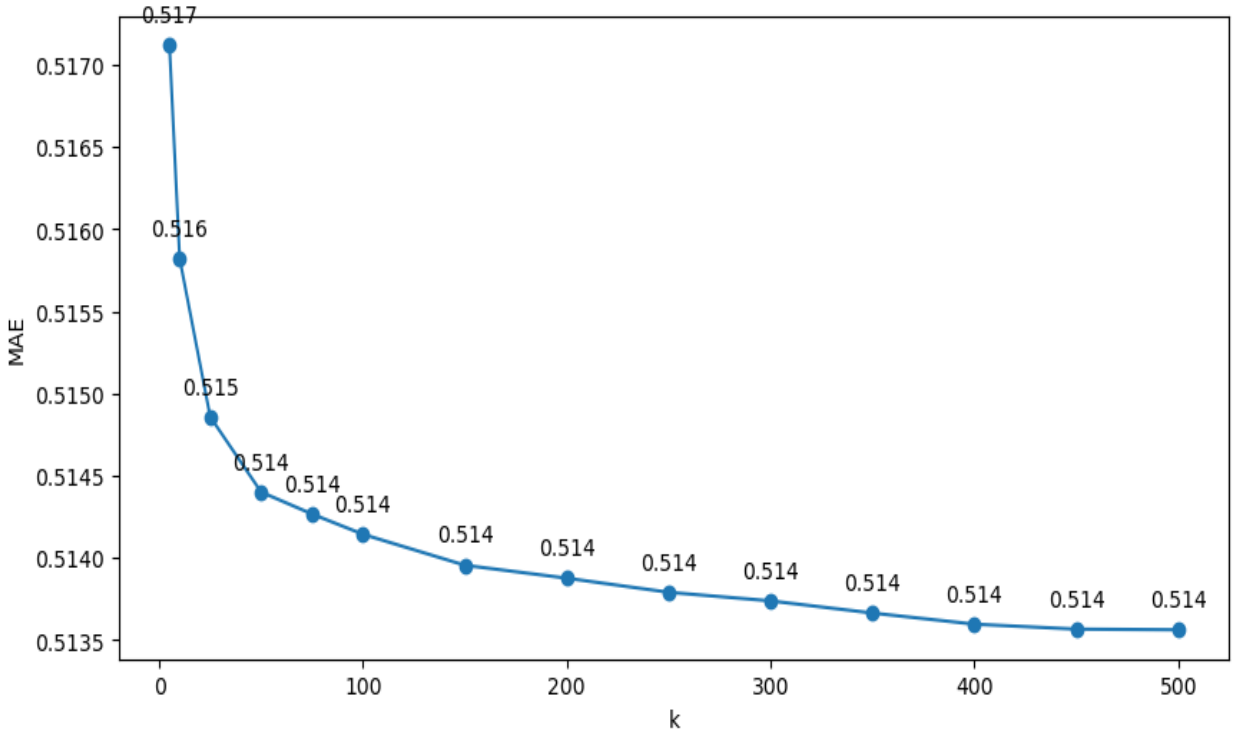


Şekil6.4. k parametresinegöre MAE sonuçları(Movielens 100k)

Şekil 6.3 ve Şekil 6.4'te Movielens 100k veriseti için, önerilen çalışmada kullanılan SVD modelindeki k parametresinin değişimine göre elde edilen performans değişimleri gösterilmektedir. Burada görüldüğü üzere k sayısının belirli bir değerden sonraki değişimleri sonucu ciddi etkilememektedir. Bu nedenle k sayısı 200 olarak belirlenmiştir. Yine aynı şekilde movielens 1M veriseti için şekil 6.5 ve şekil 6.6 incelendiğinde görüleceği üzere k sayısı 200 olarak uygun olduğu gözlenmiş ve seçilmiştir. Bunun yanısıra k sayıısnın her iki veri seti için performans değişimlerini görmek için Tablo 6.1 ve tablo 6.2'e bakılabilir.



Şekil 6.5. k parametresinegöre RMSE sonuçları(Movielens 1M)



Şekil6.6. k parametresinegöre MAE sonuçları(Movielens 1M)

Tablo 6.1.Farklı α=0.01 ve λ=0.1 durumlarındafarklı k değerlerinegöreeldeedilensonuçlar (Movielens 100k)

Sıra	Özellik Sayısı(k)	RMSE	MAE
0	5	0,699133	0,515702
1	10	0,697903	0,514902
2	25	0,696493	0,513837
3	50	0,696241	0,513499
4	75	0,695832	0,513135
5	100	0,695331	0,513213
6	150	0,695255	0,512976
7	200	0,695157	0,512784
8	250	0,695149	0,512763
9	300	0,695067	0,512665
10	350	0,695009	0,512602
11	400	0,695114	0,512644
12	450	0,695183	0,512561
13	500	0,695141	0,512555

Tablo 6.2. Farklı α=0.01 ve λ=0.1 durumlarındafarklı k değerlerinegöreeldeedilensonuçlar(Movielens 1M)

Sıra	Özellik Sayısı(k)	RMSE	MAE
0	5	0,700363	0,517119
1	10	0,698661	0,515821
2	25	0,697523	0,514858
3	50	0,696945	0,514400
4	75	0,696842	0,514266
5	100	0,696637	0,514143
6	150	0,696453	0,513955
7	200	0,696383	0,513876
8	250	0,696303	0,513790
9	300	0,696240	0,513738
10	350	0,696209	0,513664
11	400	0,696103	0,513597
12	450	0,696058	0,513566
13	500	0,696083	0,513562

Ayrıca önerilen teknik ile elde edilen sonuçları kıyaslamak maksadıyla baseline, KNN, SVD, NMF teknikleriyle deneyler gerçekleştirilmiştir. Elde edilen sonuçlar Tablo 6.3 de listelenmektedir. Movielens 100k ve Movielens 1M veri setleri için elde edilen sonuçlar sırasıyla Tablo 6.3 ve Tablo 6.4'de listelenmektedir.Buna göre önerilen modelin performansında önemli öğede artış sağlandığı görülmektedir.

Tablo 6.3. Performans Karşılaştırması

Model	**MAE**	**RMSE**
Baseline	0,9403	0,7460
KNN	0,9798	0,7727
SVD	0,9133	0,7263
NMF	0,9581	0,7519
Proposed	**0,7033**	**0,5134**

Tablo 6.4. Performans Karşılaştırması (Movielens 1M)

Model	RMSE	MAE
Baseline	0,9089	0,7192
KNN	0,9310	0,7309
SVD	0,8781	0,6845
NMF	0,9041	0,7172
Proposed	0,6963	0,5138

Önerilen çalışmanın zaman karmaşıklığı m toplam kullanıcı sayısını, n toplam öge sayısını ve k değeri de boyut indirgeme sonrası elde edilen özellik sayısını temsil etmek üzere, O(mnk)'dır. Önerilen çalışmada k değeri kullanıcı ve öge matrisinin toplam özellik sayısına göre çok küçük ve sabit değer aldığı için (örneğin bu çalışmada k=200 olarak belirlenmiştir.) zaman karmaşıklığı O(mn) olmaktadır.

7. SONUÇ VE ÖNERİLER

Elde edilen sonuçlar, önerilen modelin diğer alternatif modellere göre üstün performans sergilediğini göstermektedir. Özellikle, MAE ve RMSE metriklerine dayanarak yapılan değerlendirmeler, önerilen modelin daha düşük hata oranlarına sahip olduğunu göstermiştir. Model seçimi ve performans karşılaştırması, gridsearch tekniğiyle belirlenen öğrenme ve düzenleme parametrelerinin kullanılmasıyla gerçekleştirilmiş, bu da önerilen modelin belirgin şekilde iyileştirilmiş performans sergilediğini ortaya koymuştur.

Bu çalışma, öneri sistemleri alanındaki araştırmalara katkı sağlamak amacıyla yürütülmüştür. Özellikle, gerçek dünya senaryolarında yüksek performans elde etmek için öğrenme ve düzenleme parametrelerini optimize etme metodolojisinin etkinliği incelenmiştir. Çalışmanın sonuçları, önerilen modelin gerçek kullanım senaryolarında daha iyi öneriler sunma potansiyeline sahip olduğunu göstermektedir.
Genel olarak, bu çalışma, öneri sistemleri alanındaki araştırmalara katkıda bulunmayı hedefleyerek, gerçek dünya veri kümesi ve optimize edilmiş model parametreleri kullanarak daha etkili öneri sonuçları elde etme konusunda önemli bulgular sunmaktadır.

Bu çalışma, öneri sistemlerinin performansını artırmak ve daha etkili sonuçlar elde etmek amacıyla gerçekleştirilmiştir. Araştırmanın sonuçlarına dayanarak, gelecekte benzer çalışmalar yapacak araştırmacılara bazı öneriler sunulabilir:

Model Geliştirme ve Derinleştirme: Bu çalışmada kullanılan öneri modeli temel bir SVD modeli olup, ileriki çalışmalarda daha gelişmiş model türleri (örneğin, derin öğrenme tabanlı modeller) de incelenebilir. Farklı model yapıları ve algoritmaların performansını karşılaştıran çalışmalar, öneri sistemlerinin daha iyi anlaşılmasını sağlayabilir.

Veri Zenginleştirme ve İyileştirme: öneri sistemlerinin performansı doğrudan kullanılan veri kümesiyle ilişkilidir. Veri zenginleştirme teknikleri kullanarak eksik verileri doldurmak veya gürültülü verileri temizlemek, öneri sonuçlarını olumlu yönde etkileyebilir.

Kişiselleştirme ve Uyarlamalar: öneri sistemlerinin amacı kullanıcılara kişiselleştirilmiş öneriler sunmaktır. Gelecekteki çalışmalarda, kullanıcı davranışlarına ve tercihlerine daha fazla odaklanan kişiselleştirme stratejileri incelenebilir.
Daha Fazla Metrik ve Değerlendirme: Bu çalışmada MAE ve RMSE metrikleri kullanılmıştır. Gelecekteki araştırmalar farklı metriklerin kullanılmasıyla model performansının daha kapsamlı bir şekilde değerlendirilmesine odaklanabilir.

Gerçek Kullanım Senaryolarının İncelenmesi: öneri sistemlerinin gerçek kullanım senaryolarında nasıl performans gösterdiği, kullanıcı deneyimi, dönüşüm oranları ve diğer işletme metrikleri ile birleştirilerek incelenebilir.

Farklı veri kümesi ve alanlarında uygulama: Bu çalışma MovieLens 100k veri kümesi üzerine odaklanmıştır. Farklı veri kümesi ve farklı alanlarda (örneğin, ürün önerileri, müzik önerileri) yapılan benzer çalışmalar, genelleme kabiliyetini daha iyi anlamamıza yardımcı olabilir.

Bu öneriler, gelecekteki araştırmacıların bu alanda daha fazla derinlemesine çalışma yapmasına ve öneri sistemlerinin performansını daha da artırmasına yardımcı olabilir.

KAYNAKLAR

Adamopoulos, P., Tuzhilin, A. (2014). On unexpectedness in recommender systems: Or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(4), 1-32.

Adomavicius, G., Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions On Knowledge and Data Engineering*, 17(6), 734-749.

Bell, R. M., Koren, Y. (2007, August). Improved neighborhood-based collaborative filtering. *13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 7-14.

Bennett, J., Elkan, C., Liu, B., Smyth, P., Tikk, D. (2007). Kdd cup and workshop 2007. *ACM SIGKDD explorations newsletter*, 9(2), 51-52.

Bobadilla, J., Ortega, F., Hernando, A., Glez-de-Rivera, G. (2013). A similarity metric designed to speed up, using hardware, the recommender systems k-nearest neighbors algorithm. *Knowledge-Based Systems*, 51, 27-34.

Bobadilla, J., Ortega, F., Hernando, A., Gutiérrez, A. (2013). Recommender systems survey. *Knowledge-Based Systems*, 46, 109-132.

Bobadilla, J. E. S. U. S., Serradilla, F., Hernando, A. (2009). Collaborative filtering adapted to recommender systems of e-learning. *Knowledge-Based Systems*, 22(4), 261-265.

Bobadilla, J., Serradilla, F., Bernal, J. (2010). A new collaborative filtering metric that improves the behavior of recommender systems. *Knowledge-Based Systems*, 23(6), 520-528.

Burke, R. (2002). Hybrid recommender systems: Survey and experiments. *User Modeling and User-Adapted Interaction*, 12, 331-370.

Carrer-Neto, W., Hernández-Alcaraz, M. L., Valencia-García, R., García-Sánchez, F. (2012). Social knowledge-based recommender system. Application to the movies domain. *Expert Systems with Applications*, 39(12), 10990-11000.

Castells, P., Hurley, N., Vargas, S. (2021). Novelty and diversity in recommender systems. *In Recommender Systems Handbook.* New York, NY: Springer US. 603-646.

Chen, T., Zhang, W., Lu, Q., Chen, K., Zheng, Z., Yu, Y. (2012). SVDFeature: a toolkit for feature-based collaborative filtering. *The Journal of Machine Learning Research,* 13(1), 3619-3622.

Choi, S. M., Ko, S. K., Han, Y. S. (2012). A movie recommendation algorithm based on genre correlations. *Expert Systems with Applications,* 39(9), 8079-8085.

Corrêa, R. F., Pinheiro, B. F. (2010, October). Self-organizing maps applied to information retrieval of dissertations and theses from BDTD-UFPE. *In 2010 IEEE Eleventh Brazilian Symposium on Neural Networks*, 31-36.

Dell'Amico, M., Capra, L. (2008, June). Sofia: Social filtering for robust recommendations. *In IFIP International Conference on Trust Management*. Boston, MA: Springer US. 135-150.

Diaz-Garcia, J. A., Caro-Lopera, F. J. (2010). Shape theory via SVD decomposition I. *arXiv preprint arXiv,*1003-4780.

Ekstrand, M. D., Riedl, J. T., Konstan, J. A. (2011). Collaborative filtering recommender systems. *Foundations and Trends in Human–Computer Interaction,* 4(2), 81-173.

El Aissaoui, O., El Alami El Madani, Y., Oughdir, L., El Allioui, Y. (2019). A fuzzy classification approach for learning style prediction based on web mining technique in e-learning environments. *Education and Information Technologies,* 24, 1943-1959.

El-Bishouty, M. M., Aldraiweesh, A., Alturki, U., Tortorella, R., Yang, J., Chang, T. W., Graf, S. (2019). Use of Felder and Silverman learning style model for online course design. *Educational Technology Research and Development,* 67(1), 161-177.

Gabrielsson, S., Gabrielsson, S. (2006). *The use of self-organizing maps in recommender systems.* Master's Thesis, Computer Science, Uppsala University.

George, T., Merugu, S. (2005, November). *A scalable collaborative filtering framework based on co-clustering. In Fifth IEEE International Conference on Data Mining (ICDM'05)* (pp. 4-10). IEEE.

İnternet: Geyik, S. C., Kenthapadi, K. (2018). Building representative talent search at LinkedIn.LinkedIn engineering blog post. URL:https://engineering.linkedin.com/blog/2018/10/building-representative-talent-search-at-linkedin. Son Erişim Tarihi: 22.10.2018.

Geyik, S. C., Guo, Q., Hu, B., Ozcaglar, C., Thakkar, K., Wu, X., Kenthapadi, K. (2018, June). Talent search and recommendation systems at LinkedIn: Practical challenges and lessons learned. *In The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval*, 1353-1354.

Gauch, S., Speretta, M., Chandramouli, A., Micarelli, A. (2007). User profiles for personalized information access. *The Adaptive Web: Methods and Strategies of Web Personalization*, 54-89.

Goldberg, D., Nichols, D., Oki, B. M., Terry, D. (1992). Using collaborative filtering to weave an information tapestry. *Communications of the ACM*, 35(12), 61-70.

Guan, X., Li, C. T., Guan, Y. (2016). Enhanced SVD for collaborative filtering. In Advances in Knowledge Discovery and Data Mining. *20th Pacific-Asia Conference, PAKDD 2016*, Auckland, New Zealand, 503-514.

Hofmann, T. (2004). Latent semantic models for collaborative filtering. *ACM Transactions on Information Systems*, 22(1), 89-115.

Horasan, F. (2022). Latent Semantic Indexing-Based Hybrid Collaborative Filtering for Recommender Systems. *Arabian Journal for Science and Engineering*, 47(8), 10639-10653.

Huang, W. S., Li, S., Kim, S. W., Lee, K. (2018). Data imputation using a trust network for recommendation via matrix factorization. *Computer Science and Information Systems*, 15(2), 347-368.

Hu, L., Cao, J., Xu, G., Cao, L., Gu, Z., Zhu, C. (2013, May). Personalized recommendation via cross-domain triadic factorization. *In Proceedings of the 22nd international conference on World Wide Web*, 595-606.

Jung, J. J. (2012). Attribute selection-based recommendation framework for short-head user group: An empirical study by MovieLens and IMDB. *Expert Systems with Applications*, 39(4), 4049-4054.

Kalman, D. (1996). A singularly valuable decomposition: the SVD of a matrix. *The College Mathematics Journal*, 27(1), 2-23.

Kardan, A. A., Ebrahimi, M. (2013). A novel approach to hybrid recommendation systems based on association rules mining for content recommendation in asynchronous discussion groups. *Information Sciences,* 219, 93-110.

Kim, B. M., Li, Q., Park, C. S., Kim, S. G., Kim, J. Y. (2006). A new approach for combining content-based and collaborative filters. *Journal of Intelligent Information Systems,* 27, 79-91.

Kim, M., Park, S. O. (2013). Group affinity based social trust model for an intelligent movie recommender system. *Multimedia Tools and Applications,* 64, 505-516.

Kittler, J., Hatef, M., Duin, R. P., Matas, J. (1998). On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 20(3), 226-239.

Koren, Y. (2009, June). Collaborative filtering with temporal dynamics. *In Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 447-456.

Kunaver, M., Požrl, T. (2017). Diversity in recommender systems–A survey. *Knowledge-Based Systems,* 123, 154-162.

Lahabar, S., Narayanan, P. J. (2009, May). Singular value decomposition on GPU using CUDA. *IEEE International Symposium on Parallel & Distributed Processing,* 1-10.

Lekakos, G., Caravelas, P. (2008). A hybrid approach for movie recommendation. *Multimedia Tools and Applications,* 36, 55-70.

Li, R., Guo, X. (2010, October). An improved algorithm to term weighting in text classification. *International Conference on Multimedia Technology*, 1-3.

Li, S., Kim, S. W., Lee, K. (2018). Data imputation using a trust network for recommendation via matrix factorization. *Computer Science and Information Systems*, 15(2), 347-368.

Linden, G., Smith, B., York, J. (2003). Amazon. com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1), 76-80.

Lu, J., Wu, D., Mao, M., Wang, W., Zhang, G. (2015). Recommender system application developments: a survey. *Decision Support Systems*, 74, 12-32.

Luo, X., Zhou, M., Li, S., You, Z., Xia, Y., Zhu, Q. (2015). A nonnegative latent factor model for large-scale sparse matrices in recommender systems via alternating direction method. *IEEE Transactions on Neural Networks and Learning Systems*, 27(3), 579-592.

Ma, H., King, I., Lyu, M. R. (2011). Learning to recommend with explicit and implicit social relations. *ACM Transactions on Intelligent Systems and Technology (TIST),* 2(3), 1-19.

Mandl, M., Felfernig, A., Teppan, E., Schubert, M. (2011). Consumer decision making in knowledge-based recommendation. *Journal of Intelligent Information Systems,* 37, 1-22.

Miller, B. N., Albert, I., Lam, S. K., Konstan, J. A., Riedl, J. (2003, January). Movielens unplugged: experiences with an occasionally connected recommender system. *In Proceedings of the 8th International Conference On Intelligent User Interfaces,* 263-266.

Ricci, F., Rokach, L., Shapira, B. (2010). *In Recommender Systems Handbook.* Boston, MA: Springer US. 1-35

Ren, Y., Li, G., Zhang, J., Zhou, W. (2013). Lazy collaborative filtering for data sets with missing values. *IEEE Transactions on Cybernetics*, 43(6), 1822-1834.

Sarwar, B., Karypis, G., Konstan, J., & Riedl, J. T. (2000). *Application of dimensionality reduction in recommender system-a case study.*

Sarwar, B., Karypis, G., Konstan, J., Riedl, J. (2001, April). Item-based collaborative filtering recommendation algorithms. *In Proceedings of the 10th international conference on World Wide Web*, 285-295.

Schafer, J. B., Frankowski, D., Herlocker, J., & Sen, S. (2007). *Collaborative filtering recommender systems. In The adaptive web: methods and strategies of web personalization* ,291-324. Berlin, Heidelberg: Springer Berlin Heidelberg.

Shardanand, U., Maes, P. (1995, May). Social information filtering: Algorithms for automating "word of mouth". *In Proceedings of the SIGCHI conference on Human factors in computing systems*, 210-217.

Shi, Y., Larson, M., Hanjalic, A. (2014). Collaborative filtering beyond the user-item matrix: A survey of the state of the art and future challenges. *ACM Computing Surveys (CSUR)*, 47(1), 1-45.

Smith, B., Linden, G. (2017). Two decades of recommender systems at Amazon. com. *IEEE Internet Computing*, 21(3), 12-18.

İnternet: Spotify - For the Record. (n.d.). Company info. Retrieved March. URL: https://newsroom.spotify.com/company-info/, Son Erişim Tarihi: 16.2.2021.

Symeonidis, P., Nanopoulos, A., Manolopoulos, Y. (2009, October). MoviExplain: a recommender system with explanations. *In Proceedings of the Third ACM Conference on Recommender Systems*, 317-320.

Tan, T. F., Netessine, S. (2009). Is Tom Cruise threatened? Using Netflix Prize data to examine the long tail of electronic commerce. *Wharton Business School*, University of Pennsylvania, Philadelphia, 1-37.

Trewin, S. (2000). Knowledge-based recommender systems. *Encyclopedia of Library and Information Science*, 69(32), 180.

Vaz, P. C., Martins de Matos, D., Martins, B., Calado, P. (2012, June). Improving a hybrid literary book recommendation system through author ranking. *In Proceedings of the 12th ACM/IEEE-CS joint Conference on Digital Libraries*, 387-388.

Xie, J., Szymanski, B. K. (2013, April). Labelrank: A stabilized label propagation algorithm for community detection in networks. *In 2013 IEEE 2nd Network Science Workshop (NSW)*, 138-143.

Yuan, X., Han, L., Qian, S., Xu, G., Yan, H. (2019). Singular value decomposition based recommendation using imputed data. *Knowledge-Based Systems*, 163, 485-494.

Zhang, L. (2013). The Definition of Novelty in Recommendation System. *Journal of Engineering Science & Technology Review*, 6(3),20-32

Zhang, S., Yao, L., Sun, A., Tay, Y. (2019). Deep learning based recommender system: A survey and new perspectives. *ACM computing surveys (CSUR)*, 52(1), 1-38.

Zhou, X., Xu, Y., Li, Y., Josang, A., Cox, C. (2012). The state-of-the-art in personalized recommender systems for social networking. *Artificial Intelligence Review*, 37, 119-132.

ÖZGEÇMİŞ

Kişisel Bilgiler

Soyadı,adı : MOHAMMED, Mohammed Jameel
Uyruğu : Irak
Doğumtarihiveyeri
Medenihali
Telefon
Faks
e-mail

Eğitim

Öğe	EğitimBirimi	MezuniyetTarihi
Yükseklisans	GaziÜniversitesi/Bilgisayar Bilimleri	Devam ediyor
Lisans	MusulÜniversitesi/Bilgisayar Bilimleri	2020
Lise	Erbil/Mustekbal lisesi	2016

YabancıDil

İngilizce,Arapça

Yayınlar

A NOVEL COLLABORATIVE FILTERING-BASED RECOMMENDATION SYSTEM USING TWO-LEVEL SVD AND MATRIX COMPLETION, EGE 9th INTERNATIONAL CONFERENCE ON APPLIED SCIENCES
SEPTEMBER 22 - 24, 2023 - IZMIR

Mohammed J. ,Murat Y. , Fahrettin H. (2023). A NOVEL COLLABORATIVE FILTERING-BASED RECOMMENDATION SYSTEM USING TWO-LEVEL SVD AND MATRIX COMPLETION. *EGE 9th INTERNATIONAL*



GAZİ GELECEKTİR