



**FATİH SULTAN MEHMET VAKIF ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ PROGRAMI**

**KULLANICI VE ÖGE TEMELLİ YÖNTEMLERİN
BİRLİKTE KULLANILDIĞI İŞBİRLİKÇİ
FİLTRELEME TEKNİĞİ GELİŞTİRİLMESİ**

YÜKSEK LİSANS TEZİ

YUSUF ÖZDEMİR

İSTANBUL, 2021



**FATİH SULTAN MEHMET VAKIF ÜNİVERSİTESİ
LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
BİLGİSAYAR MÜHENDİSLİĞİ PROGRAMI**

**KULLANICI VE ÖGE TEMELLİ YÖNTEMLERİN
BİRLİKTE KULLANILDIĞI İŞBİRLİKÇİ
FİLTRELEME TEKNİĞİ GELİŞTİRİLMESİ**

YÜKSEK LİSANS TEZİ

**YUSUF ÖZDEMİR
(180221016)**

**Danışman
(Dr. Öğr. Üyesi Ali NİZAM)**

İSTANBUL, 2021

25/05/2021

LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ MÜDÜRLÜĞÜNE

Bilgisayar Mühendisliği Anabilim Dalı'nda 180221016 numaralı Yusuf ÖZDEMİR'in hazırladığı "Kullanıcı ve Öge Temelli Yöntemlerin Birlikte Kullanıldığı İşbirlikçi Filtreleme Tekniği Geliştirilmesi " konulu Bilgisayar Mühendisliği Yüksek Lisans tezi ile ilgili Tez Savunma Sınavı, 25/05/2021 Salı günü saat 14:00 'da yapılmış, sorulara alınan cevaplar sonunda adayın tezinin **KABULÜNE** karar verilmiştir.

Düzeltilme verilmesi halinde:

Adı geçen öğrencinin Tez Savunma Sınavı .../.../20... tarihinde, saat ...:.. da yapılacaktır.

Tez Adı Değişikliği Yapılması Halinde: Tez adının

.....
şeklinde değiştirilmesi uygundur.

| Jüri Üyesi | Tarih | İmza |
|-------------------------------------|------------|-------|
| (Danışman) Dr. Öğr. Üyesi Ali NİZAM | 25/05/2021 | KABUL |
| Prof. Dr. Ayşe Şima UYAR | 25/05/2021 | KABUL |
| Dr. Öğr. Üyesi Berna KİRAZ | 25/05/2021 | KABUL |

BEYAN/ ETİK BİLDİRİM

Bu tezin yazılmasında bilimsel ahlak kurallarına uyulduğunu, başkalarının eserlerinden yararlanılması durumunda bilimsel normlara uygun olarak atıfta bulunulduğunu, kullanılan verilerde herhangi bir tahrifat yapılmadığını, tezin herhangi bir kısmının bağlı olduğum üniversite veya bir başka üniversitedeki başka bir çalışma olarak sunulmadığını beyan ederim.

Yusuf ÖZDEMİR

İmza

TEŐEKKÜR

Bu tezin yazılmasında en büyük katkıyı saęlayan ve her konuda destek olan danıőman hocam Dr. Öğr. Üyesi Ali Nizam Bey'e; bu çalışmam esnasında bir baba olarak yapmam gereken görevlerimi de üstlenerek her türlü desteęi veren çok kıymetli eşim Esmâ Özdemir'e teşekkürü bir borç bilirim.

Yusuf ÖZDEMİR

İmza

**KULLANICI VE ÖGE TEMELLİ YÖNTEMLERİN BİRLİKTE
KULLANILDIĞI İŞBİRLİKÇİ FİLTRELEME TEKNİĞİ
GELİŞTİRİLMESİ
Yusuf Özdemir**

ÖZET

Bu çalışmada öneri sistemlerinde işbirlikçi filtreleme tekniklerinden “kullanıcı temelli işbirlikçi filtreleme” ve “öge temelli işbirlikçi filtreleme” birlikte kullanılıp, bunlara ek olarak kullanıcı özelliklerinden de faydalanan daha önce çalışılmamış hibrit bir öneri sistemi modellenip uygulama haline getirilmiştir. Ayrıca önerilen çözüm öneri sistemlerindeki soğuk başlangıç problemine yönelik kullanılabilir. Yeni bir üye geldiğinde geçmişini olmadığı için bu kişiye ne önereceğini sistem tam olarak bilemez. Bu probleme soğuk başlangıç problemi denir.

Öneri sistemlerinde eldeki verilere bakılıp benzerlik kurularak tavsiyeler oluşturulur. Klasik öneri sistemlerinde öncelikle pearson korelasyonu ile kullanıcı kullanıcı benzerliği veya öge öge benzerliği hesaplanır. Çalışılan modelde pearson ile kullanıcıların ögelere verdikleri oylar üzerinde benzerlik katsayıları çıkarıldı. Tahmin mekanizmalarıyla birlikte yaygın olarak kullanımı bilinen sınıflandırma algoritmalarından K ortalama kümeleme algoritmasının yanı sıra bu alanda önerilen yöntem benzer kullanımı olmayan karınca kolonisi optimizasyon algoritması (ACO), genetik algoritma ve yapay arı kolonisi algoritması gibi farklı algoritmalar araştırılmış ve ACO'ya dayalı kümeleme yapılmıştır. Bu kümelerdeki kullanıcıların birbirine benzediğini kabul edilip bunlar üzerinden tahminler yapılmaktadır. Önerilen modelde film önerisinde bulunmak için temel olarak 3 farklı türde tahmin yapılmaktadır. Bunlar klasik tahmin hesabı (Pearson, Öklid vb.), kullanıcı özelliklerini işin içine sokarak yapılan tahmin (yaş, cinsiyet vb.) ve kümeleme algoritmaları ile tahmin (karınca kolonisi algoritması, K ortalama kümeleme vb.).

Çıkan bu sonuçların tümü için doğruluk hesaplaması yapılarak modelin işlevselliği test edilmiştir. Sonuçlar önerilen yeni hibrit sistemin klasik yöntemden daha iyi sonuçlar verdiğini göstermiştir.

Anahtar kelimeler; “Öneri sistemleri”, “işbirlikçi filtreleme”, “kullanıcı temelli işbirlikçi filtreleme”, “öge temelli işbirlikçi filtreleme”, “soğuk başlangıç”.

HYBRID SYSTEM WITH COMBINING USER – BASED AND ITEM BASED COLLABORATIVE FILTERING

Yusuf Özdemir

ABSTRACT

In this study, by using collaborative filtering techniques "user-based collaborative filtering" and "item-based collaborative filtering" in recommendation systems, a hybrid recommendation system that has not been studied before has been modeled and made into an application, which also benefits from user features. Also, the proposed solution can be used to solve cold start problem. When a new member arrives, the system does not know exactly what to recommend to this person, as he has no history. This problem is called the cold start problem.

Recommendations are created by looking at the available data and establishing a similarity in the recommendation systems. In classical recommendation systems, pearson correlation and user-user similarity or item-item similarity are calculated first. In the model studied, similarity coefficients on the rates given by the users to the items with pearson were calculated. In addition to the K average clustering algorithm, which is one of the widely used classification algorithms with prediction mechanisms, different algorithms such as the ant colony optimization algorithm, genetic algorithm and artificial bee colony algorithm, which there is no similar usage with recommended system in this field, were researched and ACO was preferred to cluster data. It is accepted that the users in these clusters are alike and estimates are made on them. Basically, 3 different types of predictions are made to recommend a movie in the proposed model. These are classical prediction calculations (Pearson, Euclid etc.), prediction made by involving user characteristics (age, gender, etc.), and prediction with clustering algorithms (ant colony algorithm, K mean clustering, etc.).

The functionality of the model was tested by calculating the accuracy for all of these results. The results showed that the proposed new hybrid system gives better results than the conventional method.

Keywords; "Recommendation systems", "collaborative filtering", "user-based collaborative filtering", "item based collaborative filtering", "cold start".

ÖNSÖZ

Bir yönlendirme sistemi tarafından sağlanan öneriler, kullanıcıları hangi öğelerin satın alınacağı, hangi haberlerin okunacağı veya hangi müziğin dinleneceği gibi çeşitli karar verme süreçlerinde desteklemeyi amaçlamaktadır. Öneri oluşturma için çeşitli teknikler geliştirildi ve bunların çoğu ticari ortamlarda da başarıyla çalıştırıldı. Bu tekniklerin başında kullanıcı temelli işbirlikçi filtreleme ve öge temelli işbirlikçi filtreleme gelir. Bu iki teknik, diğer teknikler gibi ayrı ayrı uygulanmaktadır. Tezin amacı farklı teknikleri bir arada kullanmak ve yeni yöntemler geliştirmektir.

Hali hazırda öneri sistemlerindeki soğuk başlangıç probleminin yanı sıra yeni denenen yöntemlerde zorlukları beraberinde getirmiştir. Yeni fikirler ile istenen sonuçlar arasındaki makas bazen çok yıpratıcı olabilmektedir.

Bu çalışmada kullanıcı ve öge bilgileri kullanılarak Pearson uzaklığı hesaplanmıştır. Uzaklık değerleri ile K ortalama ve ACO yöntemlerinin birlikte kullanıldığı yeni bir hibrit öneri sistemi önerilmiştir. Sistem test sonuçları önerilen sistemin doğruluğu arttırdığını göstermektedir.

Bu tezin yazım aşamasında birçok problemle karşılaşmıştır. Her gerektiğinde katkı sağlayan ve hafta sonlarını da feda eden danışman hocam Dr. Öğretim Üyesi Ali Nizam Bey'e; bu çalışmam esnasında bir baba olarak yapmam gereken görevlerimi de üstlenerek her türlü desteği veren çok kıymetli eşim Esmâ Özdemir'e teşekkürü bir borç bilirim.

İÇİNDEKİLER

| | |
|---|------|
| ÖZET..... | v |
| ABSTRACT | vii |
| ÖNSÖZ..... | ix |
| SEMBOLLER | xii |
| ŞEKİL LİSTESİ..... | xiii |
| ÇİZELGE LİSTESİ..... | xiii |
| KISALTMALAR | xiv |
| GİRİŞ | 1 |
| BİRİNCİ BÖLÜM..... | 5 |
| 1. LİTERATÜR ARAŞTIRMASI | 5 |
| 1.1. ÖNERİ SİSTEMLERİ..... | 5 |
| 1.2. ÖNERİ SİSTEMLERİNİN İŞLEVİ..... | 9 |
| 1.3. VERİ VE BİLGİ KAYNAKLARI | 13 |
| 1.4. ÖNERİ TEKNİKLERİ | 16 |
| 1.5. İŞBİRLİKÇİ FİLTRELEME..... | 23 |
| 1.5.1. Kullanıcı Temelli İşbirlikçi Filtreleme | 24 |
| 1.5.2. Öge Temelli İşbirlikçi Filtreleme | 24 |
| 1.5.3. Bellek Temelli İşbirlikçi Filtreleme | 25 |
| 1.5.4. Model Temelli İşbirlikçi Filtreleme..... | 25 |
| 1.5.5. Hibrit İşbirlikçi Filtreleme | 25 |
| 1.6. ÖNERİ SİSTEMLERİ DEĞERLENDİRMESİ | 26 |
| 1.7. ÖNERİ SİSTEMLERİ UYGULAMALARI | 28 |
| 1.8. ÖNERİ SİSTEMLERİ VE İNSAN BİLGİSAYAR ETKİLEŞİMİ..... | 32 |
| 1.9. ÖNERİ SİSTEMİNDE KARŞILAŞILAN ZORLUKLAR..... | 36 |
| 1.9.1. Soğuk Başlangıç Problemi..... | 36 |
| İKİNCİ BÖLÜM | 38 |
| 2. METOTLAR | 38 |
| 2.1. SİSTEMİN TEMEL MODELİ..... | 38 |
| 2.2. KORELASYON VE BENZERLİK HESAPLAMALARI..... | 40 |

| | | |
|-----------------------------|---|-----------|
| 2.2.1. | Pearson Korelasyon Katsayısı..... | 40 |
| 2.2.2. | Öklid Uzaklığı..... | 41 |
| 2.2.3. | K – En Yakın Komşuluk | 42 |
| 2.3. | K – ORTALAMA KÜMELEME..... | 42 |
| 2.4. | BAĞLI BİLEŞEN ETİKETLEME | 43 |
| 2.5. | SÜRÜ ZEKÂSI | 43 |
| 2.5.1. | Karınca Kolonisi Optimizasyon Algoritması | 44 |
| 2.5.2. | Genetik Algoritma..... | 47 |
| 2.5.3. | Yapay Arı Kolonisi Algoritması | 48 |
| 2.6. | ACO İLE KÜMELEME VE TAHMİN YÖNTEMLERİ | 49 |
| 2.6.1. | ACO Algoritması ile Kümeleme | 49 |
| 2.6.2. | ACO ile Kümeleme Uygulaması | 50 |
| 2.7. | ACO İLE BİRİKEN KOKULARIN KMEANS KÜMELEME YÖNTEMİYLE BİRLİKTE KULLANIMI..... | 50 |
| ÜÇÜNCÜ BÖLÜM | | 51 |
| 3. | BULGULAR..... | 51 |
| 3.1. | VERİ HAZIRLAMA VE ÖN İŞLEME SÜREÇLERİ | 51 |
| 3.2. | MODEL ÖLÇÜM METRİKLERİ | 53 |
| 3.2.1. | Karesel Ortalama Hata..... | 53 |
| 3.2.2. | Kök Ortalama Kare Hatası..... | 54 |
| 3.2.3. | Ortalama Mutlak Hata (MAE) | 54 |
| 3.3. | DOĞRULANMIŞ MODEL SONUÇ KARŞILAŞTIRMALARI..... | 55 |
| DÖRDÜNCÜ BÖLÜM | | 66 |
| 4. | TARTIŞMA VE SONUÇ | 66 |
| KAYNAKÇA | | 67 |

SEMBOLLER

- ρ : Spearman sıralama korelasyon katsayısı
- r : Pearson Korelasyon katsayısı
- N_{xy} : ACO algoritmasında durumlar arası geçişin istenirliği
- β : ACO algoritmasında N_{xy} etkisini kontrol etmek için bir parametredir.

ŞEKİL LİSTESİ

| | |
|--|----|
| ŞEKİL 3-1 SİSTEMİN TEMEL MODELİ | 39 |
| ŞEKİL 4-1 ACO KÜMELEME İLE KARESEL ORTALAMA HATA. | 61 |
| ŞEKİL 4-2 ACO KÜMELEME İLE KÖK ORTALAMA KARE HATA. | 61 |
| ŞEKİL 4-3 ACO KÜMELEME İLE ORTALAMA MUTLAK HATA. | 62 |
| ŞEKİL 4-4 ACO + K ORTALAMA KÜMELEME VE KARESEL ORTALAMA HATA..... | 63 |
| ŞEKİL 4-5 ACO + K ORTALAMA KÜMELEME VE KÖK ORTALAMA KARE HATA..... | 64 |
| ŞEKİL 4-6 ACO + K ORTALAMA KÜMELEME VE ORTALAMA MUTLAK HATA. | 64 |

ÇİZELGE LİSTESİ

| | |
|---|----|
| ÇİZELGE 4-1 SEÇİLMİŞ 3000 TANE DERECELENDİRME İÇİN TAHMİN SONUÇLARI..... | 55 |
| ÇİZELGE 4-2 RASGELE %20'Sİ SEÇİLEREK YAPILAN DERECELENDİRMEDE EĞİTİM SETİ İÇİN TAHMİN SONUÇLARI..... | 56 |
| ÇİZELGE 4-3 EĞİTİM SETİ U1.BASE TEST SETİ U1.TEST İÇİN TAHMİN SONUÇLARI | 56 |
| ÇİZELGE 4-4 EĞİTİM SETİ U2.BASE TEST SETİ U2.TEST İÇİN TAHMİN SONUÇLARI | 57 |
| ÇİZELGE 4-5 EĞİTİM SETİ U3.BASE TEST SETİ U3.TEST İÇİN TAHMİN SONUÇLARI. | 57 |
| ÇİZELGE 4-6 EĞİTİM SETİ U4.BASE TEST SETİ U4.TEST İÇİN TAHMİN SONUÇLARI | 58 |
| ÇİZELGE 4-7 EĞİTİM SETİ U5.BASE; TEST SETİ U5.TEST İÇİN TAHMİN SONUÇLARI. | 58 |
| ÇİZELGE 4-8 EĞİTİM SETİ U1-5.BASE; TEST SETİ U1-5.TEST İÇİN TAHMİN SONUÇLARININ ORTALAMASI. | 59 |
| ÇİZELGE 4-9 EĞİTİM SETİ U5.BASE; TEST SETİ U5.TEST İÇİN TAHMİN SONUÇLARININ STANDART SAPMALARININ ORTALAMASI. | 59 |

KISALTMALAR

| | |
|---------------|---|
| a.e. | Aynı eser/yer |
| a.g.e. | Adı geçen eser |
| a.y. | Yazara ait son zikredilen yer |
| b.a. | Eserin bütününe atıf |
| bkz. | Bakınız |
| bkz.: aş. | Eserin kendi içinde aşağıya atıf |
| bkz.:yuk. | Eserin kendi içinde yukarıya atıf |
| C. | Cilt |
| çev. | Çeviren |
| ed. veya haz. | Editör/yayına hazırlayan |
| k.g. | Karşı görüş |
| karş. | Karşılaştırmamız |
| s. | Sayfa/sayfalar |
| t.y. | Basım tarihi yok |
| v.d. | Çok yazarlı eserlerde ilk yazardan sonrakiler |
| y.y. | Basım yeri yok |

| | |
|-------|--|
| ÖS | Öneri Sistemleri |
| İF | İşbirlikçi Filtreleme |
| UBCF | Kullanıcı Temelli İşbirlikçi Filtreleme |
| IBCF | Öge Temelli İşbirlikçi Filtreleme |
| MBCF | Bellek Temelli İşbirlikçi Filtreleme |
| MOBCF | Model Temelli İşbirlikçi Filtreleme |
| ACO | Karınca Kolonisi Algoritması |
| ABC | Yapay Arı Kolonisi Algoritması |
| BCO | Meta-Sezgisel Arı Kolonisi Optimizasyonu |
| GA | Genetik Algoritması |
| EA | Evrimsel Algoritması |
| KNN | K En Yakın Komşuluk Algoritması |
| CCL | Bağlı Bileşen Etiketleme |
| MSE | Karesel Ortalama Hata |
| RMSE | Kök Ortalama Kare Hatası |
| MAE | Ortalama Mutlak Hata |

GİRİŞ

Öneri Sistemleri (ÖS), belirli bir kullanıcının ilgisini en çok çekebilecek öğeler için öneriler sağlayan yazılım araçları ve tekniklerdir[1]. Öneri sistemi, öğelere ilişkin tavsiyelerde bulunmak için öğeler ve / veya yardımcı bilgiler üzerindeki geçmiş kullanıcı derecelendirmelerini kullanan belirli bir akıllı sistem türüdür. Çok çeşitli çevrimiçi alışveriş, e-ticaret hizmetleri ve sosyal ağ uygulamalarında kritik bir rol oynar. Öneri sistemleri internet ile paralel olarak gelişmiştir. İnternetin gelişmesiyle birlikte, çevrimiçi olarak mevcut bilgiler katlanarak arttı. Bu aşırı bilgi yüklemesi, gereksiz bilgileri kaldırabilecek ve minimum zamanda bir kullanıcıya en değerli bilgileri sağlayabilecek bir sistem gerektiriyordu. İşbirlikçi Filtreleme (İF), bu tür bilgileri sağlamak için en doğru ve yaygın olarak benimsenen yaklaşımlardan biridir. İF uygulama alanları e-ticaret ve e-öğrenmeden sosyal ağlara ve web aramaya kadar değişen geniş bir yelpazededir. İşbirlikçi filtrelemeyle ilgili geniş alanı, teknikleri ve zorlukları nedeniyle, temel ilkesini, çalışmasını ve bütünüyle uygulamasını anlamak için bütün bilgilerin tek bir yerde toplanmasını gerektirir[2].

İşbirlikçi filtreleme için motivasyon, insanların genellikle kendilerine benzer zevklere sahip birinden en iyi önerileri aldıkları fikrinden gelir. İşbirlikçi filtreleme, benzer ilgi alanlarına sahip kişileri eşleştirme ve bu temelde tavsiyelerde bulunma tekniklerini kapsar. İşbirlikçi filtreleme, tavsiye eden sistemler için kullanılan en popüler yaklaşımlardır, ancak derecelendirme kaydının bulunmadığı tam soğuk başlatma sorunundan ve yalnızca az sayıda derecelendirme kaydının bulunduğu eksik soğuk başlatma sorunundan mustarıptır.

Sürü zekâsı, kendi kendine organize olabilen etkileşimli ajanların veya sürülerin popülasyonunu modelleyen bir araştırma dalıdır. Bir karınca kolonisi, bir kuş sürüsü veya bir bağışıklık sistemi, bir sürü sisteminin tipik bir örneğidir. Arıların kovanlarının etrafında toplanması, sürü zekâsının bir başka örneğidir. Yapay Arı Kolonisi (ABC) Algoritması, bal arısı sürüsünün akıllı davranışına dayalı bir

optimizasyon algoritmasıdır. Kombinasyonel ve sayısal optimizasyon problemlerini çözmek için çeşitli modern sezgisel algoritmalar geliştirilmiştir.

Bu makale farklı bölümlere ayrılmıştır. Temel olarak 5 bölümden oluşmaktadır. Bunlar: giriş; öneri sistemleri (literatür araştırması); tavsiye sistemleri yöntemleri (kullanıcı temelli tavsiye sistemleri, öge temelli tavsiye sistemleri, soğuk başlangıç problemi); çalışmanın uygulama modeli ve çözüm önerileri (model çalışması); sonuç bölümüdür.

Birinci bölüm giriş bölümüdür. Giriş bölümünde çalışılan konu olan öneri sisteminin kısaca tanıtımı yapıldıktan sonra işbirlikçi filtreleme, kullanıcı kullanıcı işbirlikçi filtreleme, öge öge işbirlikçi filtreleme, soğuk başlangıç problemi, üzerinde çalışılacak problem, bu problemin çözümü için üretilen model, tezin ana başlıkları ve içeriği hakkında bilgi verilmektedir.

İkinci bölümde genel olarak literatür çalışması yapılmıştır. Çalışmanın en üst başlığı öneri sistemlerinden, öneri sistemlerinin işlevinden, veri bilgi kaynaklarından, bu önerilerde bulunabilmek için geliştirilen tekniklerden, üzerinde çalışılan işbirlikçi filtrelemeden ve çeşitlerinden, bu alanda kullanılan algoritmalarından, öneri sistemlerinin sonuçlarının değerlendirilmesinin nasıl yapıldığından, piyasada kullanılan öneri sistemleri uygulamalarından, öneri sistemleri ve insan bilgisayar etkileşiminden ve karşılaşılan zorluklardan bahsedilmektedir. Bu bölüm işbirlikçi filtreleme gibi öneri sistemi oluşturmak için günümüzde en popüler olarak kullanılan teknikleri sunar; içerik tabanlı, bağlama duyarlı yöntemler, kullanıcı temelli tavsiye sistemleri, öge temelli tavsiye sistemleri. Ayrıca bir literatür çalışması olarak yeni bilginin edinilmesine rehberlik etmek için aktif öğrenme ilkelerinin kullanılması; önerilerdeki yenilik ve çeşitlilik; bir tavsiye sistemini kötü niyetli kullanıcıların saldırılarına karşı korumak için uygun teknikler ve daha güvenilir öneriler oluşturmak için birden çok türde kullanıcı geri bildirim ve tercihlerini bir araya getiren öneri sistemlerini de anlatıyoruz.

Üçüncü bölüm, uygulamamızda kullanılan teknikleri, yöntemleri ve yaklaşımları anlatmaktadır. Bu bölüm ayrıca ÖS tasarımını etkileyebilecek hususları da dikkate alır. Geliştirilen sistemlerin kullanıcı deneyleri ile değerlendirilmesinde

klasik pearson yaklaşımını, K ortalama küme algoritması, K en yakın komşuluk algoritması, genetik algoritma, arı algoritması, karınca koloni algoritması ve soğuk başlangıç problemini, zorlukları ve önlemlerini anlatır.

Dördüncü bölüm (model çalışması), tavsiyelerin nasıl sunulduğu, gözden geçirildiği, açıklandığı ve görselleştirildiği ile ilgili bir dizi konuyu ele alan belgeleri içerir. Bunlar arasında, bu bölüm kullanıcının bir tavsiye sistemi tarafından desteklenen karar verme sürecine odaklanmaktadır. Buna ek olarak tavsiye eden sistemlerin doğruluğunun ölçüm metrikleri de anlatılmıştır. Bu teknikler müzik, kitap, mobil bilgisayar, sosyal ağlar, eğitim ve filmler gibi birbirinden çok farklı alan ve geniş bir yelpazede kullanılabilir.

Soğuk başlatma önerisi, yeni kullanıcılar veya ögeler için seyrek kullanıcı-öge etkileşimleri nedeniyle zorlu bir sorun olmuştur. Mevcut çabalar, çoğu soruna veri düzeyinde yaklaşan soğuk başlangıç sorununu bir dereceye kadar hafifletmiştir. Daha önceki yöntemler genellikle yardımcı verileri kullanıcı veya öge özellikleri olarak içerirken, daha yeni yöntemler, daha yüksek dereceli grafik yapıları aracılığıyla daha zengin anlambilim yakalamak için heterojen bilgi ağlarından yararlanmaktadır. Biz burada pearson ile benzerlik teknikleri ve tahmin mekanizmalarıyla birlikte yaygın olarak bilinen sınıflandırma algoritmalarının yanı sıra Karınca Optimizasyon Algoritması (ACO) gibi farklı yaklaşımları ekleyerek bir model öneriyoruz. Yeni kullanıcılar olması durumunda, tavsiyelerde bulunmak için sistem tercihleri hakkında bilgiye sahip değildir. Önerilen yaklaşım, sınıflandırma yöntemlerini saf bir İF sisteminde birleştirirken, demografik verilerin kullanımı benzer davranışa sahip diğer kullanıcıların tanımlanmasına yardımcı olur. Karınca algoritması ile kümeleme çalışması yapıyoruz. Bu kümelerdeki kullanıcıların birbirine benzediğini kabul edip bunlar üzerinden tahminler yapıyoruz. Burada 3 farklı türde tahmin yapıyoruz. Bunlar klasik pearson hesabı ile tahmin, kullanıcı özelliklerini kullanarak yapılan tahmin ve kümeleme algoritmaları ile gruplayarak yapılan tahminler.

Son bölümde çalışmanın sonuçlarından bahsedilmiştir. Sonuçlar, önerilen sistemin doğruluğu artırma konusunda işe yaradığı göstermektedir. GroupLens araştırma grubu tarafından sağlanan yaygın olarak bilinen veri kümesi kullanılmıştır.

Farklı deneysel senaryolarda tatmin edici sayısal sonuçlar sağlayarak önerilen çözümün avantajları ortaya koyuyoruz.

1. LİTERATÜR ARAŞTIRMASI

1.1. ÖNERİ SİSTEMLERİ

Öneri Sistemleri (ÖS), belirli bir kullanıcının ilgisini en çok çekebilecek öğeler için öneriler sağlayan yazılım araçları ve tekniklerdir[1]. Öneriler, hangi ürünlerin satın alınacağı, hangi müziğin dinlenmesi gerektiği veya hangi çevrimiçi haberlerin okunacağı gibi çeşitli karar verme süreçleriyle ilgilidir. Tavsiye sistemlerinin geliştirilmesi, veri madenciliği, istatistik, karar destek sistemleri, yapay zekâ, insan bilgisayar etkileşimi, pazarlama ve tüketici davranışı gibi çeşitli alanlardan uzmanları içeren çok disiplinli bir çalışmadır. Öneri sistemleri, çevrimiçi kullanıcıların aşırı bilgi yüklemesiyle başa çıkmalarına ve daha iyi seçimler yapmalarına yardımcı olan değerli araçlardır. Artık netteki en güçlü ve popüler bilgi keşif araçlarından biridir. Öneri oluşturma için çeşitli teknikler geliştirildi ve son 20 yıl içinde bunların çoğu ticari ortamlarda da başarıyla uygulandı.

"Öge", sistemin kullanıcılara ne önerdiğini belirtmek için kullanılan genel terimdir. Bir ÖS normalde belirli bir öge türüne odaklanır ve buna göre tasarımı, grafik kullanıcı arabirimi ve önerileri oluşturmak için kullanılan temel öneri tekniğinin tümü, o belirli öge için yararlı ve etkili öneriler sağlamak üzere özelleştirilir[1].

Öneri sistemleri öncelikle, örneğin bir web sitesinin sunabileceği potansiyel olarak çok fazla sayıda alternatif ögeyi değerlendirmek için yeterli kişisel deneyim veya yeterliliğe sahip olmayan kişilere yöneliktir. En iyi örnek, kullanıcılara okumak için bir kitap seçmede yardımcı olan bir kitap tavsiye sistemidir. Popüler web sitesi Amazon.com'da site, her müşteri için çevrimiçi mağazayı kişiselleştirmek için bir ÖS kullanır. Öneriler genellikle kişiselleştirildiğinden, farklı kullanıcılar veya kullanıcı grupları çeşitli, özel önerilerden yararlanır[3]. Bunun yanında kişiselleştirilmemiş öneriler de vardır. Bunlar çok daha basit üretilir ve normalde dergi veya gazetelerde yer alır. Tipik örnekler arasında ilk on kitap seçimi, CD'ler vb yer alır. Belirli

durumlarda yararlı ve etkili olsalar da, bu tür kişiselleştirilmemiş öneriler tipik olarak ÖS araştırmasında ele alınmaz.

En basit haliyle, kişiselleştirilmiş öneriler, öğelerin sıralı listeleri olarak sunulur. Bu sıralamayı gerçekleştirirken, ÖS kullanıcının tercihlerine ve kısıtlamalarına göre en uygun ürün veya hizmetlerin ne olduğunu tahmin etmeye çalışır. Böyle bir hesaplama görevini tamamlamak için, ÖS kullanıcılarından bilgi toplar. Bunlar ürünler için derecelendirme olarak açıkça ifade edilen veya kullanıcının eylemlerinin yorumlanmasıyla elde edilen tercihleridir. Örneğin, ÖS belirli bir ürün sayfasına gitmeyi o sayfada gösterilen öğeler için örtük bir tercih işareti olarak kabul edebilir[3].

ÖS gelişimi oldukça basit bir gözlemden başlamıştır: bireyler genellikle rutin, günlük kararlar alırken başkaları tarafından sağlanan tavsiyelere güvenirlir. Örneğin, okumak için bir kitap seçerken akranlarının ne tavsiye ettiğine güvenmek yaygındır; işverenler işe alma kararlarında tavsiye mektuplarına güvenirlir ve izlemek için bir film seçerken, bireyler bir film eleştirmeninin yazdığı, okudukları gazetede yer alan film eleştirilerini okuma ve bunlara güvenme eğilimindedir[3].

Bu davranışı taklit etmeye çalışırken, ilk ÖS bir kullanıcı topluluğu tarafından üretilen önerilerden yararlanmak ve bu önerileri "aktif" bir kullanıcıya veya öneri arayan bir kullanıcıya sunmak için algoritmalar uyguladı. Öneriler benzer kullanıcıların veya benzer zevklere sahip kişilerin hoşuna giden öğeler içindi. Bu yaklaşım, iş birlikçi filtreleme olarak adlandırılır ve mantığı, eğer aktif kullanıcı geçmişte belirli kullanıcılarla mutabık kalmışsa, bu benzer kullanıcılardan gelen diğer tavsiyelerin, aktif kullanıcının ilgisini çekmesinin yanı sıra alakalı olması gerektiği şeklindedir[4].

E-ticaret web siteleri gelişmeye başladıkça, mevcut tüm alternatifleri filtrelemekten türetilen öneriler sağlamak için acil bir ihtiyaç ortaya çıktı. Kullanıcılar, bu web sitelerinin sunduğu çok çeşitli öğelerden en uygun seçimlere ulaşmakta zorlandılar.

İnternette mevcut olan hızlı büyüme ve bilgi çeşitliliği ve ürün satışı, ürün karşılaştırması, müzayedeler gibi yeni e-ticaret hizmetlerinin hızlı tanıtımı,

kullanıcıları sık sık bunaltmış ve kötü kararlar almalarına yol açmıştır. Bir fayda üretmek yerine seçeneklerin mevcudiyeti, kullanıcıların refahını azaltmaya başladı. Seçim iyi olsa da daha fazla seçeneğin her zaman daha iyi olmadığı anlaşıldı. Aslında, özgürlük, özerklik ve kendi kaderini tayin etme imalarıyla birlikte seçim aşırı hale gelebilir ve nihayetinde özgürlüğün bir tür sefalet uyandıran tiranlık olarak görülebileceği hissini oluşturabilir[4].

Son yıllarda, ÖS aşırı bilgi yüklemesi sorunuyla başa çıkmanın değerli bir yolu olduğu kanıtlanmıştır. Nihayetinde bir ÖS, kullanıcıyı yeni, henüz deneyimlenmemiş, kullanıcının mevcut göreviyle ilgili olabilecek öğelere yönlendirerek bu fenomeni ele alır. Bir kullanıcının talebi üzerine ÖS, kullanıcı bağlamına ve ihtiyacına göre öneri yaklaşımına bağlı olarak ifade edilebilen kullanıcılar, mevcut öğeler ve özelleştirilmiş veri tabanlarında depolanan önceki işlemler hakkında çeşitli bilgi ve verileri kullanarak öneriler üretir[3].

Kullanıcı daha sonra önerilere göz atabilir. Bunları kabul edebilir veya etmeyebilir ve hemen veya daha sonraki bir aşamada örtük veya açık geri bildirim sağlayabilir. Bu kullanıcı eylemi ve geri bildirim, öneri veri tabanında saklanabilir ve gelecek kullanıcı-sistem etkileşimlerinde yeni öneriler oluşturmak için kullanılabilir.

Daha önce belirtildiği gibi, tavsiye sistemleri çalışması, diğer klasik bilgi sistemi araçları ve tekniklerindeki araştırmaya kıyasla nispeten yenidir. Öneri sistemleri, 1990'ların ortalarında bağımsız bir araştırma alanı olarak ortaya çıktı[3]. Son yıllarda, aşağıdaki gerçeklerin gösterdiği gibi, tavsiye sistemlerine olan ilgi çarpıcı biçimde artmıştır.

Öneri sistemleri, Amazon.com, YouTube, Netflix, Spotify, LinkedIn, Facebook, Tripadvisor, Last.fm ve IMDB gibi yüksek puan alan web sitelerinde önemli bir rol oynar. Dahası, birçok medya şirketi şu anda abonelerine sağladıkları hizmetlerin bir parçası olarak işletim sistemi geliştiriyor ve dağıtıyor. Örneğin, isteğe bağlı akış medyasının çevrimiçi sağlayıcısı olan Netflix, öneri sisteminin performansını önemli ölçüde artırmayı ilk başaran takıma bir milyon dolarlık ödül verdi[5].

Bu alana özel olarak ayrılmış konferanslar ve atölyeler yapılmaktadır. Örneğin 2007'de kurulan Computing Machinery's (ACM) Öneri Sistemleri Üzerine Konferans Serisi bunlardan bir tanesidir. Bu konferans, öneri teknolojisi araştırma ve uygulamalarında en önemli yıllık etkinliktir. Ek olarak, öneri sistemlerine ayrılan oturumlar sıklıkla veri tabanları, bilgi sistemleri ve uyarlanabilir sistemler alanındaki daha geleneksel konferanslara dâhil edilir. Bu kapsamdaki diğer kayda değer saymamız gereken konferanslar: ACM'nin Bilgi Edinme Özel İlgi Grubu (SIGIR); Kullanıcı Modelleme, Uyarlama ve Kişiselleştirme (UMAP); Akıllı Kullanıcı Ara yüzleri (IUI); Dünya Çapında Ağ (WWW); ve ACM'nin Veri Yönetimi Özel İlgi Grubu[6].

Dünyanın dört bir yanındaki yükseköğretim kurumlarında, lisans ve lisansüstü kursları ciddi miktardaki kaynakları öneri sistemlerine ayırmıştır, Öneri sistemleri hakkındaki eğitimler bilgisayar bilimi konferanslarında çok popülerdir ve ÖS tekniklerini tanıtan kitaplarda yayınlanmıştır. Springer, kendi serisinde öneri sistemlerindeki belirli konularda birkaç kitap yayınlamaktadır: Springer Briefs in Electrical and Computer Engineering. Yazılım mühendisliğine yönelik sistem uygulamalarını tavsiye etmeye adanmış büyük, yeni bir makale koleksiyonu da yayınlanmıştır[6].

Akademik dergilerde öneri sistemleri alanındaki araştırma ve gelişmeleri kapsayan birkaç özel konu vardır. ÖS için belirli konuları olan dergilerden bazıları şunlardır: Yapay Zekâ İletişimi (AI Communications IEEE Akıllı Sistemler (IEEE Intelligent Systems); Uluslararası Elektronik Ticaret Dergisi (International Journal of Electronic Commerce); Uluslararası Bilgisayar Bilimi ve Uygulamaları Dergisi (International Journal of Computer Science and Applications); Bilgisayar İnsan Etkileşiminde ACM İşlemleri (ACM Transactions on Computer Human Interaction); Bilgi Sistemlerinde ACM İşlemleri (ACM Transactions on Information Systems); Kullanıcı Modelleme ve Kullanıcıya Uyarlanmış Etkileşim (User Modeling and User-Adapted Interaction); Etkileşimli Akıllı Sistemlerde ACM İşlemleri (ACM Transactions on Interactive Intelligent Systems); ve Akıllı Sistemler ve Teknolojiyle İlgili ACM İşlemleri (ACM Transactions on Intelligent Systems and Technology) [6].

Bu bölümde, ÖS fikirlerini ve kavramlarını kısaca tartışacağız. Bu bölümün sonunda, ÖS'nin geleceği için özellikle önemli olduğuna inandığımız bazı araştırma zorluklarını belirledik.

1.2. ÖNERİ SİSTEMLERİNİN İŞLEVİ

Öneri sistemleri, kullanıcılara bir kullanıcının kullanmak isteyebileceği öğeler için öneriler sağlayan yazılım araçları ve teknikleri olarak tanımlamıştık. Şimdi, bir öneri sisteminin oynayabileceği bir dizi olası rolü göstermek için bu tanımlı iyileştirmek istiyoruz. İlk olarak, öneri sisteminin hizmet sağlayıcı adına oynadığı rol ile öneri sistemleri kullanıcısının rolünü ayırmalıyız.

Örneğin, bir seyahat tavsiye sistemi tipik olarak ETS Tur gibi bir seyahat aracısı veya tripadvisor.com gibi bir varış yeri yönetimi kuruluşu tarafından cirosunu artırmak veya ETS Tur durumunda daha fazla otel odası satmak için uygulamaya konulur ve destinasyon yönetimi organizasyonu olması durumunda destinasyona gelen turist sayısını artırmak ister.

Kullanıcının bu iki sisteme erişimdeki birincil motivasyonu, bir varış noktasını ziyaret ederken uygun bir otel ve ilginç etkinlikler veya ilgi çekici yerler bulmak olacaktır. Aslında, hizmet sağlayıcıların bu teknolojiden yararlanmak isteyebilmelerinin çeşitli nedenleri vardır. ÖS'nin yardımcı olacağı alanların bazıları şunlardır: satılan öğelerin sayısını artırma, daha çeşitli ürünler satabilme, kullanıcı memnuniyetini artırma, kullanıcı sadakatini artırma, yardımcı olunabilecek yeni görevler tanımlama, bazı iyi öğeleri bulma, tüm iyi öğeleri bulma, bağlam içinde yorum yapma, bir sıra önerme, bir paket önerme, sadece göz atanlar için çalışma yapma, güvenilir bir öneri bulma, profili iyileştirme, kullanıcın kendisini ifade etmesini sağlama, başkalarına yardım etmeye teşvik etme, kullanıcıların başkalarını etkilemesine ortam hazırlama[7]. Bunları biraz örnekler ile açıklayalım.

Satılan öğelerin sayısını artırın: Bu, muhtemelen ticari bir ÖS için en önemli işlevdir, yani, genellikle herhangi bir öneri olmadan satılanlara kıyasla ek bir ürün seti satabilmek. Bu hedefe ulaşılır çünkü önerilen öğeler muhtemelen kullanıcının ihtiyaç ve isteklerine uygun olur. Muhtemelen kullanıcı, birkaç öneriyi denedikten sonra bunu anlayacaktır. Ticari olmayan uygulamaların, kullanıcı için bir öğenin seçilmesiyle

ilişkilendirilmiş bir maliyeti olmasa bile benzer hedefleri vardır. Örneğin, bir içerik ağı, sitesinde okunan haberlerin sayısını artırmayı amaçlamaktadır. Genel olarak, servis sağlayıcının bakış açısından, ÖS tanıtmanın birincil amacının dönüşüm oranını artırmak olduğunu söyleyebiliriz, yani, sadece bilgilere göz atan basit ziyaretçilerin sayısına kıyasla, tavsiyeyi kabul eden ve bir ögeyi tüketen kullanıcıların sayısını arttırın[7].

Daha çeşitli ürünler satın: Bir ÖS'nin diğer bir ana işlevi, kullanıcının kesin bir öneri olmadan bulunması zor olabilecek öğeleri seçmesini sağlamaktır. Örneğin, bir turist ÖS'de hizmet sağlayıcı, sadece en popüler olanları değil, bir turistik bölgedeki tüm ilgi çekici yerleri tanıtmakla ilgilenir. Bu, ÖS olmadan zor olabilir çünkü hizmet sağlayıcı, belirli bir kullanıcının zevkine uymayan yerlerin reklamını yapma riskini karşılayamaz. Bu nedenle, ÖS popüler olmayan yerleri doğru kullanıcılara önerir veya reklamını yapar[3].

Kullanıcı memnuniyetini artırın: İyi tasarlanmış bir ÖS, kullanıcının site veya uygulama ile olan deneyimini de geliştirebilir. Kullanıcı, önerileri ilginç ve ilgili bulacak ve uygun şekilde tasarlanmış bir insan bilgisayar etkileşimi ile sistemi kullanmaktan da keyif alacaktır. Etkili, doğru öneriler ve kullanılabilir bir ara yüzün kombinasyonu, kullanıcının sisteme ilişkin öznel değerlendirmesini artıracaktır. Bu da sistem kullanımını ve önerilerin kabul edilme olasılığını artıracaktır[4].

Kullanıcı sadakatini artırın: Bir kullanıcı, ziyaret edildiğinde eski müşteriyi tanıyan ve ona değerli bir ziyaretçi olarak davranan bir web sitesine sadık olmalıdır. Bu, ÖS'nin standart bir özelliğidir, çünkü birçok ÖS tavsiyeleri hesaplar ve böylece kullanıcının öge derecelendirmeleri gibi önceki etkileşimler sırasında kullanıcıdan elde edilen bilgilerden yararlanır. Sonuç olarak, kullanıcı siteyle ne kadar uzun süre etkileşimde bulunursa, kullanıcının modeli o kadar rafine hale gelir: sistemin kullanıcının tercihlerinin temsili gelişir ve öneri çıktısının kullanıcının tercihlerini özelleştirip eşleştirme etkinliği artar[8].

E-hizmet sağlayıcılarının neden ÖS'leri tanıttığına dair bazı önemli motivasyonlardan yukarıda bahsetmiştik. Ancak kullanıcılar, görevlerini veya hedeflerini etkili bir şekilde destekleyecekse bir ÖS isteyebilir. Sonuç olarak, bir ÖS

bu iki oyuncunun ihtiyalarını dengelemeli ve her ikisi iin de deęerli bir hizmet sunmalıdır.

ÖS'nin uygulanmasında yardımcı olabileceęi görevler tanımlayın: Bazıları, bir kullanıcı iin yararlı olabilecek ögeler iin öneriler sunmak gibi normalde bir ÖS ile ilişkilendirilen ana veya temel görevler olarak kabul edilebilen görevler tanımlayın. Diğerleri, bir ÖS'yi tüketmenin daha fırsatçı yolları olarak düşünülebilir. Nitekim bu görev farklılaşması bir arama motorunda olana çok benzer. Birincil işlevi, kullanıcının bilgi ihtiyacıyla ilgili belgeleri bulmaktır, ancak bir web sayfasının önemini kontrol etmek iin bir sorgunun sonuç listesindeki sayfanın konumuna bakarak veya çeşitli bilgileri keşfetmek iin de kullanılabilir. Bir belge koleksiyonundaki bir kelimenin tüm kullanımlarına bakarak yapar bunu[9].

Bazı İyi Ögeleri Bulun: Bir kullanıcıya bazı ögeleri sıralı liste olarak önerin ve kullanıcının onlardan ne kadar hoşlanacağına dair tahminler yapın. Bu, birçok ticari sistemin ele aldığı ana öneri görevidir. Bazı sistemler tahmin edilen derecelendirmeyi göstermez.

Tüm iyi ögeleri bulun: Bazı kullanıcı ihtiyalarını karşılayabilecek tüm ögeleri önerin. Bu gibi durumlarda sadece bazı iyi ögeler bulmak yetersizdir. Bu, özellikle ögelerin sayısı nispeten az olduğunda veya ÖS tıbbi veya finansal uygulamalar gibi görev açısından kritik olduğunda geçerlidir. Bu durumlarda, tüm olasılıkların dikkatlice incelenmesinden elde edilen faydaya ek olarak kullanıcı, bu ögelerin ÖS sıralamasından veya ÖS'nin oluşturduğu ek açıklamalardan da yararlanabilir[9].

Baęlam iinde yorum yapın: Mevcut bir baęlam verildiğinde, örneğin bir öge listesi verildiğinde, kullanıcının uzun vadeli tercihlerine baęlı olarak bazılarını vurgulayın. Örneğin, bir TV tavsiye sistemi, elektronik program rehberinde görüntülenen hangi TV şovlarının izlemeye deęer olduğuna açıklama ekleyebilir.

Bir sıra önerin: Tek bir önerinin oluşturulmasına odaklanmak yerine, fikir bir bütün olarak hoşla giden bir dizi öge önermektir. Tipik örnekler arasında bir TV dizisi önermek, veri madencilięi üzerine bir kitap tavsiye ettikten sonra ÖS üzerine bir kitap veya müzik paralarının bir derlemesi yer alabilir.

Bir paket önerin: Birbirine iyi uyan bir grup öge önerin. Örneğin, bir seyahat planı, sınırlandırılmış bir alanda bulunan çeşitli turistik yerler, destinasyonlar ve konaklama hizmetlerinden oluşabilir. Kullanıcı açısından bu çeşitli alternatifler tek bir seyahat destinasyonu olarak düşünülebilir ve seçilebilir[9].

Sadece göz atanlar için çalışma yapın: Bu görevde, kullanıcı herhangi bir ürün satın alma niyetinde olmadan kataloga göz atar. Önericinin görevi, kullanıcının söz konusu tarama oturumu için kullanıcının ilgi alanlarının kapsamına girme olasılığı daha yüksek olan öğelere göz atmasına yardımcı olmaktır. Bu, uyarlanabilir büyük ortam teknikleriyle de desteklenen bir görevdir.

Güvenilir bir öneri bulun: Bazı kullanıcılar öneri sistemlerine güvenmez, bu nedenle tavsiyelerde ne kadar iyi olduklarını görmek için onlarla oynarlar. Bu nedenle, belirli bir sistem, sadece tavsiye almak için gerekli olanlara ek olarak, kullanıcıların davranışını test etmesine izin veren belirli işlevler de sunabilir.

Profili iyileştirin: Bu, kullanıcının neyi sevip neyi sevmediği hakkında tavsiye sistemine bilgi sağlama yeteneği ile ilgilidir. Bu, kişiselleştirilmiş öneriler sağlamak için kesinlikle gerekli olan temel bir görevdir. Sistemin aktif kullanıcı hakkında belirli bir bilgisi yoksa o zaman sadece "ortalama" bir kullanıcıya verilecek tavsiyelerin aynısını sağlayabilir.

Kendini ifade et: Bazı kullanıcılar önerilerle hiç ilgilenmeyebilir. Aksine, onlar için önemli olan, derecelendirmeleriyle katkıda bulunmalarına ve fikir ve inançlarını ifade etmelerine izin verilmesidir. Bu etkinlik için kullanıcı memnuniyeti, yine de bir kaldıraç görevi görebilir ve bu da kullanıcının uygulamaya olan bağlılığının devam etmesiyle sonuçlanır[9].

Başkalarına yardım edin: Bazı kullanıcılar bilgi ile katkıda bulunmaktan mutlu olurlar, mesela öğelerin değerlendirilmesinde. Çünkü topluluğun katkılarından yararlandığına inanırlar. Bu, rutin olarak kullanılmayan bir tavsiye sistemine bilgi girmek için önemli bir motivasyon olabilir. Örneğin, bir otomobil ÖS'nde, hâlihazırda yeni bir araba satın almış olan bir kullanıcı, bir dahaki sefere başka biri yeni bir araba satın alacağından sisteme gireceği derecelendirmenin kendisi yerine diğer kullanıcılar için yararlı olabileceğinin farkındadır.

Başkalarını etkileyin: Web tabanlı ÖS'lerde, asıl amacı diğer kullanıcıları belirli ürünleri satın almaları için açıkça etkilemek olan kullanıcılar vardır. Nitekim sistemi yalnızca belirli öğeleri teşvik etmek veya cezalandırmak için kullanabilen kötü niyetli kullanıcılar da vardır.

Bu çeşitli noktaların gösterdiği gibi, ÖS'nin bir bilgi sistemindeki rolü oldukça çeşitli olabilir. Bu çeşitlilik, bir dizi farklı bilgi kaynağı ve tekniğin kullanılmasını gerektirir.

1.3. VERİ VE BİLGİ KAYNAKLARI

ÖS'ler, tavsiyelerini oluşturmak için çeşitli türde verileri aktif olarak toplayan bilgi işleme sistemleridir. Veriler, öncelikle önerilecek öğeler ve bu önerileri alacak kullanıcılar hakkındadır. Ancak, tavsiye eden sistemler için mevcut olan veri ve bilgi kaynakları çok çeşitli olabileceğinden, nihayetinde kullanılıp kullanılmayacağı tavsiye tekniğine bağlıdır[10].

Genel olarak, bilgi eksikliği olan, yani çok basit ve temel verileri kullanan, örneğin kullanıcı derecelendirmeleri veya maddeler için değerlendirmeler gibi tavsiye teknikleri vardır. Diğer teknikler, kullanıcıların ontolojik tanımlarını veya öğelerin kısıtlamaları veya kullanıcıların sosyal ilişkilerini ve etkinliklerini kullandıkları için çok daha fazla bilgiye bağlıdır. Her durumda, genel bir sınıflandırma olarak, ÖS'ler tarafından kullanılan veriler üç tür nesneye atıfta bulunur: öğeler, kullanıcılar ve işlemler, yani kullanıcılar ve öğeler arasındaki ilişkiler[9].

Öğeler tavsiye edilen nesnelere dir. Öğeler, karmaşıklıkları ve değerleri veya faydaları ile karakterize edilebilir. Bir öğenin değeri, öge kullanıcı için yararlıysa olumlu, öge uygun değilse ve kullanıcı onu seçerken yanlış karar vermişse olumsuz olabilir. Bir kullanıcı bir öğeyi alırken her zaman, öğeyi aramanın bilişsel maliyetini ve sonunda öge için ödenen gerçek parasal maliyeti içeren bir maliyete maruz kalır.

Örneğin, bir haber ÖS'nin tasarımcısı bir haber ögesinin karmaşıklığını, yani yapısını, metinsel gösterimini ve herhangi bir haber ögesinin zamana bağlı önemini hesaba katmalıdır. Ancak aynı zamanda, ÖS tasarımcısı, kullanıcı haberleri okumak için ödeme yapmasa bile, haber öğelerini aramak ve okumakla ilgili her zaman bilişsel

bir maliyet olduğunu anlamalıdır. Seçilen bir öge kullanıcıyla ilgiliyse, bu maliyete yararlı bilgiler edinmiş olmanın yararı hâkimdir. Oysa öge alakalı değilse, o kalemin kullanıcı için net değeri ve tavsiyesi negatiftir. Arabalar veya finansal yatırımlar gibi diğer alanlarda, ögelerin gerçek parasal maliyeti, en uygun tavsiye yaklaşımını seçerken dikkate alınması gereken önemli bir unsur haline gelir[11].

Haberler, web sayfaları, kitaplar, CD'ler ve filmler düşük karmaşıklığa ve değere sahip ögelerdir. Daha büyük karmaşıklığa ve değere sahip ögeler şunlardır: dijital kameralar, cep telefonları, PC'ler vb. göz önünde bulundurulmuş en karmaşık ögeler sigorta poliçeleri, finansal yatırımlar, seyahat ve işlerdir.

ÖS'ler, çekirdek teknolojilerine göre, ögelerin bir dizi özelliğini kullanabilir. Örneğin, bir film tavsiye sisteminde, tür, oyuncular, yönetmen, bir filmi tanımlamak ve bir ögenin faydasının özelliklerine nasıl bağlı olduğunu öğrenmek için kullanılabilir. Ögeler, çeşitli bilgi ve temsil yaklaşımları kullanılarak, örneğin en basit şekilde tek bir kimlik kodu olarak veya daha zengin bir biçimde, bir özellikler kümesi olarak ve hatta alanın ontolojik bir temsilinde bir kavram olarak temsil edilebilir[12].

Kullanıcılar, yukarıda belirtildiği gibi ÖS kullanıcıları çok çeşitli hedeflere ve özelliklere sahip olabilir. Önerileri ve insan-bilgisayar etkileşimini kişiselleştirmek için, ÖS'ler kullanıcılar hakkında bir dizi bilgiyi kullanır. Bu bilgiler çeşitli şekillerde yapılandırılabilir ve yine hangi bilgilerin modelleneyeceğinin seçimi öneri tekniğine bağlıdır.

Örneğin, işbirliğine dayalı filtrelemede kullanıcılar, belirli ögeler için kullanıcı tarafından sağlanan derecelendirmeleri içeren basit bir liste olarak modellenir. Demografik bir ÖS'de yaş, cinsiyet, meslek ve eğitim gibi sosyo-demografik özellikler kullanılır. Kullanıcı verilerinin kullanıcı modelini oluşturduğu söylenir. Kullanıcı modeli, kullanıcının profilini çıkarır, yani tercihlerini ve ihtiyaçlarını kodlar. Çeşitli kullanıcı modelleme yaklaşımları vardır ve bir anlamda bir ÖS, bunları sağlayan bir araç olarak görülebilir. Kullanıcı modellerini oluşturarak ve kullanarak öneriler üretir. Uygun bir kullanıcı modeli olmadan kişiselleştirme mümkün olmadığından, kullanıcı modeli her zaman merkezi bir rol oynayacaktır. Örneğin, işbirlikçi bir filtreleme yaklaşımını yeniden değerlendirirken, kullanıcı ya doğrudan öge derecelendirmelerine

göre profillenir ya da bu derecelendirmeleri kullanarak, sistem, kullanıcıların modellerinde her bir faktörün nasıl ağırlıklandığına göre farklılık gösterdiği bir faktör değerleri vektörü türetir[13].

Kullanıcılar; ayrıca davranış örüntü verileriyle de tanımlanabilir, örneğin, site tarama modelleri veya seyahat arama modelleri. Ayrıca kullanıcı verileri, kullanıcılar arasındaki bu ilişkilerin güven düzeyi gibi kullanıcılar arasındaki ilişkileri de içerebilir. Bir ÖS, benzer veya güvenilir kullanıcılar tarafından tercih edilen öğeleri kullanıcılara önermek için bu bilgileri kullanabilir.

İşlemler; genel olarak bir işlemi, bir kullanıcı ile ÖS arasında kaydedilmiş bir etkileşim olarak adlandırırız. İşlemler, insan-bilgisayar etkileşimi sırasında üretilen önemli bilgileri depolayan ve sistemin kullandığı öneri oluşturma algoritması için yararlı olan log benzeri verilerdir. Örneğin, bir işlem günlüğü, kullanıcı tarafından seçilen öğeye bir referans ve bu belirli tavsiye için bağlamın bir açıklamasını içerebilir. Varsa bu işlem seçilen öğe için derecelendirme gibi kullanıcının sağladığı açık geri bildirimleri de içerebilir[8].

Aslında derecelendirmeler, ÖS'nin topladığı en popüler işlem verisi biçimidir. Bu derecelendirmeler açıkça veya dolaylı olarak toplanabilir. Açıkça derecelendirmelerin toplanmasında, kullanıcıdan derecelendirme ölçeğindeki bir öğe hakkında fikir vermesi istenir. Buna göre, derecelendirmeler çeşitli şekillerde olabilir:

- Amazon.com ile ilişkili kitap önerisinde sağlanan 1-5 yıldız gibi sayısal derecelendirmeler[9].
- Kullanıcıdan bir öğe hakkındaki fikrini en iyi belirten terimi seçmesinin istendiği “kesinlikle katılıyorum, katılıyorum, tarafsız, katılmıyorum, kesinlikle katılmıyorum” gibi sıralı derecelendirmeler[9].
- Kullanıcıdan belirli bir ögenin iyi mi yoksa kötü mü olduğuna karar vermesinin istendiği seçimleri modelleyen ikili derecelendirmeler[9].
- Tekli derecelendirmeler, bir kullanıcının bir öğeyi gözlemlediğini veya satın aldığını veya başka bir şekilde öğeyi olumlu olarak derecelendirdiğini gösterebilir. Bu gibi durumlarda, bir

derecelendirmenin olmaması, kullanıcıyı ürünle ilgili hiçbir bilginin olmadığını gösterir[9].

Kullanıcı değerlendirmesinin başka bir biçimi, kullanıcı tarafından sistemin sunduğu öğelerle ilişkilendirilen etiketlerden oluşur. Örneğin, MovieLens'te ÖS etiketleri, MovieLens kullanıcılarının bir film hakkında ne düşündüklerini temsil eder, örneğin: "çok uzun" veya "korku" gibi.

Örtülü derecelendirme toplayan işlemlerde sistem, kullanıcının eylemlerine dayanarak kullanıcının fikrini çıkarmayı amaçlar. Örneğin, bir kullanıcı Amazon.com'da "Python" anahtar kelimesini girerse, uzun bir kitap listesi sağlanacaktır. Buna karşılık, kullanıcı ek bilgi almak için listedeki belirli bir kitaba tıklayabilir. Bu noktada, sistem, kullanıcının bir şekilde o kitapla ilgilendiği sonucuna varabilir[14].

Konuşma sistemlerinde, yani etkileşimli bir süreci destekleyen sistemlerde işlem modeli daha rafine edilir. Bu sistemlerde, kullanıcı talepleri sistem eylemleriyle dönüşümlü olarak değişir. Yani kullanıcı bir öneri isteyebilir ve sistem bir öneri listesi oluşturabilir. Ancak, kullanıcıya daha iyi, daha hassas sonuçlar sağlamak için ek kullanıcı tercihleri de talep edebilir. Burada, işlem modelinde, sistem çeşitli istek yanıtlarını toplar ve sonunda öneri sürecinin sonucunu gözlemleyerek etkileşim stratejisini değiştirmeyi öğrenebilir.

1.4. ÖNERİ TEKNİKLERİ

Kullanıcı için yararlı öğeleri belirleyen temel işlevini uygulamak için, ÖS bir öğenin tavsiye edilmeye değer olduğunu tahmin etmelidir. Bunu yapabilmek için, sistemin bazı öğelerin faydasını tahmin edebilmesi veya en azından bazı öğelerin faydasını karşılaştırması ve ardından bu karşılaştırmaya dayanarak hangi öğelerin önerileceğine karar vermesi gerekir. Tahmin adımı, öneri algoritmasında açık olmayabilir, ancak yine de bir ÖS'nin genel rolünü tanımlamak için bu birleştirici modeli uygulayabiliriz. Buradaki amacımız, okuyucuya gösterilecek olan tüm farklı öneri yaklaşımlarının bir açıklaması yerine birleştirici bir bakış açısı sağlamaktır.

Bir ÖS'nin tahmin adımını göstermek için, örneğin, yalnızca en popüler şarkıları öneren basit ve kişiselleştirilmemiş bir öneri algoritmasını düşünün. Bu yaklaşımı kullanmanın mantığı, kullanıcının tercihleri hakkında daha kesin bilgilerin yokluğunda, popüler bir şarkının, yani birçok kullanıcı tarafından sevilen bir şarkının da büyük olasılıkla genel bir kullanıcıya hitap edeceğidir veya en azından rastgele seçilen başka bir şarkıdan daha yüksek olasılıkla. Bu nedenle, bu tür popüler şarkıların faydasının bu genel kullanıcı için oldukça yüksek olacağı tahmin edilmektedir.

Yukarıda bahsedildiği gibi, bazı tavsiye sistemleri bir tavsiyede bulunmadan önce faydayı tam olarak tahmin etmez, ancak bir ögenin bir kullanıcı için yararlı olabileceğini varsaymak için bazı buluşsal yöntemler uygulayabilirler. Bu, örneğin bilgiye dayalı sistemlerde tipiktir. Bu yardımcı program tahminleri belirli algoritmalarla hesaplanır ve kullanıcılar, öğeler ve yardımcı program işlevinin kendisi hakkında çeşitli bilgiler kullanır. Örneğin, sistem, yardımcı program işlevinin evet – hayır gibi ikili değişken olduğunu varsayabilir ve bu nedenle, yalnızca bir ögenin kullanıcı için yararlı olup olmadığını belirleyecektir. Sonuç olarak, tavsiyeyi isteyen kullanıcı hakkında bir miktar bilgi olduğunu veya muhtemelen hiç olmadığını varsayarak, öğeler ve diğer kullanıcılar hakkındaki bilgilerin yanı sıra sistem, çeşitli yardımcı program tahminleri ve dolayısıyla öneriler oluşturmak için bu bilgiyi uygun bir algoritma ile kullanacaktır[8].

Bazen bir öge için kullanıcı yararının, genel olarak "bağlamsal" olarak adlandırdığımız diğer değişkenlere bağlı olduğu gözlemlendiğini de not etmek önemlidir. Örneğin, bir ögenin bir kullanıcı için faydası, kullanıcının alan bilgisinden etkilenebilir veya tavsiyenin istendiği zamana bağlı olabilir. Aynı şekilde, kullanıcılar mevcut konularına daha yakın öğelerle daha fazla ilgilenebilir. Sonuç olarak, tavsiyeler bu özel ek ayrıntılara uyarlanmalıdır ve sonuç olarak doğru tavsiyelerin ne olduğunu doğru bir şekilde tahmin etmek giderek daha zor hale gelmektedir[8].

ÖS adreslenen alan ve kullanılan bilgi açısından değişen, ancak özellikle tavsiye algoritmasına, yani bir tavsiyenin faydasının tahmininin nasıl yapıldığına göre değişen birkaç farklı türde tavsiye sistemi sunmaktadır. Diğer farklılıklar, önerilerin nihai olarak nasıl bir araya getirildiği ve kullanıcı isteklerine yanıt olarak kullanıcıya sunulduğu ile ilgilidir. Bu yönler de daha detaylı olarak tartışılması gereken konulardır.

Farklı ÖS türlerine ilk bir genel bakış sağlamak için, Burke tarafından sağlanan ve tavsiye eden sistemler arasında ayrım yapmanın ve bunlara atıfta bulunmanın klasik bir yolu haline gelen bir taksonomiden alıntı yapmak istiyoruz. Burke beş farklı öneri yaklaşımı sınıfını birbirinden ayırmaktadır[15]: İşbirlikçi Filtreleme (Collaborative Filtering, CF), İçerik Temelli Sistem (Content – Based, CN), Demografik Sistem (Demographic, DM), Bilgi Temelli Sistemler (Knowledge-Based, KB), Fayda Temelli Sistem (Utility-based, UT).

İF, Bu yaklaşımın orijinal ve en basit uygulaması, geçmişte benzer zevklere sahip diğer kullanıcıların sevdiği öğelere dayalı olarak aktif kullanıcıya tavsiyelerde bulunur. İki kullanıcının beğenilerindeki benzerlik, kullanıcıların derecelendirme geçmişindeki benzerliğe göre hesaplanır. İşbirlikçi filtrelemeden "insanlar arası korelasyon" olarak bahsedilmesinin nedeni budur. İşbirlikçi filtreleme, ÖS'de en popüler ve yaygın olarak uygulanan teknik olarak kabul edilir.

İşbirlikçi filtrelemede komşu - temelli yöntemler kapsamlı bir yer tutar. Komşu - temelli yöntemler, öğeler arasındaki veya alternatif olarak kullanıcılar arasındaki ilişkilere odaklanır. Bir öğe-öge yaklaşımı, aynı kullanıcı tarafından benzer öğelerin derecelendirmelerine dayalı olarak bir kullanıcının bir öğeye tercihini modeller. Komşu - temelli yöntemler, basitlik, verimlilik ve doğru ve kişiselleştirilmiş öneriler üretme yeteneği sunar. Bu tür yöntemlerin temel faydalarının yanı sıra temel özelliklerini de ilerde daha detaylı açıklayacağız.

Dahası, komşu - temelli bir tavsiye sistemi uygularken gerekli olan temel kararları ele alıp, bu tür kararların nasıl alınacağına dair pratik bilgiler vermemiz gerekir. Belki de tavsiye eden sistemin derecelendirme tahmini ve hesaplama performansı üzerinde en büyük etkiye sahip olan karar, kullanıcı tabanlı ve ürün tabanlı bir yöntem arasındaki seçimdir. Kullanıcı sayısının mevcut öğe sayısını aştığı tipik ticari tavsiye sistemlerinde, öğe temelli yaklaşımlar, daha doğru öneriler sağlarken, hesaplama açısından daha verimli oldukları ve daha az sıklıkta güncelleme gerektirdikleri için tercih edilmelidir. Öte yandan, kullanıcı tabanlı yöntemler genellikle daha orijinal öneriler sağlar ve bu da kullanıcıları daha tatmin edici bir deneyime yönlendirebilir.

Genellikle büyük ticari tavsiye sistemlerinde gözlemlenen seyreklik ve sınırlı kapsam sorunları, iki araştırma yönünü keşfederek tartışılmaktadır: boyutsallık azaltma ve grafik tabanlı teknikler. Boyut azaltma, kullanıcıların ve ögelerin en önemli özelliklerini yakalayan kompakt bir temsiliyi sağlar. Böyle bir yaklaşımın bir avantajı, bu kullanıcılar farklı ögeleri derecelendirmiş veya bu ögeler farklı kullanıcılar tarafından derecelendirilmiş olsa bile, kullanıcı çiftleri veya ögeler arasında anlamlı ilişkiler elde etmeye izin vermesidir. Öte yandan, grafik tabanlı teknikler verilerdeki geçişli ilişkilerden yararlanır. Bu teknikler ayrıca kullanıcılar veya doğrudan bağlantılı olmayan ögeler arasındaki ilişkiyi değerlendirerek seyreklik ve sınırlı kapsam sorunlarından da kaçınır. Bununla birlikte, boyut azaltmanın aksine, grafiğe dayalı yöntemler de verilerdeki bazı “yerel” ilişkileri korur.

İşbirlikçi filtreleme önerileri oluşturmak için mevcut olan birkaç yeni uzantı vardır. Özellikle, matris çarpanlara ayırma (ör. Tekil Değer Ayırıştırma (TDA)) gibi gizli faktör modelleri vardır. Bu yöntemler hem ögeleri hem de kullanıcıları aynı gizli faktör uzayına dönüştürür. Gizli alan daha sonra hem ürünleri hem de kullanıcıları kullanıcı geri bildirimlerinden otomatik olarak çıkarılan faktörler açısından karakterize ederek derecelendirmeleri açıklamak için kullanılır. TDA örtük geri bildirim ve geçici bilgiler dâhil olmak üzere verilerin ek özelliklerini ele alabilir. Ayrıca, küresel optimizasyon tekniklerini kullanarak daha titiz formüller önererek komşuluk tekniklerinin eksikliklerini gidermek için teknikleri açıklarlar. Bu tür tekniklerin kullanılması, komşu büyüklüğündeki sınırı kaldırmayı ve örtük geri bildirim ve zamansal dinamikleri ele almayı mümkün kılar. Elde edilen doğruluk, bir dizi pratik avantaj sunarken, matris çarpanlara ayırma modellerine yakındır.

İçerik Temelli Sistem, kullanıcının geçmişte sevdiğilerine benzer ögeler önermeyi öğrenir. Ögelerin benzerliği, karşılaştırılan ögelerle ilişkili özelliklere göre hesaplanır. Örneğin, bir kullanıcı korku türüne ait bir filmi olumlu olarak derecelendirdiyse, sistem bu türden başka filmler önermeyi öğrenebilir.

Klasik içerik tabanlı öneri teknikleri, kullanıcı profilinin özelliklerini ögelerin özellikleriyle eşleştirmeyi amaçlar. Çoğu durumda, ögelerin öznelikleri yalnızca ögelerin açıklamalarından çıkarılan anahtar kelimelerdir. Anlamsal indeksleme

teknikleri, anahtar kelimeler yerine kavramlar kullanan öge ve kullanıcı profillerini temsil eder.

Klasik anahtar kelime tabanlı öneri teknikleri sistemlerin temel sorunlarının üstesinden gelmek için indeksleme teknikleri kullanır. İki ana grup anlamsal indeksleme tekniği vardır: yukarıdan aşağıya ve aşağıdan yukarıya. Önceki gruptaki teknikler, ontolojiler, Wikipedia gibi ansiklopedik bilgi ve bağlı Veri bulutundan gelen veriler gibi harici bilgi kaynaklarının entegrasyonuna dayanırken, ikinci gruptaki teknikler hipoteze dayalı, kelimelerin anlamının büyük metinsel belgelerde kullanımına bağlı olduğu hafif bir anlamsal temsile dayanır. Yeni nesil anlamsal içerik tabanlı tavsiye sistemlerini gerçekleştirmek için anlamsal yaklaşımların nasıl kullanılacağını, ana potansiyellerinin ve sınırlamalarının bir tanımını sağlayarak daha detaylı çalışmalar yapılmaktadır.

Demografik Sistem, kullanıcının demografik profiline dayalı ögeler önerir. Varsayım, farklı demografik dokular için farklı tavsiyelerin üretilmesi gerektiğidir. Birçok web sitesi, demografiye dayalı basit ve etkili kişiselleştirme çözümlerini benimser. Örneğin, kullanıcılar dillerine veya ülkelerine göre belirli web sitelerine gönderilir. Veya öneriler kullanıcının yaşına göre özelleştirilebilir. Bu yaklaşımlar pazarlama literatüründe oldukça popüler olmakla birlikte, demografik sistemler üzerine daha fazla ÖS araştırması yapılabilir.

Bilgi Temelli Sistemler, belirli öge özelliklerinin kullanıcıların ihtiyaçlarını ve tercihlerini nasıl karşıladığı ve nihayetinde ögenin kullanıcı için nasıl yararlı olduğu hakkında belirli alan bilgisine dayalı ögeler önerir. Önemli bilgi temelli tavsiye sistemleri vaka temellidir. Bu sistemlerde, bir benzerlik işlevi, kullanıcının ihtiyaçlarının önerilerle ne kadar eşleştiğini tahmin eder. Burada benzerlik puanı, tavsiyenin kullanıcı için faydası olarak doğrudan yorumlanabilir. Bilgi tabanlı sistemler, dağıtımlarının başlangıcında diğerlerinden daha iyi çalışma eğilimindedir, ancak öğrenme bileşenleri ile donatılmamışlarsa, insan - bilgisayar etkileşiminin günlüklerinden yararlanabilen diğer sığ yöntemlerle aşılabirler.

Fayda Temelli Sistemler, tavsiye edilecek nesnelere için bir fayda değeri hesaplar ve ilke olarak, bu tür hesaplamalar fonksiyonel bilgiye dayalı olabilir.

Bununla birlikte, mevcut sistemler bu tür çıkarımları kullanmaz, kullanıcıların ihtiyaçları ile ürünlerin özellikleri arasında kendi eşleştirmelerini yapmalarını gerektirir.

Fayda Temelli ve bilgi temelli tavsiye verenler, kullanıcıları hakkında uzun vadeli genellemeler oluşturmaya çalışmazlar, bunun yerine tavsiyelerini, bir kullanıcının ihtiyacı ile mevcut seçenekler arasındaki eşleşmenin bir değerlendirmesine dayandırır. Fayda tabanlı tavsiye verenler, kullanıcı için her nesnenin faydasının hesaplanmasına dayalı olarak önerilerde bulunur[15].

Kısıtlama temelli sistemler başka bir bilgi tabanlı ÖS türüdür. Kullanılan bilgi açısından, her iki sistem de benzerdir: kullanıcı gereksinimleri toplanır, çözüm bulunamayan durumlarda otomatik olarak tutarsız gereksinimlere yönelik onarımlar önerilir ve öneri sonuçları açıklanır. En büyük fark, çözümlerin hesaplanma biçiminde yatmaktadır. Vaka temelli tavsiyeciler, benzerlik ölçütleri temelinde tavsiyeleri belirlerken, kısıtlama temelli tavsiyeciler ağırlıklı olarak müşteri gereksinimlerinin ürün özellikleriyle nasıl ilişkilendirileceğine dair açık kurallar içeren önceden tanımlanmış bilgi tabanlarından yararlanır[15].

Kısıtlama temelli öneri yaklaşımlarını gözden geçirmekte ve uygun araç desteği pratik ortamlarda çok önemli olabileceğinden, kısıtlama temelli tavsiye verenler için bilgi tabanlarının geliştirilmesine yönelik teknolojilere genel bir bakış sunmaktadır. Kısıtlama temelli yöntemlerin finansal hizmetler veya elektronik tüketim malları gibi karmaşık ürünleri tavsiye etmek için özellikle uygun olduğunu gösteriyor.

Ayrıca, kısıtlama temelli öneri uygulamaları tarafından desteklenen olası kullanıcı etkileşimi biçimleri de vardır, bunlar kısıtlama temelli öneri uygulamalarının başarıyla uygulandığı senaryoları rapor eder ve farklı teknik çözüm yaklaşımlarını gözden geçirir.

Topluluk temelli sistem türü, kullanıcının arkadaşlarının tercihlerine göre öneriler sunar. Bu teknik, "Bana arkadaşlarınızın kim olduğunu söyleyin, ben de size kim olduğunuzu söyleyeyim" epigramını takip eder. Kanıtlar, insanların benzer ancak anonim bireylerin tavsiyelerinden çok arkadaşlarının tavsiyelerine güvenme eğiliminde olduklarını göstermektedir. Açık sosyal ağların artan popülaritesi ile

birlikte, topluluk temelli sistemlere veya genellikle atıfta bulunulduğu gibi, sosyal tavsiye sistemlerine artan bir ilgi yaratmaktadır. Bu tür ÖS modelleri ve kullanıcıların sosyal ilişkileri ve tercihleri hakkında bilgi edinir. Tavsiye, kullanıcının arkadaşları tarafından sağlanan derecelendirmelere dayanmaktadır. Aslında bu ÖS'ler sosyal ağların yükselişini takip etmekte ve kullanıcıların sosyal ilişkileriyle ilgili basit ve kapsamlı bir veri elde edilmesini sağlamaktadır.

Hibrit Öneri Sistemleri, yukarıda bahsedilen tekniklerin kombinasyonuna dayanmaktadır. A ve B tekniklerini birleştiren bir hibrit sistem, B'nin dezavantajlarını düzeltmek için A'nın avantajlarını kullanmaya çalışır. Örneğin, İF yöntemleri yeni öge sorunlarından mustarıptir veya derecelendirme içermeyen ögeleri öneremezler. Bu, içerik tabanlı yaklaşımları sınırlamaz çünkü yeni ögeler için tahmin, tipik olarak kolayca erişilebilen özelliklerine dayalıdır. İki veya daha fazla ÖS tekniği göz önüne alındığında, yeni bir hibrit sistem oluşturmak için bunları birleştirmek için birkaç yol önerilmiştir.

Daha önce de belirttiğimiz gibi, bir öneri ararken kullanıcının bağlamı, sistemin çıktısını daha iyi kişiselleştirmek için kullanılabilir. Örneğin, geçici bir bağlamda, kışın tatil önerileri yaz aylarında sunulanlardan çok farklı olmalıdır. Ya da birinin arkadaşlarıyla bir Cumartesi akşamı için bir restoran tavsiyesi, iş arkadaşlarıyla bir iş günü öğle yemeği için önerilenden farklı olmalıdır.

Bağlamsal bilgileri öneri sürecine dahil etmek için üç popüler farklı algoritmik paradigma vardır: indirgeme temelli (ön filtreleme), bağlamsal son filtreleme ve bağlam modelleme. İndirgeme temelli (ön filtreleme) yöntemlerde, tavsiyeleri hesaplamak için yalnızca mevcut kullanım bağlamıyla eşleşen bilgiler, örneğin aynı bağlamda değerlendirilen ögelerin derecelendirmeleri kullanılır. Bağlamsal sonradan filtrelemede, öneri algoritması bağlam bilgisini yok sayar. Algoritmanın çıktısı, yalnızca hedef bağlamla ilgili önerileri içerecek şekilde filtrelenir / ayarlanır. Bağlamsal modellemede, üç yaklaşımdan daha karmaşık olanı bağlam verileri, tahmin modelinde açıkça kullanılır[8].

Veri Madenciliği alanında geliştirilen teknikler yardımıyla öneri görevleri çözülebilir. Öneri Sistemleri bağlamında kullanılan ana Veri Madenciliği teknikleri ve

bu tekniklerin başarıyla uygulandığı vakalar vardır. Özellikle başlıca teknikler: örnekleme veya boyut azaltma gibi ön işleme teknikleri; Bayes Ağları, Karar Ağaçları ve Destek Vektör Makineleri gibi sınıflandırma teknikleri; k-araçları gibi kümeleme teknikleri ve son olarak ilişkilendirme kuralları.

1.5. İŞBİRLİKÇİ FİLTRELEME

Öneri sistemleri web ile paralel olarak gelişmiştir. İnternetin büyümesi, tüm mevcut çevrimiçi bilgilerden yararlı bilgileri etkili bir şekilde çıkarmayı çok daha zor hale getirdi. Çok büyük miktarda veri, verimli bilgi filtreleme için mekanizmalar gerektirir. İşbirlikçi filtreleme, bu problemin üstesinden gelmek için kullanılan tekniklerden biridir. Uygulamasını e-ticaret ve e-öğrenmeden sosyal ağlara ve web aramaya kadar değişen alanlarda bulmuştur. İşbirlikçi filtrelemeyle ilgili geniş alanı, teknikleri ve zorlukları nedeniyle, temel ilkesini, çalışmasını ve bütünüyle uygulamasını anlamak için tek bir yerde toplanmasını gerektirir[2].

İşbirlikçi filtreleme, kullanıcı tarafından aranan bilgileri ve kalıpları, bakış açıları, çoklu araçlar ve matrislerde depolanan kullanıcıların davranışları hakkında önceden var olan veriler gibi çoklu veri kümeleri ile işbirliği yaparak filtreleyen bir teknik olarak tanımlanır. Çok büyük bir veri kümesi mevcut olduğunda işbirliğine dayalı filtreleme gereklidir. İşbirlikçi filtreleme yöntemleri, verilerin algılanması ve izlenmesi, kontrol hattı ve maden arama gibi çeşitli biçimlerde çok sayıda veriye sahip çok çeşitli alanlar için tavsiye sistemleri oluşturmak için kullanılır. E-ticaret ve web siteleri, satışları artırmak için kullanıcılara ürün önermek; bankalar ve borsalar gibi finansal hizmetler sunan kurumların finansal verileri; her türlü sensör ve faaliyetten verilerin alındığı geniş coğrafi alanların algılanması; bunlardan birkaçıdır[2].

Biraz daha yeni ve biraz dar olan işbirlikçi filtrelemenin bir tanımı da filtreleme olarak bilinen bir süreç olan tahmin yapma sürecini, bir kullanıcının tercihleri ve hoşlanmadıkları hakkında veri toplayarak otomatikleştirmenin bir yolu olduğunu belirtir. İşbirlikçi filtreleme yaklaşımının altında yatan varsayım, bir A kişinin bir konu hakkında B kişisiyle aynı görüşe sahip olması durumunda, A'nın B ile ilişkili olduğu ve farklı bir konu hakkındaki görüşüne benzer bir görüşe sahip olma olasılığının daha yüksek olduğudur. Bu tür tahminlerin kullanıcıya özgü olması dikkat

çekicidir, ancak çok sayıda kullanıcıdan gelen veriler kullanılarak oluşturulmuştur. Kullanıcının yaş, cinsiyet ve konum gibi kişisel bilgileri genellikle işbirlikçi filtrelemede kullanılmaz, ancak kısmen gözlemlenen bir derecelendirme matrisi kullanılır. Derecelendirme matrisi ikili veya sıralı olabilir. İkili matris, kullanıcının adı veya kimliği satırlarda yer alırken, kullanıcıların beğenme veya beğenmeme şeklinde sütunlarda derecelendirmelerini içerir. Sıralı matris son zamanlarda sıklıkla kullanılan bir sistem olan mükemmel, çok iyi, iyi, ortalama, zayıf veya basitçe beş veya on yıldız şeklinde kullanıcıdan gelen bir dizi yanıt şeklinde derecelendirmeleri içerir. Derecelendirme matrisi, web sitesinin sunucusu tarafından, örneğin tıklama akışı günlüğü kullanılarak kolayca örtük olarak toplanabilir. Sağlanan mal veya hizmetlerin sayfalarına yönelik bağlantılara yapılan tıklamalar, kullanıcının olumlu bir incelemesi olarak kabul edilebilir. Derecelendirme matrisleri yararlı olabilirken, önemli bir dezavantajı, son derece seyrek olmalarıdır, bu nedenle benzer kullanıcıları sınıflarda bir araya toplamak çok zordur. Bu, her kullanıcının her bir ürün hakkında yorum yapmamasından kaynaklanmaktadır. Bu nedenle, işbirlikçi filtreleme, bu seyrek verilerin depolanması ve bir öneri sistemi oluşturmak için analiz edilmesinden oluşur[15].

1.5.1. Kullanıcı Temelli İşbirlikçi Filtreleme

Kullanıcı Bazlı İşbirlikçi Filtreleme, hedef kullanıcının beğenisine benzer zevke sahip diğer kullanıcılar tarafından o öğeye verilen derecelendirmeler temelinde bir kullanıcının beğenebileceği öğeleri tahmin etmek için kullanılan bir tekniktir. Birçok web sitesi, öneri sistemlerini oluşturmak için işbirliğine dayalı filtreleme kullanır.

1.5.2. Öge Temelli İşbirlikçi Filtreleme

Öneri sistemi, kullanıcı tarafından daha önce satın alınan bir dizi öge gibi, kullanıcının ilgisini çekeceği bilinen bir dizi öğeye dayalı olarak öğeleri tek tek kullanıcılara önerir. Ürünler bir tüccarın Web sitesinin kullanıcılarına tavsiye etmek için kullanılır. Hizmet, öğeleri "benzer" öğelerin listelerine eşleyen önceden oluşturulmuş bir tabloyu kullanarak öneriler üretir. Tabloda yansıtılan benzerlikler, kullanıcı topluluğunun ortak çıkarlarına dayanmaktadır. Örneğin, bir düzenlemede,

Benzerlikler, kullanıcılar tarafından satın alınan öğeler arasındaki korelasyonlara dayanmaktadır (örneğin, A ve B öğeleri benzerdir çünkü A öğesini satın alan kullanıcıların nispeten büyük bir kısmı B öğesini de satın almıştır). Tabloda ayrıca tek tek öğeler arasındaki benzerlik derecelerini gösteren puanlar da yer almaktadır. Sistem kişisel tavsiyeler üretmek için, kullanıcının ilgisini çekeceği bilinen öğelere karşılık gelen benzer öğeler listelerini tablodan alır. Bu benzer öğeler listeleri uygun şekilde tek bir liste halinde birleştirilir ve bu liste daha sonra sıralanır (birleşik benzerlik Puanlarına göre) ve önerilen öğelerin bir listesini oluşturmak için filtrelenir. Ayrıca tavsiyelerde bulunmak için bir kullanıcının elektronik alışveriş sepetinin mevcut ve / veya geçmiş içeriklerinin kullanılmasına yönelik çeşitli yöntemler de vardır. Bir düzenlemede, kullanıcı birden çok alışveriş sepeti oluşturabilir ve belirlenmiş bir alışveriş sepetine özel tavsiyeler almak için öneri sistemini kullanabilir. Başka bir düzenlemede tavsiyeler, bir kullanıcının alışveriş sepetinin mevcut içeriğine dayalı olarak üretilir, böylece tavsiyeler, kullanıcı tarafından gerçekleştirilen mevcut alışveriş görevine karşılık gelme eğilimindedir[16].

1.5.3. Bellek Temelli İşbirlikçi Filtreleme

CF için iki ana yaklaşım vardır: Bellek temelli CF ve model temelli CF. Bellek temelli algoritma, tüm veri tabanını sistem belleğine yükler ve bu tür bir bellek veri tabanına dayalı olarak öneri için tahmin yapar. Basittir ancak büyük veri sorunuyla karşılaşır.

1.5.4. Model Temelli İşbirlikçi Filtreleme

Model temelli algoritma, devasa veri tabanını bir modele sıkıştırmaya çalışır ve bu modele referans mekanizması uygulayarak öneri görevini yerine getirir. Model temelli CF, kullanıcının isteğine anında yanıt verebilir.

1.5.5. Hibrit İşbirlikçi Filtreleme

A ve B tekniklerini birleştiren bir hibrit sistem, B'nin dezavantajlarını düzeltmek için A'nın avantajlarını kullanmaya çalışır. Örneğin, İF yöntemleri yeni öğe sorunlarından mustarıptır veya derecelendirme içermeyen öğeleri öneremezler. Bu, içerik tabanlı yaklaşımları sınırlamaz çünkü yeni öğeler için tahmin, tipik olarak kolayca erişilebilen özelliklerine dayalıdır. İki veya daha fazla ÖS tekniği göz önüne

alındığında, yeni bir hibrit sistem oluşturmak için bunları birleştirmek için değişik yollar önerilmiştir.

1.6. ÖNERİ SİSTEMLERİ DEĞERLENDİRMESİ

Öneri sistemleri araştırması, uygulamaya ve ticari uygulamalara güçlü bir vurgu yapılarak yürütülmektedir. ÖS dağıtımının pratik tarafı ile ilgili çok önemli bir konu, sistemlerin kalitesini ve değerini değerlendirme gerekliliğidir. Sistemin yaşam döngüsünün farklı aşamalarında ve çeşitli amaçlar için değerlendirme gereklidir. Tasarım zamanında, uygun tavsiye yaklaşımının seçimini doğrulamak için değerlendirme gereklidir. Tasarım aşamasında, değerlendirme çevrimdışı olarak uygulanmalıdır ve öneri algoritmaları, yani hesaplanan önerileri, depolanan kullanıcı etkileşimleriyle karşılaştırır. Çevrimdışı bir değerlendirme, kullanıcı etkileşimlerinin aynı veri kümelerinde birkaç algoritmanın çalıştırılması ve performanslarının karşılaştırılmasından oluşur. Bu tür bir değerlendirme, uygun veriler mevcutsa, genellikle mevcut anonim karşılaştırma verileri üzerinde veya aksi takdirde toplanan veriler üzerinde gerçekleştirilir. Çevrimdışı deneylerin tasarımı, güvenilir sonuçlar sağlamak için bilinen deney tasarım uygulamalarını takip etmelidir. Çevrimdışı deneyler, tavsiye görevini yerine getirirken seçilen algoritmanın kalitesini ölçebilir. Ancak, bu tür bir değerlendirme, kullanıcı memnuniyeti, sistemle ilgili kabul veya deneyim hakkında herhangi bir fikir veremez. Algoritmalar, temel öneri problemini çözmede, yani kullanıcı derecelendirmelerini tahmin etmede çok doğru olabilir, ancak başka bir nedenle, örneğin sistemin performansı beklendiği gibi olmadığı için sistem kullanıcılar tarafından kabul edilmeyebilir[17].

Bu nedenle, kullanıcı merkezli bir değerlendirme de gereklidir. Sistem başlatıldıktan sonra çevrimiçi olarak veya odaklanmış bir kullanıcı çalışması olarak gerçekleştirilebilir. Çevrimiçi değerlendirme sırasında, gerçek kullanıcılar, arka planda çalışan deneyin tam yapısının farkında olmadan sistemle etkileşime girer. Sistem performansını artırmak için, sistem günlüklerinin karşılaştırılması ve analizi için farklı kullanıcı grupları üzerinde algoritmaların çeşitli sürümlerini çalıştırmak mümkündür. Ek olarak, algoritmaların çoğu ağırlık eşikleri, komşuların sayısı vb. sabit ayarlama ve ölçümleme gerektiren parametreleri içerir[8].

Çevrimiçi değerlendirme mümkün olmadığında veya çok riskli olduğunda odaklanmış kullanıcı çalışmaları yapılır. Bu tür bir değerlendirmede, küçük bir kullanıcı grubundan sistemin çeşitli sürümleriyle farklı görevleri yerine getirmesinin istendiği kontrollü bir deney planlanır. Bu durumda, kullanıcının performansını analiz etmek ve kullanıcıların deneyimleri hakkında rapor verebilmesi için anketler dağıtmak mümkündür. Bu tür deneylerde, sistemler hakkında hem nicel hem de nitel bilgi toplamak mümkündür[17].

Son yıllarda, öneri sistemleri için kullanıcı merkezli değerlendirme prosedürlerine ve ölçülerine ilgi artmıştır. Araştırmacılar, tavsiye eden sistemlerin hedeflerinin, kullanıcıları elde tutmaya ve memnuniyetine yol açan yararlı ve eğlenceli, kişiselleştirilmiş bir deneyim sağlamak için araçlar olarak algoritmaların doğruluğunun ötesine geçtiğini fark ettiler. Bu yaklaşım, bir ÖS'nin değerlendirilen yönlerini, tercihlerin ortaya çıkarılma biçimi, önerilen sonuçların sunumu ve son olarak, kullanıcılara verilen açıklamalardır. Açıklamalar birkaç amaca hizmet edebilir: en popüler olanı sonuçların gerekçelendirilmesidir, yani kullanıcıya sistemin neden belirli bir ögeyi önermeye karar verdiğini açıklamaktır. Diğer hedefler arasında sisteme olan güveni artırmak, kullanıcıyı önerilen ögeyi satın almaya ikna etmek ve kullanıcının karar vermesine yardımcı olmak yer alabilir. Öneri açıklamasının değerlendirmesini tasarlarırken, açıklamanın amacını belirlemek ve bunu ölçmek için uygun bir metriği ayarlamak önemlidir[17].

Tavsiye eden sistemleri, yani çevrim dışı, çevrim içi ve kullanıcı çalışmalarını değerlendirmek için yapılabilecek daha önce bahsedilen üç deney türü vardır. Avantajlarını ve dezavantajlarını ve değerlendirilecek özellikleri dikkate alarak değerlendirme yöntemlerini seçmeye yönelik kılavuzlar da vardır. Genellikle bir algoritmanın öngörüsünün doğruluğuna ve ilgili ölçümlere odaklanan literatürdeki mevcut değerlendirme tartışmalarının aksine, özelliğe yönelik değerlendirmeye odaklanmak daha yerinde olabilir. Sistemin başarısıyla alakalı çok sayıda özellik vardır. Özelliklerin her biri için, uygun deney türü ve ilgili önlemler önerilir. Özellikler arasında şunlar yer alır: kapsam, soğuk başlangıç, güven, yenilik, risk ve şans[17].

Kullanıcı merkezli değerlendirme önemlidir. Kullanıcının sistemle ilgili deneyimini değerlendirmek için kullanıcı merkezli deneylerin nasıl yapılacağına dair

ayrıntılı ve pratik yönergeler vardır. Tavsiye eden sistemlerin yönlerini ve değerlendirilmesi gereken kullanıcılarla etkileşimlerini haritalayan teorik bir kullanıcı merkezli değerlendirme çerçevesi vardır. Öğrenciler ve araştırmacılar için kullanıcı deneyleri yapmak için pratik yönergeler gerekir. Hipotezleri belirtmek, katılımcıları işe almak, deneylerin tasarımı ve sonuçların istatistiksel analizini yapmak gerekir. Son olarak, ilgili literatürden gerçek sistemlerin değerlendirmelerinin birçok örneğini araştırmak gerekir[17].

1.7. ÖNERİ SİSTEMLERİ UYGULAMALARI

Öneri sistemleri araştırması, teorik katkısının yanı sıra, genellikle endüstriyel ÖS'leri pratik olarak iyileştirmeyi amaçlamaktadır ve sistemlerin uygulanması için geçerli olan çeşitli pratik hususlar hakkında araştırmayı içerir. Aslında, ÖS, makine öğrenimi ve veri madenciliği algoritmalarının büyük ölçekli kullanımına bir örnektir. Hem araştırma topluluğundan hem de endüstriyel alandaki ortak ilgi, bir yandan araştırma için verilerin kullanılabilirliğini, diğer yandan gelişmiş algoritmaların daha da gelişmesini güçlendirdi. ÖS'lerdeki pratik araştırmalar, bir ÖS'nin yaşam döngüsündeki farklı aşamalarla ilgili olan yönleri inceler, yani sistemin tasarımı, uygulanması, değerlendirilmesi, bakımı ve sistem çalışması sırasında iyileştirilmesi. 2006 yılında açıklanan Netflix Ödülü, tavsiye sistemleri araştırma topluluğu ve endüstrisi ve bunların karşılıklı etkileşimi için önemli bir olaydı. Ögelerin kullanıcılara önerilmesinin önemini vurguladı ve birçok yeni veri madenciliği öneri tekniğinin geliştirilmesini hızlandırdı. Netflix Ödülü pek çok araştırma faaliyeti başlatmış olsa da, ödül tam öneri sorununun basitleştirilmesiydi. Tahmin edilen ve gerçek derecelendirmeler arasında Kök Ortalama Kare Hatasını (RMSE) optimize ederken kullanıcı derecelendirmelerini tahmin etmekten oluşuyordu. 2009 yılında verilen ödülünden öğrenilen dersler oldu ve gerçek dünya tavsiye sistemleri için bir vaka çalışması olarak Netflix sistemini kullanan ÖS'lerin endüstriyel ayarları hakkında bilgi geliştirdi. Ek olarak, gerçek dünya ÖS geliştirirken dikkat edilmesi gereken ÖS uygulama sorunları hakkında endüstri perspektifi de edindi. Kullanıcıya optimize edilmiş kişiselleştirilmiş bir deneyim sağlamak için Netflix bir sorun için en iyi modeli seçmek için veri merkezli yaklaşımı kullanır. Netflix geliştirmesinden öğrendiğimiz

dersi kısaca özetleyecek olursak yenilikçi algoritmaların geliştirme ve değerlendirme sürecini destekleyebilecek ve büyük hacimli mevcut verilere dayalı öneriler sunabilecek uygun bir ölçeklenebilir sistem mimarisine ihtiyaç vardır[8].

Bir ÖS tasarlarırken dikkate alınması gereken ilk faktör, uygulanması gereken algoritmik yaklaşım üzerinde büyük bir etkiye sahip olduğu için uygulamanın alanıdır. Montaner ÖS'lerin bir sınıflandırmasını sağlar ve mevcut ÖS uygulamalarını belirli uygulama alanlarına göre sınıflandırır. Bu özel uygulama alanlarına dayanarak, en yaygın öneri sistemi uygulamaları için daha genel alan sınıfları tanımlayabiliriz[18]:

- Eğlence - filmler, müzik, oyunlar ve IPTV için öneriler.
- E-ticaret — tüketiciler için kitap, kamera, PC gibi satın alınabilecek ürün önerileri.
- Sosyal - sosyal ağlardaki kişilerin tavsiyeleri, Facebook, tweetler beslemeleri, LinkedIn güncellemeleri ve diğerleri gibi içerik sosyal medya içeriği önerileri.
- İçerik - kişiselleştirilmiş gazeteler, belgeler için öneriler.
- Web sayfaları, e-öğrenme uygulamaları ve e-posta filtreleri.
- Hizmetler — seyahat hizmetleri tavsiyeleri, danışma için uzman tavsiyeleri, kiralanacak ev tavsiyeleri veya çöpçatanlık hizmetleri.

Tavsiye sistemleri daha popüler hale geldikçe, sigortalıları tavsiye etme veya soru cevaplama sistemleri için sorular önerme gibi yeni ve çeşitli uygulamaların potansiyel avantajlarına ilgi artmaktadır. Yukarıdaki liste şu anda ÖS teknikleriyle ele alınan tüm uygulama alanlarını kapsayamayacağından, çeşitli uygulama alanlarının sadece birkaç örneğini vermektedir. Belirli bir uygulama alanı için ÖS'nin geliştiricisi, alanın belirli yönlerini, gereksinimlerini, uygulama zorluklarını ve sınırlamalarını anlamalıdır. Ancak bu faktörleri analiz ettikten sonra kişi en uygun tavsiye algoritmasını seçebilir ve etkili bir insan-bilgisayar etkileşimi tasarlayabilir. Burada belirli alanlardaki tavsiye sistemlerinin uygulamaları açıklanmaktadır. Bu uygulamaların her biri, ÖS'nin belirli bir alan için gereksinimlerini, kesin zorluklarını ve bunları ele almak için uygun teknolojileri ve algoritmaları ifade etmektedir.

Belirli bir alan için tasarlanmış bir ÖS'nin ayrıntılı bir örneği, teknolojiyle geliştirilmiş öğrenme (TEL) için öneri sistemlerini açıklar [19]. Genel olarak her tür öğretim ve öğrenme faaliyetini destekleyen teknolojileri kapsayan TEL, hem bireylerin hem de kuruluşların öğrenme uygulamalarını geliştirmek için yeni yöntem ve teknolojileri tasarlamayı, geliştirmeyi ve test etmeyi amaçlamaktadır. Eğitim ve öğrenmenin dijital yolu daha popüler hale geldikçe, içeriğin kişiselleştirilmesini öğrenme sürecine entegre etme ihtiyacı ve algoritmaların kalitesini değerlendirmek için mevcut veriler TEL ÖS'lerin popülerliğinin artması için fırsatlar yarattı. TEL, öğrenme sürecini kişiselleştirmek ve bunu kullanıcının önceki bilgilerine, yeteneklerine ve tercihlerine göre ayarlamak için tavsiye sistemleri teknolojisini entegre etmekten büyük ölçüde yararlanabileceğinden, TEL'e uygulanan ÖS'lerde önemli bir artış vardır. TEL için üç ana kategoriden oluşan bir sınıflandırma çerçevesi kullanarak kategorize edebiliriz: Desteklenen Görevler, Yaklaşım ve Çalışma. Çerçeve kullanılarak sistemlerin analizi, her bir kümenin alana benzersiz bir katkı biçimini temsil ettiği yedi TEL ÖS kümesine ayrılabilir. Burada, değerlendirme ayarlarını ve önlemlerini anlatmak için TEL ÖS alanı hakkında bir genel bakış sağlamak amaçlanmaktadır[19].

Tavsiye için bir başka popüler alan ise, müzik için ÖS'ler tasarlanırken ve değerlendirilirken, öneriler için çeşitli zorluklar ortaya çıkaran müzik öğelerinin benzersiz özellikleri dikkate alınmalıdır. Bu tür zorluklar, örneğin, bir kullanıcının bir film veya kitaba kıyasla önerilen bir öğe hakkında fikir edinmesi için geçen kısa süreyi veya aynı öğenin birçok kez önerilebileceği gerçeğini içerir. Ek olarak, müzik tek bir öğe, bir çalma listesi olarak önerilebilir ve türe, sanatçıya veya gruba göre gruplanabilir. Müzik ÖS'leri, diğer birçok alanın aksine, büyük ölçüde, alana özgü zorluklar anlamına gelen içerik temelli önerilere dayanmaktadır.

Çok popüler hale gelen yeni teknolojilerin ortaya çıkmasıyla birlikte öneri için bazı yeni uygulama alanları gelişti. Örneğin GPS gibi cihazların özel yeteneklerini kullanan mobil cihazlar için belirli ÖS'lerin geliştirilmesini hızlandıran mobil teknolojinin evrimidir. ÖS konum temelli bir mobil tavsiye sisteminin ana bileşenlerini gözden geçirmektedir. Kullanıcı profillerini zenginleştirmek için mobil sensörlerden yararlanabilecek mobil bağlam temelli ÖS'lerin çeşitli uygulama alanları varken,

mobil konum temelli önerileri vurgulamak gerekir. Bu tür ÖS uygulamaları, kullanıcının konumuna ve davranış ve tercih geçmişine göre yerler ve mekânlar önerir. Tavsiye mekânları için uygulanan algoritmaları ve tavsiyelerin kalitesini değerlendirmek için değerlendirme prosedürlerini açıklar. Teslim alma yerlerindeki taksi şoförleri için öneriler veya bir perakende mağazasının nerede bulunacağına dair öneriler gibi ek konum tabanlı uygulamalar da yapılabilir[20].

Yeni teknolojilerle ortaya çıkan yeni ÖS'lere bir başka örnek, sosyal web ile ilgili ve özellikle sosyal medya alanını hedefleyen tavsiye sistemleridir. Sosyal ağların yükselişiyle kullanıcılar bilgi, faaliyet ve etkileşimlerle aşırı yüklenir. Sosyal tavsiye sistemleri, kullanıcıya ilgili içeriği tanımlamada yardımcı olmayı ve yalnızca ilgili faaliyetler ve etkileşimlerde yer almayı amaçlayan öneriler sunmalıdır. Sosyal medyaya özel olarak geliştirilmiş ÖS'lerin dışında, diğer alanların tavsiye sistemleri, standart ÖS'lerin kalitesini artırmak için sosyal medyanın kullanıcılar hakkında sunduğu yeni veri türlerinden yararlanabilir. Sosyal ÖS terimi, bulunduğu sosyal medya platformuyla ilgili birçok ÖS türünü kapsar. İki ana tür vardır: sosyal medya içeriği önerileri ve insanların önerileri[21].

Özel bir sosyal tavsiye biçimini, bir tavsiyede yer alan iki tarafın tatmin olmasını gerektiren karşılıklı kişiler arası tavsiyeleri vurgulamaktadır. Bazı örnekler şunlardır: kitap önerileri, insan kaynakları önerileri, çalışanların ve işverenlerin önerileri ve öğrenme grupları için öğrenci gruplarının önerilmesi. Her iki tarafın da ihtiyaç duyduğu benzersiz karşılıklı memnuniyet tarzının yanı sıra, kullanıcıların açık geri bildirim sağlama istekliliği, kullanıcıların genellikle sistemle uzun süre meşgul olması gibi geleneksel ve karşılıklı öneriler arasındaki diğer farklılıkları ve kullanıcıları önerilerle aşırı yüklememe gerekliliğini vurgulamak gerekir[21].

Sosyal Web ayrıca, Web arama zorluklarını ele almak ve gelişmiş arama özelliklerini uygulamak için öneri tekniklerine güvenen modern arama motorları tarafından da kullanılmaktadır. Spesifik olarak, çeşitli motorlar, bir kullanıcı sorgusuna yalnızca sorgu terimleriyle ilgili olmayan, aynı zamanda kullanıcının sosyal Web'deki etkinlikleri kendi geçmişinden çıkarılan şekilde kullanıcının arama geçmişine, itibarına ve tercihlerine uygun sonuçlar üreterek bir tür kişiselleştirme ve işbirliği uygulamaya çalışır.

Bilgi Erişiminin (Information Retrieval- IR) araştırma hedeflerini ve ÖS perspektifinden kişiselleştirilmiş Web araştırmasını tartışmak gerekir. Son ÖS araştırmaları ortaya çıkan tekniklerin arama motoru zorluklarını ele almak için nasıl uygulanabileceğini gösteriyor. Arama motoru iyileştirmesi için iki gelecek vaat eden fikir vardır: kişiselleştirme ve işbirliği. Kullanıcı tercihlerini kullanarak web aramalarını kişiselleştirmek için bir dizi farklı yaklaşım sunup, arama sonuçlarını etkileyecek bağlam bilgilerini açıklamak gerekir. Buna ek olarak, çeşitli bilgi arama görevlerini yerine getirirken arkadaşlar, meslektaşlar veya benzer ihtiyaçları olan kullanıcılar arasındaki potansiyel işbirliği potansiyelinden yararlanmaya çalışan işbirlikçi bilgi erişimi alanındaki son çalışmaları detaylandırmak da gerekir. Bu yeni çizgi sosyal arama olarak adlandırılan araştırma, benzer kullanıcıların deneyimlerinden ve tercihlerinden etkilenen arama sonuçları sağlamada Web'in sosyal medya özelliğinden yararlanır. Bir arama motorunun modern öneri teknolojileri için benzersiz bir platform sağlayacağı bir öneri sistemleri ve arama motorları yakınsaması öngörülmüyor. Bu kaynakları arama motoru algoritmalarına entegre etmenin ve kullanıcıların ihtiyaçlarını anlamaya çalışan proaktif bir arama deneyimi tarzının, doğru zamanda doğru bilgileri alabilen oldukça memnun kullanıcılar ile sonuçlanacağına inanılıyor. Arama motorlarını etkileyen bir diğer eğilim, mobil cihazlar üzerinden yapılan arama faaliyetlerinin artmasıdır. Bu, arama ve keşif ara yüzlerine yeni kısıtlamalar getirir, ancak aynı zamanda gelişmiş kişiselleştirmeye izin veren mobil sensörleri kullanan yenilikler için fırsatlar da getirir[22].

1.8. ÖNERİ SİSTEMLERİ VE İNSAN BİLGİSAYAR ETKİLEŞİMİ

Önceki bölümlerde gösterdiğimiz gibi, araştırmacılar temel olarak bir dizi teknik çözüm tasarlamak ve hedef kullanıcı tarafından neyin sevildiği ve ne kadar beğenildiği hakkında daha iyi tahminler elde etmek için çeşitli bilgi kaynaklarını kullanmakla ilgilenmişlerdir. Bu araştırma faaliyetinin altında yatan varsayım, basitçe bu doğru önerileri veya en iyi seçenekleri sunmanın yeterli olmadığıdır. Diğer bir deyişle, öneriler kendi adına konuşmalı ve kullanıcı, eğer doğru ise önerileri kesinlikle kabul etmelidir. Bu açıkça tavsiye sorununun aşırı basitleştirilmiş bir açıklamasıdır ve tavsiyelerin verilmesi o kadar kolay değildir.

Uygulamada, özerk bir karar vermek için yeterli bilgiye sahip olmadıkları için kullanıcıların tavsiyelere ihtiyacı vardır. Sonuç olarak, önerilen tavsiyeyi değerlendirmek onlar için kolay olmayabilir. Bu nedenle, çeşitli araştırmacılar bir tavsiyenin belirli bir kullanıcı tarafından kabul edilmesine yol açan faktörleri anlamaya çalışmışlardır.

Swearingen ve Sinha, bir ÖS'nin etkinliğinin tahmin algoritmasının kalitesinin ötesine geçen faktörlere bağlı olduğunu ilk belirtenler arasındaydı. Aslında, tavsiye eden kişi ayrıca kullanıcıları önerilen öğeleri denemeye veya okumaya, satın almaya, dinlemeye, seyretmeye vb. ikna etmelidir. Bu, elbette, seçilen öğelerin bireysel özelliklerine ve dolayısıyla öneri algoritmasına bağlıdır. Süreç aynı zamanda öğeler sunulduğunda, karşılaştırıldığında ve açıklandığında sistem tarafından desteklenen belirli insan / bilgisayar etkileşimine de bağlıdır. Swearingen ve Sinha, bir kullanıcının bakış açısından, etkili bir tavsiye sisteminin sisteme güven vermesi gerektiğini ve en azından biraz şeffaf bir sistem mantığına sahip olması gerektiğini buldu. Ek olarak bu çalışmada kullanıcıları yeni, henüz deneyimlenmemiş öğelere yönlendirmek gerektiğini ve resimler ve topluluk derecelendirmeleri dâhil önerilen öğeler hakkında ayrıntılar verilmesi gerektiğini ve son olarak önerileri iyileştirmenin yollarının sunulmasının gerektiği belirtiyorlar[23].

Swearingen ve Sinha ve benzer şekilde yönlendirilmiş diğer araştırmacılar, öneri algoritmasının önemini azaltmazlar, ancak etkinliğinin yalnızca tahminin doğruluğu açısından, yani standart ve popüler IR ölçümleri açısından değerlendirilmemesi gerektiğini iddia eder, örneğin Ortalama Mutlak Hata (MAE), kesinlik veya Normalize İndirgenmiş Kümülatif Kazanç (NDCG). Tavsiye eden sistemin kabulü ve tavsiyeleri ile ilgili diğer boyutlar ölçülmelidir. Bu fikirler oldukça iyi sunulmuş ve diğer araştırmacılar tarafından da tartışılmıştır. Bu çalışmada tavsiye listelerinin benzerliği, öneri tesadüfleri ve bir tavsiyede kullanıcı ihtiyaç ve beklentilerinin önemi dâhil olmak üzere tavsiye sistemlerini değerlendirmek için kullanıcı merkezli talimatlar önerilmektedir[23].

Öneri sistemleri, öneri amaçları için gerekli olan muazzam miktarda kullanıcı verisi toplar. Ancak, bu verilerin kullanılabilirliği, bu verilerin, özellikle güvenilmeyen taraflarca erişiliyorsa veya kötü amaçlı araçlar tarafından kötüye kullanılıyorsa, son

kullanıcının gizlilik beklentilerini ihlal edecek şekilde kullanılmasına neden olabilir. Tavsiye eden sistemler tarafından uygulanan kullanıcı gizliliğine yönelik riskleri analiz edilmeli, mevcut çözümler araştırılıp, tavsiye edenlerin kullanıcıları için gizlilik etkilerini arttırmaları gerekir. Özellikle, tek bir kullanıcı modelleme verisi deposunu ortadan kaldıran çeşitli merkezi olmayan çözümler kullanarak kullanıcı gizliliğini koruyan mimari çözümler kullanılmalı, aksi takdirde tavsiye edene yönelik kötü niyetli saldırıların hedefi olur. Ek olarak, orijinal kullanıcı modelleme verilerini bozan veya resmi şifreleme yöntemlerini uygulayan algoritmik çözümler de kullanılabilir. Bunlar, güvenilmeyen bir tarafça erişilse bile, orijinal veriler yerine yalnızca değiştirilmiş veya şifrelenmiş kullanıcı verilerinin açığa çıkacağını garanti eder. Son olarak, politika odaklı çözümler de kullanılabilir. Bu çözümler, kişisel kullanıcı verilerinin depolanmasını, aktarılmasını ve kullanılmasını sınırlayan direktiflere ve mevzuat girişimlerine yöneliktir[24].

Tavsiye sistemlerinin temel amacı, kullanıcıların daha iyi seçimler yapmalarına yardımcı olmaktır. Bu nedenle, insanların nasıl seçim yaptıklarını ve insan karar verme sürecinin nasıl desteklenebileceğini anlamak önemlidir. Bu konuda, geniş bir psikolojik araştırma literatürüne dayanan ve tavsiye eden sistemler araştırma topluluğu için uygun ve erişilebilir olacak şekilde formüle edilmiş, günlük seçim ve karar verme psikolojisine genel bir bakış atabiliriz. "İnsanlar nasıl seçim yapar?" Sorusu Psikoloji, ekonomi ve diğer alanlarda bu konudaki geniş ve etkileyici bilimsel literatüre aşina olsanız bile, yanıtlanması şaşırtıcı derecede zordur. Öneri sistemleri insanlar, "Bir bilgisayar programı nasıl tavsiyelerde bulunabilir?" Sorusuyla aynı zorluk derecesine çıkacağından, bunun nedenini anlamak için iyi bir konumdadır. Her iki durumda da en üst düzey cevap şudur: "Çok sayıda farklı yaklaşım vardır ve bunlar çeşitli şekillerde birleştirilebilir." Öneri sistemlerinin, seçimi kolaylaştırmak için mevcut birçok araçtan biri olarak nasıl görülebileceğini hesaplamak gerekir. Ardından, insanların daha iyi seçimler yapmalarına yardımcı olacak stratejilere üst düzey bir genel bakış atmak gerekir, bu tavsiye eden sistemlerin daha büyük seçim resmine nasıl uyduğunu gösterir. Ek olarak, ÖS'lerin temel işlevleri: tercih modellerini oluşturmak için bilgileri ortaya çıkarmak, geniş bir seçenek kümesini daraltmak, kullanıcıların küçük bir önerilen seçenekler kümesi arasından seçim yapmalarına yardımcı olmak ve kullanıcıların

geniş seçenek alanlarını keşfetmelerine yardımcı olmak. İnsan karar verme anlayışının bu süreçlerle ilgili araştırma ve uygulamaları nasıl aydınlayabileceği konusunu daha net bir şekilde görmüş oluyoruz[25].

Önceki bölümlerde tartışıldığı gibi, tavsiye sistemleri genellikle tavsiyelerde bulunmak için karmaşık algoritmalar kullanır. Ancak, ÖS, bir sistem tarafından sağlanan tavsiyenin her zaman kullanıcıları tarafından kabul edileceğini varsayamaz. Bir önerinin güvenilir bir tavsiye olarak görülüp dikkate alınmaması, yalnızca kullanıcıların öneriye ilişkin algılarına değil, aynı zamanda bir tavsiye veren olarak sisteme de bağlıdır.

Bir tavsiyenin güvenilir bir tavsiye olarak görüldüğünü ve aslında sadece kullanıcının öneriye ilişkin algıları nedeniyle değil, aynı zamanda bir tavsiye veren olarak algılanan sistemin temel rolü nedeniyle de dikkate alındığını vurgulamak gerekir. Aslında, ikna ile ilgili literatür, insanların güvenilir kaynaklardan gelen önerileri kabul etme olasılığının yüksek olduğunu öne sürüyor ve bu nedenle, ÖS'nin güvenilirliğinin, tavsiye kabul etme olasılığını artırmak için hayati önem taşıdığı sonucuna varıyoruz. Bu nedenle, ÖS'lerin güvenilirliğinin nasıl artırılacağı üzerinde daha fazla çalışmak daha iyi sonuçlar kazandırabilir.

İnsan-insan, insan-teknoloji ve insan - tavsiye eden sistem etkileşimleri bağlamında kaynak faktörlere ilişkin mevcut literatürü gözden geçirmek gerekir. Ayrıca, sosyal teknolojinin artan popülaritesi ışığında sistem güvenilirliği değerlendirmesini de ele almak gerekir. İnsan ve teknoloji etkileşimi bağlamında ve özellikle öneri sistemleri alanında incelenen kaynak özellikleri de tartışılmaktadır. Son olarak, diğer bağlamlarda etkili olduğu tespit edilen birçok sosyal ipucunun, tavsiye eden sistemlere göre henüz uygulanmamış ve test edilmemiş olduğu sonucuna varılabilir.

Kişilik, bireylerin kalıcı duygusal, kişilerarası, deneyimsel, tutumsal ve motivasyon tarzlarında farklılık gösterdiği en önemli yolu açıklar. Araştırmalar, kişiliğin özellikle soğuk başlangıç problemi ve çeşitli önerilerle başa çıkmada yararlı olduğunu göstermiştir[26].

Kişiliğin kullanıcı tercihleriyle nasıl ilişkili olduğunu ve kişiliğin tavsiye sistemlerinde nasıl kullanılacağını tartışmak gerekir. Kişiliğin Beş Faktör Modeli vardır. Bu model, ana faktörlere karşılık gelen özellikler açısından kolayca ölçülebildiği için tavsiye sistemlerinde kullanım için uygun görünmektedir. Gözlemlenen bir kullanıcı için kişilik faktörlerinin edinimi, açık bir şekilde anketler yoluyla veya sosyal medya akışları veya cep telefonu çağrı günlükleri gibi modalitelerde makine öğrenimi yaklaşımları kullanılarak dolaylı olarak yapılabilir[26].

1.9. ÖNERİ SİSTEMİNDE KARŞILAŞILAN ZORLUKLAR

1.9.1. Soğuk Başlangıç Problemi

Öneri sistemi, öğelere ilişkin tavsiyelerde bulunmak için öğeler ve / veya yardımcı bilgiler üzerindeki geçmiş kullanıcı derecelendirmelerini kullanan belirli bir akıllı sistem türüdür. Çok çeşitli çevrimiçi alışveriş, e-ticaret hizmetleri ve sosyal ağ uygulamalarında kritik bir rol oynar. İşbirlikçi filtreleme, tavsiye eden sistemler için kullanılan en popüler yaklaşımlardır, ancak derecelendirme kaydının bulunmadığı tam soğuk başlatma (CCS) sorunundan ve yalnızca az sayıda derecelendirme kaydının bulunduğu eksik soğuk başlatma (ICS) sorunundan mustarıdır. Soğuk başlatma önerisi, yeni kullanıcılar veya öğeler için seyrek kullanıcı-öge etkileşimleri nedeniyle zorlu bir sorun olmuştur. Mevcut çabalar, çoğu soruna veri düzeyinde yaklaşan soğuk başlangıç sorununu bir dereceye kadar hafifletmiştir. Daha önceki yöntemler genellikle yardımcı verileri kullanıcı veya öge özellikleri olarak içerirken, daha yeni yöntemler, daha yüksek dereceli grafik yapıları aracılığıyla daha zengin anlambilim yakalamak için heterojen bilgi ağlarından yararlanmaktadır.

Soğuk başlatma, bir dereceye kadar otomatikleştirilmiş veri modellemesini içeren bilgisayar tabanlı bilgi sistemlerinde potansiyel bir sorundur. Özellikle, sistemin kullanıcılar veya hakkında henüz yeterli bilgi toplamadığı öğeler için herhangi bir çıkarımda bulunamaması sorunuyla ilgilidir.

Bir ÖS, kullanıcılara müzik, kitaplar gibi belirli öğeler için kişiselleştirilmiş öneriler sunmayı amaçlamaktadır. Popüler teknikler, içerik temelli modelleri ve İF

yaklaşımını içerir. Burada karşımıza çıkan çok önemli bir sorun soğuk başlatma sorunu. Bu sorun, yeni kullanıcılar veya yeni öğeler için önerilerle ilgilidir.

2. METOTLAR

2.1. SİSTEMİN TEMEL MODELİ

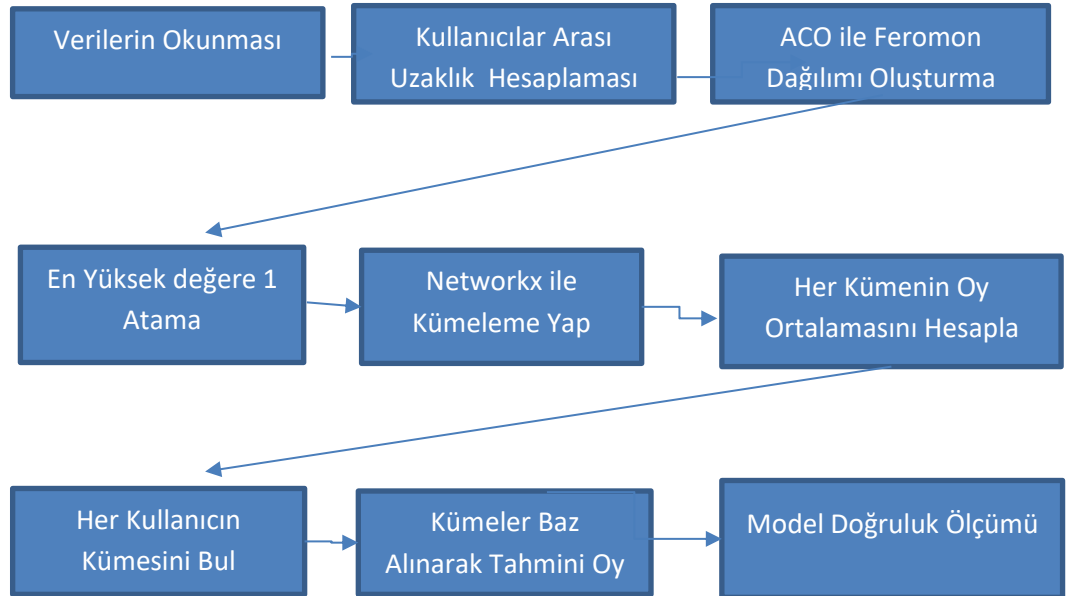
Klasik öneri sistemlerinde Pearson veya Öklid bağlantısı ile kullanıcı kullanıcı benzerliğine bakılır veya ögenin diğer ögeye benzerliğine bakılır. Daha önceki yöntemler genellikle yardımcı verileri kullanıcı veya öge özellikleri olarak içerirken daha yeni yöntemler ise daha yüksek dereceli grafik yapıları kullanıyor. Bu çalışmada önerilen model daha önce denenmemiş karma bir modeldir. Bu modelin üretim aşamasında denenen algoritmalar pearson korelasyonu, K – en yakın komşuluk, K-ortalama kümeleme, bağlı bileşen etiketleme, karınca optimizasyon algoritması, genetik algoritma ve yapay arı kolonisi algoritmasıdır. Çalışılan modelde pearson ile kullanıcıların ögelere verdikleri oylar üzerinde uzaklık mesafeleri hesaplandı. Tahmin mekanizmalarıyla birlikte yaygın olarak kullanımı bilinen sınıflandırma algoritmalarından K ortalama kümeleme algoritmasının yanı sıra bu alanda kullanılmamış olan karınca kolonisi optimizasyon algoritması (ACO), genetik algoritma ve yapay arı kolonisi algoritması gibi farklı algoritmalar araştırılmış ve ACO ile kümeleme yapılmıştır. Yeni kullanıcılar olması durumunda, tavsiyelerde bulunmak için sistem tercihleri hakkında bilgiye sahip değildir. Önerilen yaklaşım, sınıflandırma yöntemlerini saf bir işbirlikçi filtreleme sisteminde birleştirirken, kullanıcı özellikleri verilerinin kullanımı ile benzer davranışa sahip diğer kullanıcıların tanımlanmasına yardımcı olur. Bu sayede soğuk kullanıcı problemine de çözüm sunulmuştur. Bu kümelerdeki kullanıcıların birbirine benzediğini kabul edilip bunlar üzerinden tahminler yapılmaktadır. Önerilen modelde film önerisinde bulunmak için temel olarak 3 farklı türde tahmin yapılmaktadır. Bunlar klasik tahmin hesabı (Pearson, Öklid vb.), kullanıcı özelliklerini işin içine sokarak yapılan tahmin (yaş, cinsiyet vb.) ve kümeleme algoritmaları ile tahmin (karınca kolonisi algoritması, K ortalama kümeleme vb.).

Klasik oy tahmini şöyle yapılmaktadır: pearson hesabı ile kullanıcı – kullanıcı benzerliği ve öge – öge benzerlik oranları çıkmaktadır. En çok birbirine benzeyen kullanıcıların belirli bir ögeye vereceği oyların ortalaması ve kullanıcının kendi ortalamasının aritmetiksel ortalaması yeni oy tahminini verir.

Yeni modelde kümeleme algoritmaları ile oy tahmini şöyle yapılmaktadır; bkz. Şekil 3-1: pearson hesabı ile çıkan kullanıcı kullanıcı benzerlik oranını kümeleme algoritmaları ile kümeledikten sonra her kullanıcı kümesinin her öge için ortalamaları bulunur. Tahminde bulunurken kullanıcının kendi ortalaması ile bulunduğu kümenin bu ögeye verdiği oyların ortalaması hesaba katılarak yeni oy tahmin edilir.

Demografik modelde kullanıcı özellikleri ile oy tahmini şöyle yapılmaktadır: kullanıcılar yaşa ve cinsiyete göre kümelenir. Her kümenin ortalamaları hesaplanır. Tahminde bulunulacak kullanıcı hangi kümede ise o kümenin ortalaması tahmin olarak kaydedilir.

Çıkan bu sonuçların tümü için model ölçüm metrikleri ile doğruluk hesaplaması yapılarak model çalışması sonuçlandırılır.



Şekil 3-1 Sistemin temel modeli

2.2. KORELASYON VE BENZERLİK HESAPLAMALARI

Korelasyon, olasılık kuramı ve istatistikte iki rastsal deęişken arasındaki doğrusal ilişkinin yönünü ve gücünü belirtir. Genel istatistiksel kullanımda korelasyon, bağımsızlık durumundan ne kadar uzaklaştığını gösterir. İki nokta arasındaki uzaklık, bu iki noktanın birbiri ile olan ilişkisini bize verir. Farklı durumlar için farklı korelasyon katsayıları geliştirilmiştir.

2.2.1. Pearson Korelasyon Katsayısı

Rastgele gözlemler arasındaki bağımlılığı ölçmek, şüphesiz istatistikte merkezi bir rol oynar. Bağımlılıkları tam olarak anlamak ve tanımlamak çok zor olduğundan, genellikle bağımlılığın gücünü tek bir sayıya yoğunlaştırmakla ilgilenir. Klasik ve tartışmasız en popüler korelasyon katsayısı, sonlu ve pozitif varyanslı rastgele X ve Y deęişkenleri için tanımlanan katsayı Pearson katsayısıdır. Makine öğrenmesinde de, Pearson korelasyon katsayısı, çoklu veri deęişkenlerinin benzerliğini ölçmek için önemli bir yöntemdir. İki sayısal nokta arasındaki ilişkiyi, ilişkinin yönünü ve şiddetini belirlemek için kullanılır. Daha çok verilerin normal dağılım gösterdiği zamanlarda tercih edilir. Deęeri $[-1,1]$ arasındadır. Korelasyon katsayısı 1'e eşit olduğunda, tamamen pozitif korelasyon haline gelir; -1'e eşit olduğunda, tamamen negatif korelasyon haline gelir. Yani korelasyon katsayısının negatif olması sayılar arasındaki ters ilişkiyi gösterir iken, katsayının pozitif olması sayılar arasındaki doğru ilişkiyi anlatır. Bu, korelasyon katsayısının mutlak deęeri ne kadar büyükse, korelasyonun o kadar güçlü olduğu ve ne kadar küçük ise bunun tersi olduğu anlamına gelir. Verilerin normal dağılım göstermediğinde ise Spearman sıralama korelasyonu tercih edilir[27].

Denklem 3-1 Pearson korelasyon hesaplama.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

| | | |
|----------------|---|--|
| r | = | Korelasyon katsayısı |
| x _i | = | Örnekteki x değişkeninin değeri |
| \bar{x} | = | X değişkeninin değerlerinin ortalaması |
| y _i | = | Örnekteki y değişkeninin değeri |
| \bar{y} | = | Y değişkeninin değerlerinin ortalaması |

2.2.2. Öklid Uzaklığı

Öklid uzaklığı, eski Yunan matematikçiler Öklid ve Pisagor'un bulduğu bazı hesaplamalar üzerinden yapılmaktadır. Öklid mesafeleri sayı olarak temsil etmiyordu ve Pisagor teoreminden mesafe hesaplamasına bağlantı 18. yüzyıla kadar yapılmamıştı. Matematikte Öklid uzaklığı, iki nokta arasındaki bir doğru parçasının uzunluğudur. Pisagor teoremi kullanılarak noktaların Kartezyen koordinatlarından hesaplanabilir, bu nedenle bazen Pisagor mesafesi olarak adlandırılır. Nokta olmayan iki nesne arasındaki mesafe, genellikle iki nesneye olan nokta çiftleri arasındaki en küçük mesafe olarak tanımlanır. Formüller, bir noktadan bir çizgiye olan mesafe gibi farklı türdeki nesnelere arasındaki mesafeleri hesaplamak için bilinir. İleri matematikte, uzaklık kavramı soyut metrik uzaylara genelleştirilmiştir ve Öklid dışındaki diğer mesafeler incelenmiştir. İstatistik ve optimizasyondaki bazı uygulamalarda, mesafenin kendisi yerine Öklid mesafesinin karesi kullanılır. Öklid uzaklığının değerlendirilmesi, makine öğrenimi, iletişim, biyoinformatik vb. gibi birçok farklı mühendislik alanında gereklidir. Öklid uzaklığının hesaplanması, donanımda pahalı operatörler olan kareleme ve karekök gerektirir. Bu nedenle literatürde farklı uygulamalar için alternatif mesafeler tanıtılmıştır. Bu alternatifler arasında en popüler olanı Manhattan mesafesidir[28].

Denklem 3-2 Öklid hesaplama.

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

p,q = Öklid uzayındaki iki nokta

qi, pi = Öklid vektörleri (başlangıç noktası)

n = n uzay

2.2.3. K – En Yakın Komşuluk

İstatistikte, K-en yakın komşu algoritması (K-NN), ilk olarak Evelyn Fix ve Joseph Hodges tarafından 1951'de geliştirilen ve daha sonra Thomas Cover tarafından genişletilen parametrik olmayan bir sınıflandırma yöntemidir. Sınıflandırma ve regresyon için kullanılır. Her iki durumda da girdi, veri setindeki en yakın eğitim örneklerinden oluşur[29].

K-NN, fonksiyonun yalnızca yerel olarak tahmin edildiği ve tüm hesaplamaların fonksiyon değerlendirmesine kadar ertelendiği bir sınıflandırma türüdür. Bu algoritma, sınıflandırma için mesafeye dayandığından, özellikler farklı fiziksel birimleri temsil ediyorsa veya çok farklı ölçeklerde geliyorsa, eğitim verilerini normalleştirmek, doğruluğunu önemli ölçüde artırabilir[29].

K-NN, öneri sistemleri için en yaygın kullanılan algoritmalarından biridir. Bununla birlikte, tahmin edilen her konum için hata, algoritma için kullanılan K değerine bağlı olarak fark edilir şekilde değişir. Bu nedenle, K sabit bir değer ise, pozisyonlar için tahmin hatası daha fazla azaltılamaz[29].

2.3. K – ORTALAMA KÜMELEME

K-ortalımalı kümeleme algoritması, araştırma topluluğundaki en güçlü ve popüler veri madenciliği algoritmalarından biri olarak kabul edilir. Bununla birlikte, popüleritesine rağmen, algoritmanın, beklenmedik yakınsamaya yol açan merkezlerin rastgele başlatılmasıyla ilişkili sorunlar da dâhil olmak üzere belirli sınırlamaları vardır. Ek olarak, böyle bir kümeleme algoritması, farklı küme şekillerinden ve aykırı

değer etkilerinden sorumlu olan küme sayısının önceden tanımlanmasını gerektirir. K-ortalama algoritmasının temel bir sorunu, çeşitli veri türlerini işleyememesidir. K-ortalama algoritma k değerine bağlıdır. Herhangi bir kümeleme analizi gerçekleştirmek için her zaman belirtilmesi gerekir. Farklı k değerleriyle kümeleme, sonunda farklı sonuçlar üretecektir[30].

K-ortalama kümeleme, n gözlemi, her bir gözlemin en yakın ortalamaya (merkeze) ait olduğu ve kümenin bir örneği olarak hizmet ettiği k kümelerine bölmeyi amaçlayan bir yöntemdir. Bu kümeleme işlemi ile elde edilmek istenen, kümelerin birbirlerine benzerliklerinin minimum ve küme içindeki benzerliklerinin ise maksimum olmasını temin etmektir.

2.4. BAĞLI BİLEŞEN ETİKETLEME

Bağlı bileşen etiketleme (CCL – Connected Component Labeling) bir görüntüyü tarar ve piksellerini piksel bağlantısına göre bileşenlere gruplandırır, yani bağlı bir bileşendeki tüm pikseller benzer piksel yoğunluğu değerlerini paylaşır ve bir şekilde birbirine bağlıdır. Tüm gruplar belirlendikten sonra, her piksel, atandığı bileşene göre bir gri seviyesi veya bir renkle etiketlenir. Bir görüntüdeki çeşitli ayrık ve bağlantılı bileşenlerin çıkarılması ve etiketlenmesi, birçok otomatik görüntü analizi uygulamasının merkezinde yer alır[31]. Çalışmada küme oluşturma için kullanılmıştır.

Bağlı bileşen etiketleme, bağlı piksel bölgelerini, yani aynı yoğunluk değerlerini paylaşan bitişik piksel bölgelerini belirlemek için bir görüntüyü piksel piksel yukarıdan aşağıya ve soldan sağa tarayarak çalışır. Bağlı bileşen etiketleme, ikili veya gri düzey görüntülerde çalışır ve farklı bağlantı ölçüleri mümkündür[31].

2.5. SÜRÜ ZEKÂSI

Sürü zekâsı, kendi kendine organize olabilen etkileşimli ajanların veya sürülerin popülasyonunu modelleyen bir araştırma dalıdır. Bir karınca kolonisi, bir kuş sürüsü veya bir bağışıklık sistemi, bir sürü sisteminin tipik bir örneğidir. Arıların kovanlarının etrafında toplanması, sürü zekâsının bir başka örneğidir. Yapay Arı Kolonisi (ABC) Algoritması, bal arısı sürüsünün akıllı davranışına dayalı bir optimizasyon algoritmasıdır. Kombinasyonel ve sayısal optimizasyon problemlerini

çözmek için çeşitli modern sezgisel algoritmalar geliştirilmiştir. Bu algoritmalar, göz önünde bulundurulmuş kriterlere bağlı olarak popülasyon temelli, yinelemeli, stokastik, deterministik vb. gibi farklı gruplara ayrılabilir. Bir dizi çözümle çalışan ve bunları iyileştirmeye çalışan bir algoritmaya popülasyon temelli denir. Aranan çözüme yaklaşmak için birden çok yinelemenin kullanılması yinelemeli algoritma olarak adlandırılır. Bir algoritma, bir çözümü geliştirmek için olasılıklı bir kural kullanıyorsa, buna olasılık veya stokastik denir. Algoritma tarafından simüle edilen olgunun doğasına bağlı olarak başka bir sınıflandırma yapılabilir. Bu tür bir sınıflandırma esas olarak iki önemli popülasyon tabanlı algoritma grubuna sahiptir: evrimsel algoritmalar (EA) ve sürü zekâsı tabanlı algoritmalar. En popüler EA Genetik Algoritmadır (GA)[32].

Sürü terimi, genel bir şekilde, etkileşen ajanların veya bireylerin herhangi bir kısıtlanmış koleksiyonuna atıfta bulunmak için kullanılır. Bir sürünün klasik örneği, kovanlarının etrafında dolaşan arılardır, ancak metafor, benzer bir mimariye sahip diğer sistemlere kolayca genişletilebilir. Bir karınca kolonisi, tek tek etkenleri karıncalar olan bir sürü olarak düşünülebilir; bir kuş sürüsü, kuş sürüsüdür. Bir bağışıklık sistemi, bir hücre ve molekül sürüsüdür, bir kalabalık ise bir insan sürüsüdür.

2.5.1. Karınca Kolonisi Optimizasyon Algoritması

Karıncalar, uzun zamandır insanların dikkatini çeken karmaşık sosyal davranışlar sergilerler. Muhtemelen gözümüze çarpan en dikkat çekici davranışlardan biri, sözde karınca sokaklarının oluşmasıdır. Gençken, birçoğumuz böyle bir karınca otoyoluna adım atmış olabilir veya sadece karıncaların bu tür rahatsızlıklara nasıl tepki vereceğini görmek için yoluna bir engel koymuş olabiliriz. Bu karınca yollarının nereye gittiğini ve hatta nasıl oluştuğunu da merak etmiş olabiliriz. Bu tür sorular, yaşımız ilerledikçe ve üniversiteye giderken, bilgisayar bilimi, matematik vb. Gibi diğer konularda çalıştıkça çoğumuz için daha az acil hale gelebilir. Bununla birlikte, karıncaların davranışlarını ayrıntılı olarak inceleyen, başta biyologlar olmak üzere önemli sayıda araştırmacı vardır[33].

Pek çok karınca türünün görsel algılama yeteneği yalnızca ilkel olarak gelişmiştir ve tamamen kör olan karınca türleri vardır. Aslında, karıncaların

davranışları üzerine yapılan ilk araştırmaların önemli bir kavrayışı, bireyler arasındaki veya bireyler ile çevre arasındaki iletişimin çoğunun, karıncalar tarafından üretilen kimyasalların kullanımına dayandığıydı. Bu kimyasallara feromon denir. Bu, örneğin insanlarda ve en önemli duyuları görsel veya akustik olan diğer yüksek türlerde olanlardan farklıdır. Bazı karınca türlerinin sosyal yaşamı için özellikle önemli olan, iz feromonudur. Mesela yiyeceklerden gelen yolları işaretlemek için kullandıkları özel bir feromon salgırlar. Arayıcılar, feromon izlerini algılayarak, diğer karıncalar tarafından keşfedilen yiyeceklere giden yolu takip edebilirler. Bir karıncanın diğer karıncaların bıraktığı kimyasal bir izden etkilendiği bu toplu iz sürme davranışı, Karınca Optimizasyon Algoritmasının (ACO) ilham verici kaynağıdır[34]. Feromon izi buharlaşması, tüm karıncaların yetersiz bir yola hızla yaklaşmasını önleyen bir keşif mekanizması olarak görülebilir. Aslında, düşüş, ilmeklerin elimine edildiği sıraya bağlıdır.

Bugün, ACO'nun uygulamaları hakkında çok fazla sayıda çalışma yapılmıştır. Düzinelerce uygulama alanıyla gerçek bir sezgisel yöntemdir. Hem ACO algoritmalarının performansı hem de çalışmalarına ilişkin teorik anlayışımız önemli ölçüde artmış olsa da, şimdiye kadar sadece ön adımların atıldığını düşünerek çok daha fazla araştırmanın yapılması gerekir. ACO algoritmalarının, objektif fonksiyon değerleri, karar parametreleri veya kısıtlamalar gibi örnek verilerinin problemi çözerken değişebileceği dinamik problemleri içeren daha karmaşık optimizasyon problemlerine genişletilmektedir. Belirsizlik, gürültü, yaklaşıklık veya diğer faktörler nedeniyle objektif fonksiyon değerleri, karar değişken değerleri veya kısıtlama sınırları hakkında yalnızca olasılıksal bilgiye sahip olunan problemler (stochastic problems) işi daha zor hale getirmektedir[34].

Yapay bir karınca, Karınca kolonisi optimizasyon algoritmalarında belirli bir optimizasyon problemine iyi çözümler arayan basit bir hesaplama aracıdır. Bir karınca kolonisi algoritması uygulamak için optimizasyon probleminin ağırlıklı bir grafikte en kısa yolu bulma problemine dönüştürülmesi gerekir. Her yapay karınca yinelemenin ilk adımında grafikteki kenarların izlenmesi gereken stokastik olarak bir çözüm oluşturur. Stokastik; rastgele belirlenir, istatistiksel olarak analiz edilebilen ancak kesin olarak tahmin edilemeyen rastgele bir olasılık dağılımına veya modeline sahip olmak

anlamlarına gelir. İkinci adımda, farklı karıncaların bulduğu yollar karşılaştırılır. Son adım, her bir kenardaki feromon seviyelerinin güncellenmesinden oluşur.

Her yapay karınca, grafikte hareket etmek için bir yol bulmalıdır. Bir karınca bir sonraki kenarı seçmek için turunu atarken, mevcut konumundaki her kenarın uzunluğunun yanı sıra denk geldiği feromon seviyesini dikkate alacaktır. Her yapay karınca bu algoritmanın her adımında, bir durumdan hareket eder. Durumlar ve daha eksiksiz çözümler vardır. Bunlar her karınca gezisi için hesaplanır. Her karınca turu yani güzergâhı durumlar ve daha eksiksiz yeni çözümlere göre belirlenir. Bulunduğu durumdan yeni kenar seçimini karınca aşağıdaki formüle göre yapar.

Denklem 3-3 Yeni kenar seçimi.

$$p_{xy}^k = \frac{(T_{xy}^\alpha)(N_{xy}^\beta)}{\sum_{z \in allowed_x} (T_{xz}^\alpha)(N_{xz}^\beta)}$$

x = Karıncanın durumu

y = Daha iyi bir çözüm

k = Karınca

α = Txy etkisini kontrol etmek için bir parametredir

β = Nxy etkisini kontrol etmek için bir parametredir

Nxy = xy durum geçişinin istenirliği

Txy = x durumdan y'ye geçerken geçiş için biriktirilen feromon miktarı

p = İhtimal

İzler yani feromon miktarı tüm karıncalar çözümlerini tamamladığında güncellenir, daha iyi çözümlerden daha çok karınca gideceği için feromon miktarı artarken daha kötü çözümlere karıncalara gitmeyeceği için feromon buharlaşması meydana gelir. Global feromon güncelleme aşağıdaki şekilde hesaplanabilir:

Denklem 3-4 ACO feromon güncelleme.

$$T_{xy} \leftarrow (1 - \rho)T_{xy} + \sum_k^m \Delta T_{xy}$$

T_{xy} = Bir durum geçişi için biriktirilen feromon miktarıdır

ρ = Feromon buharlaşma katsayısı

m = Karıncaların sayısı

ΔT_{xy} = Karınca tarafından salgılanan feromon miktarı

k = Karınca

2.5.2. Genetik Algoritma

Genetik Algoritmalar, evrimden esinlenen bir hesaplama modelleri ailesidir. Bu algoritmalar, basit bir kromozom benzeri veri yapısı üzerindeki belirli bir soruna olası bir çözümü kodlar ve kritik bilgileri korumak için bu yapılara rekombinasyon operatörleri uygular. Genetik algoritmalar, uygulandığı problemlerin aralığı oldukça geniş olmasına rağmen, genellikle işlev iyileştiricileri olarak görülüyor. Genetik bir algoritmanın uygulanması, tipik olarak rastgele kromozomlardan oluşan bir popülasyonla başlar. Daha sonra bu yapılar değerlendirilir ve üreme fırsatları öyle bir şekilde tahsis edilir ki, hedef soruna daha iyi bir çözüm sunan kromozomlara, daha kötü çözümler olan kromozomlardan daha fazla "üreme" şansı verilir. Bu çözümün iyi tarafı tipik olarak mevcut nüfusa göre tanımlanır. Bir genetik algoritmanın bu özel açıklaması kasıtlı olarak soyuttur çünkü bir anlamda genetik algoritma teriminin iki anlamı vardır. 1960'ların başından beri bu ilkeleri optimizasyon rutinlerine dahil etmek için ara sıra girişimlerde bulunulmasına rağmen, GA'lar ilk olarak Holland tarafından sağlam bir teorik temelde oluşturulmuştur. İlk olarak John Holland ve onun öğrencileri tarafından tanıtılan ve araştırılan bir modeli ifade eder. Genetik algoritmalar için mevcut teorinin çoğu, ya sadece ya da esas olarak Holland tarafından sunulan modele ve kanonik genetik algoritma olarak anılan varyasyonlara uygulanmıştır. Genetik algoritmaların modellenmesindeki son teorik ilerlemeler, temel olarak kanonik genetik algoritma için de geçerlidir. Terimin daha geniş bir kullanımında, bir genetik algoritma, bir arama alanında yeni numune noktaları oluşturmak için seçim ve

rekombinasyon operatörlerini kullanan herhangi bir popülasyon tabanlı modeldir. Birçok genetik algoritma modeli, araştırmacılar tarafından büyük ölçüde deneysel bir bakış açısıyla tanıtılmıştır. Bu araştırmacıların çoğu uygulama odaklıdır ve genellikle optimizasyon araçları olarak genetik algoritmalarla ilgilenirler[35].

GA, doğal evrim fenomenini simüle etmeye çalışır. Doğal evrimde her tür, sürekli değişen bir ortamda faydalı adaptasyonlar arar. Türler geliştikçe, yeni özellikler bireysel üyelerin kromozomlarında kodlanır. Bu bilgi rastgele mutasyonla değişir, ancak evrimsel gelişimin arkasındaki gerçek itici güç, üreme sırasında kromozomla ilgili materyallerin kombinasyonu ve değişimidir.

2.5.3. Yapay Arı Kolonisi Algoritması

Bal arısı sürülerinin belirli zeki davranışlarını modellemek için çeşitli yaklaşımlar önerilmiş ve kombinasyonel tip problemleri çözmek için uygulanmıştır. Tereshko ve Loengarov, bir arı kolonisini, bir ortamdan bilgi toplayan ve davranışını buna göre ayarlayan dinamik bir sistem olarak görüyordu. Arıların yiyecek arama davranışları üzerine robotik bir fikir geliştirdiler[36]. Genellikle, tüm bu robotlar fiziksel ve işlevsel olarak aynıdır, böylece herhangi bir robot rastgele bir şekilde diğerleri ile değiştirilebilir. Sürü, önemli bir toleransa sahiptir; tek bir aracıdaki başarısızlık tüm sistemin performansını durdurmaz. Böcekler gibi bireysel robotlar sınırlı yeteneklere ve çevre hakkında sınırlı bilgiye sahiptir. Öte yandan sürü, kolektif zekâ geliştirir. Deneyler böcek benzeri robotların gerçek robotik görevlerde başarılı olduğunu gösterdi. Ayrıca, üç temel bileşenden oluşan kolektif zekânın ortaya çıkmasına yol açan minimal bir yem seçimi modeli geliştirdiler: gıda kaynakları, istihdam edilmiş toplayıcılar ve işsiz toplayıcılar. Bu model, davranışın iki ana yöntemini tanımlar: bir nektar kaynağına yetiştirme ve bir kaynaktan vazgeçme[36]. Teodorovic, trafik ve ulaşımdaki karmaşık sorunları çözmeyi amaçlayan yapay sistemlerin geliştirilmesinde arı sürüsü zekasını kullanmayı önerdi. Teodorovic ayrıca belirleyici kombinasyonel problemleri ve belirsizlikle karakterize kombinasyonel problemleri çözebilen Meta-Sezgisel Arı Kolonisi Optimizasyonunu (BCO) önerdi[37]. Drias v.d. gerçek arıların davranışlarından esinlenen ve bunu maksimum ağırlıklı doyma (max-sat) probleminin özelliklerine uyarlayan yeni bir akıllı yaklaşım veya meta-sezgisel olan Arı Kolonisi Optimizasyonu (BSO) tanıttı[38]. Benzer

şekilde, Benatchba v.d. arıların üreme sürecine dayalı 3-sat problemini çözmek için bir meta-sezgisel yöntem geliştirdi[39]. Wedde v.d. bal arılarının iletişimsel ve değerlendirme yöntemlerinden ve prosedürlerinden esinlenen BeeHive adlı yeni bir yönlendirme algoritması sundu. BeeHive algoritmasında, arı ajanları yiyecek arama bölgeleri adı verilen ağ bölgeleri boyunca seyahat eder. Yolda, ağ durumu hakkındaki bilgileri yerel yönlendirme tablolarını güncellemek için teslim edilir[40].

Bütün bu çalışmalar, kombinasyonel problem türlerini içerir. Literatürde bal arısı kolonisinin akıllı davranışına dayanan tek bir sayısal optimizasyon algoritması vardır. Yang, sayısal fonksiyon optimizasyonlarını çözmek için sanal bir arı algoritması (VBA) geliştirdi. İki parametrelili işlevler için, bir sanal arı sürüsü oluşturulur ve sürü, faz uzayında rastgele hareket etmeye başlar. Bu arılar, işlevin kodlanmış değerlerine karşılık gelen bazı hedef nektar bulduklarında etkileşime girerler. Optimizasyon probleminin çözümü arı etkileşimlerinin yoğunluğundan elde edilebilir[41]. Çok değişkenli fonksiyonları optimize etmek için Karaboğa, sanal arı algoritmasından farklı bir yapay arı kolonisi (ABC) algoritması tanımladı. Baştürk ve Karaboğa, ABC algoritmasının performansını genetik algoritmanın performansı ile karşılaştırdı[42].

2.6. ACO İLE KÜMELEME VE TAHMİN YÖNTEMLERİ

Keşif ekibindeki karıncalar izledikleri yollar ile arkalarından gelenler için koku bırakırlar. En çok yiyeceğin olduğu yere daha fazla karınca gelir ve daha fazla koku birikir. Böylece kümeleme oluşur. Bu kümelemeler ile kullanıcılar arasında iyi bir korelasyon kurulduğunda iyi tahminler yapılabilir.

2.6.1. ACO Algoritması ile Kümeleme

Karınca kolonisi optimizasyon algoritmalarında, yapay bir karınca, belirli bir optimizasyon problemine iyi çözümler arayan basit bir hesaplama aracıdır. Bir karınca kolonisi algoritması uygulamak için en uygun şekilde sokma probleminin ağırlıklı bir grafikte en kısa yolu bulma problemine dönüştürülmesi gerekir. Derecelendirme verilerinde kullanıcı ve öge verileri mevcuttur. Bu veriler üzerinden pearson korelasyonu ile uzaklık hesabı yapılmıştır. Bu uzaklık katsayılarını karınca kolonisi

algoritmasına girdi olarak verilir. Hesaplanan uzaklıklar ACO algoritmasındaki yol uzunluğu gibi kullanılarak bu yollar üzerinde karıncaların koku biriktirmeleri sağlanmıştır. Karınca algoritmasında belirli sayıda karınca belirli yineleme ile çalıştırılır. Her yinelemenin ilk adımında, her karınca rasgele olarak bir çözüm oluşturur, yani grafikteki kenarların izlenmesi gereken sıra. İkinci adımda, farklı karıncaların bulduğu yollar karşılaştırılır. Son adım, her bir kenardaki feromon seviyelerinin güncellenmesinden oluşur. Karıncalar yola çıktıklarında en çok kokunun biriktiği yerlerde gezmeye başlarlar. Böylece karıncalar kümeler halinde belirli güzergâhları takip ederler.

2.6.2. ACO ile Kümeleme Uygulaması

Karıncaların yoğun olarak izlediği yollarda daha fazla koku birikir. Belirli güzergâhlarda en fazla koku nerede birikti ise o referans alınır, diğer kokular silinir. Az koku biriken kenarlar ihmal edildiğinde geride kalan güzergâhlar kümelenmiş kullanıcı grupları olur. Böyle bir ayırımdan sonra toplam kümeleme sayısı makul bir değere düşer (dört ile on arası). Böylece her bir kullanıcı bu kümelerden birine dâhil edilir. Her kümedeki kullanıcıların bir ögeye verdiği oyların ortalaması kaydedilir. Bu bütün ögeler için tekrarlanır. Böylece öge sayısı ve küme sayısı matrisi elde edilir. Bir kullanıcı için tahmin yapılacağına, kullanıcının hangi kümede olduğu bulunur. Sonra o kümenin o öge için derecelendirme ortalaması tahmin olarak yazılır.

2.7. ACO İLE BİRİKEN KOKULARIN KMEANS KÜMELEME YÖNTEMİYLE BİRLİKTE KULLANIMI

Belirli sayıda karınca belirli yineleme ile gezdikten sonra arkalarında kokular birikir. Bu kokular K - means algoritmasına girdi olarak verilir. K - means algoritması ile kümeleme yapılır. Bu kümeleme sonucunda kullanıcılar belirli kümelere dâhil edilir. Eldeki veriler üzerinden her kümenin her öge için ortalaması alınır. Böylece küme sayısı ve öge sayısı matrisi oluşur. Yeni yapılacak tahminde bu matris kullanılır. Örneğin bir kullanıcının bir ögeye vereceği derecelendirme tahmini yapıldığında, kullanıcı hangi kümede ise o kümenin o ögeye verdiği derecelendirmenin ortalamasını kullanıcının derecelendirmesi olarak yazılır.

3. BULGULAR

3.1. VERİ HAZIRLAMA VE ÖN İŞLEME SÜREÇLERİ

MovieLens veri setleri, Minnesota Üniversitesi'ndeki GroupLens Araştırma Projesi tarafından toplandı[43].

Bu veri seti şunlardan oluşur:

* 1682 filmde 943 kullanıcıdan 100.000 derecelendirme (1-5).

* Her kullanıcı en az 20 filme oy vermiştir.

* Kullanıcılar için basit demografik bilgiler (yaş, cinsiyet, meslek, posta kodu)

Veriler, 19 Eylül 1997'den 22 Nisan 1998'e kadar olan yedi aylık dönemde MovieLens web sitesi (movielens.umn.edu) aracılığıyla toplandı. Bu veriler temizlendi - 20'den az derecelendirmeye sahip olan veya olmayan kullanıcılar bu veri setinden tam demografik bilgiler çıkarıldı.

İşte verilerin kısa açıklamaları.

ml-data.tar.gz - Sıkıştırılmış tar dosyası. U veri dosyalarını yeniden oluşturmak için aşağıdaki işlemi yapın:

```
gunzip ml-data.tar.gz
```

```
tar xvf ml-veri.tar
```

```
mku.sh
```

u.data - Tam u veri kümesi, 1682 öge üzerinde 943 kullanıcı tarafından 100.000 değerlendirme.

Her kullanıcı en az 20 filme oy vermiştir. Kullanıcılar ve ögeler, 1'den ardışık olarak numaralandırılır. Veriler rastgele sıralanır. Bu, kullanıcı kimliğinin

sekmeye ayrılmış bir listesidir | öge kimliği | derecelendirme | zaman damgası. Zaman damgaları 1/1/1970 UTC'den bu yana unix saniyedir.

u.info - u veri kümesindeki kullanıcıların, ögelerin ve derecelendirmelerin sayısı.

u.item - Ögeler (filmler) hakkında bilgiler; bu sekmeye ayrılmış bir listedir

Film Id | Film Adı | Çıkış Tarihi | Video Çıkış Tarihi | IMDb URL'i | Bilinmeyen | Eylem | Macera | Animasyon | Çocuk | Komedi | Suç | Belgesel | Dram | Fantezi | Kara Film | Korku | Müzikal | Gizem | Romantizm | Bilim Kurgu | Gerilim | Savaş | Batı |

Son 19 alan türlerdir, "1" filmin o türe ait olduğunu, "0" olmadığını belirtir; filmler aynı anda birkaç türde olabilir. Film Id u.data veri kümesinde kullanılanlardır.

u.genre - Türlerin listesi.

u.user - Kullanıcılar hakkında demografik bilgiler; bu, kullanıcı detaylarının sekmeye ayrılmış bir listesidir | yaş | cinsiyet | meslek | posta kodu

Kullanıcı id, u.data veri setinde kullanılanlardır.

u1.base - Veri, u1.base ve u1.test ile u5.base ve u5.test arasında ayarlar.

u1.test, u verilerinin eğitim ve test verilerine % 80 / % 20 bölünmesidir. U1, ..., u5'ten her birinin ayrık test setleri vardır; bu, 5 kat çapraz doğrulama için ise deneyinizi her eğitim ve test seti ile tekrarlar ve sonuçların ortalamasını alırsınız. Bu veri setleri u.data'dan mku.sh tarafından oluşturulabilir. (u1.test, u2.base, u2.test, u3.base, u3.test, u4.base, u4.test, u5.base, u5.test)

ua.base, ua.test, ub.base, ub.test - Ua.base, ua.test, ub.base ve ub.test veri kümeleri, u verilerini bir eğitim setine ve test setinde kullanıcı başına tam olarak 10 derecelendirmeye sahip bir test setine böler. Ua.test ve ub.test kümeleri ayrıktır. Bu veri setleri u.data'dan mku.sh tarafından oluşturulabilir.

allbut.pl - Bir kullanıcı derecelendirmesinin n'si hariç tümünün eğitim verilerinde bulunduğu eğitim ve test kümelerini oluşturan komut dosyası.

mku.sh - u.data'dan tüm u veri kümelerini oluşturmak için bir komut dosyası.

3.2. MODEL ÖLÇÜM METRİKLERİ

Son yirmi yılda, çalışılan modellerin sayısı ve türlerinde bir artış oldu. Hangi formüller ile yapılan çalışmalar ile ilgili değişkenlerin daha doğru ve kesin tahminlerini ürettiğini belirlemeye olan ilgi de artmıştır; buna karşılık, model tarafından üretilen tahminleri bağımsız, güvenilir gözlemlerle karşılaştırmak için kullanılabilen hata istatistikleri daha geniş bir şekilde uygulanmıştır. İki veya daha fazla karşılaştırılabilir tahmin seti arasındaki farklılıkları değerlendirmeye olan ilgi, hangi setin en güvenilir olduğu bilinmediğinde daha fazla artmıştır ve bu da hata veya fark istatistiklerinin uygulanmasını artırma eğiliminde olmuştur. Bu ölçümlerin her biri, ilgilenilen değişkenin birimlerindeki ortalama model-tahmin hatasını ifade etmesi açısından çeşit çeşittir. Bu ölçümler ayrıca, herhangi bir tahmin kümesinin güvenilir olduğu bilinmediğinde ortalama farkı ortalama hata yerine temsil etmek için kullanılmıştır. Bu yöntemler yıllardır model performansını değerlendirmek için kullanılmış olsa da model hataları için en uygun metrik üzerinde fikir birliği yoktur[44]. Biz kendi modelimizin doğruluğunu teyit etmek için “Karesel Ortalama Hata”, “Kök Ortalama Kare Hatası” ve “Ortalama Mutlak Hata” metriklerini kullandık. Şimdi bunları kısaca açıklayalım.

3.2.1. Karesel Ortalama Hata

İstatistikte bir tahmincinin gözlenmeyen bir miktarı tahmin etme prosedürünün karesel ortalama hatası (Mean Square Error – MSE), hataların karelerinin ortalamasını ölçer, yani tahmin edilen değerler ile gerçek değer arasındaki ortalama kare farkıdır. MSE kare hata kaybının beklenen değerine karşılık gelen bir risk fonksiyonudur. MSE'nin neredeyse her zaman kesin olarak pozitif olması ve sıfır olmaması, rastlantısallıktan veya tahmin edicinin daha doğru bir tahmin üretebilecek bilgileri hesaba katmamasından kaynaklanmaktadır. MSE, tahmincinin kalitesinin bir ölçüsüdür, her zaman pozitif bir değerdir ve sıfıra yakın değerler daha iyidir[45].

Denklem 4-1 Karesel ortalama hata.

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - X_i)^2$$

MSE = Karesel Ortalama Hata

Y_i = Gerçek değer

X_i = Tahmin edilen değer

n = Toplam sayı

3.2.2. Kök Ortalama Kare Hatası

Kök Ortalama Kare Hatası (RMSE), meteoroloji, hava kalitesi ve iklim araştırma çalışmalarında model performansını ölçmek için standart bir istatistiksel ölçümdür. Karesel hata toplamının örnek sayısına bölünerek ortalamasının alınmasına dayanır.

Denklem 4-2 Kök ortalama kare hatası.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Y_i - X_i)^2}$$

RMSE = Kök Ortalama Kare Hatası

Y_i = Gerçek değer

X_i = Tahmin edilen değer

n = Toplam sayı

i = Değişken sayı

3.2.3. Ortalama Mutlak Hata (MAE)

Ortalama Mutlak Hata (MAE), model değerlendirmelerinde yaygın olarak kullanılan başka bir yararlı ölçüdür. İstatistikte ortalama mutlak hata, aynı olguyu ifade eden eşleştirilmiş gözlemler arasındaki hataların bir ölçüsüdür. Hataların mutlak değerinin aritmetik ortalamasıdır. Burada amaç iki sürekli değişkenin arasındaki farkı ölçmektir. Bir liste ile diğer örnekleri, tahmin edilenle eldeki veriyi, sonraki zamanla

ilk zaman arasındaki karşılaştırmaları şeklinde bir ölçüm tekniğini içerir. Ortalama mutlak hata aşağıdaki şekilde hesaplanır. Tahmin edilen değer ile gerçek değer arasındaki farkın mutlak değeri alınır. Bütün tahmin edilen ve gerçek değerler için bu işlem tekrar edilerek, bütün sayılar toplanır. Toplam değeri toplam tahmin yaptığımız sayıya böldüğümüzde elde ettiğimiz sayı Ortalama Mutlak Değer'imizdir.

Denklem 4-3 Ortalama mutlak hata.

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n}$$

MAE = Ortalama Mutlak Hata

y_i = Tahmin edilen değer

x_i = Gerçek değer

n = Toplam sayı

i = Değişken sayı

3.3. DOĞRULANMIŞ MODEL SONUÇ KARŞILAŞTIRMALARI

Çalışmanın bu bölümünde doğrulanmış modelin sonuç karşılaştırmalarını gözle rahatça kıyaslayabilmek için çizelgede gösterilecektir. Çizelge 4-1'de görüldüğü üzere hibrit öneri sistemi sonuçları klasik öneri sistemine göre daha iyi sonuçlar vermektedir.

Çizelge 4-1 Seçilmiş 3000 tane derecelendirme için tahmin sonuçları.

| | Pearson | K - Means | ACO | ACO + K - Means | Yaşa Göre | Cinsiyete Göre |
|------|---------|-----------|-------|-----------------|-----------|----------------|
| MSE | 14.45 | 0.97 | 12.42 | 0.80 | 1.47 | 1.47 |
| RMSE | 3.80 | 0.99 | 3.52 | 0.89 | 1.21 | 1.21 |
| MAE | 3.70 | 0.80 | 3.36 | 0.67 | 1.05 | 1.05 |

Çizelge 4-2'de Movielens verilerinden 90570 derecelendirme kullanılmıştır. Bunların % 20 si rasgele olarak seçilmiştir. Daha sonra bu seçilen derecelendirmeler

için 6 farklı yöntem ile oy tahmini yapılmıştır. Elde edilen tahminler daha önce elde mevcut olup seçilen veriler ile karşılaştırılarak doğruluk oranları hesaplanmıştır. Sonuçları tekrar test etmek için bu işlem beş defa tekrar edilmiştir. Beş işlemin sonuçlarının aritmetik ortalaması alınmıştır.

Çizelge 4-2 Rasgele %20'si seçilerek yapılan derecelendirmede eğitim seti için tahmin sonuçları.

| | Pearson | K - Means | ACO | ACO + K - Means | Yaşa Göre | Cinsiyete Göre |
|------|---------|-----------|-------|-----------------|-----------|----------------|
| MSE | 13.48 | 0.74 | 11.32 | 0.73 | 0.75 | 0.85 |
| RMSE | 3.67 | 0.86 | 3.15 | 0.85 | 0.86 | 0.92 |
| MAE | 3.49 | 0.69 | 2.23 | 0.64 | 0.66 | 0.72 |

Çizelge 4-3'de Movielens verilerinden 100000 derecelendirme kullanılmıştır. Veriler %80 eğitim %20 test olacak şekilde Movielens tarafından ayrılmıştır. Movielens bu amaçla beş ayrı eğitim ve test seti sağlamıştır. ACO sonuçları 20 karınca 10 yinleme ile elde edilmiştir. Eğitim seti ile tavsiye sisteminin eğitilmesi sağlanmıştır. Eğitilen tavsiye sisteminden daha sonra bu test verisi için 6 farklı yöntem ile oy tahmini yapılması istenmiştir. Elde edilen tahminler daha önce elde mevcut olup seçilen veriler ile karşılaştırılarak doğruluk oranları hesaplanmıştır. Aynı işlem 5 ayrı eğitim ve test seti için tekrarlanmıştır.

Çizelge 4-3 Eğitim seti u1.base test seti u1.test için tahmin sonuçları.

| | Pearson | K - Means | ACO | ACO + K - Means | Yaşa Göre | Cinsiyete Göre |
|------|---------|-----------|------|-----------------|-----------|----------------|
| MSE | 13.83 | 1.15 | 3.85 | 1.11 | 1.20 | 1.12 |
| RMSE | 3.71 | 1.07 | 1.96 | 1.05 | 1.09 | 1.06 |
| MAE | 3.53 | 0.85 | 1.47 | 0.83 | 0.85 | 0.84 |

Çizelge 4-4 Eğitim seti u2.base test seti u2.test için tahmin sonuçları.

| | Pearson | K - Means | ACO | ACO + K - Means | Yaşa Göre | Cinsiyete Göre |
|------|---------|-----------|------|-----------------|--------------|-------------------|
| MSE | 13.83 | 1.12 | 6.46 | 1.10 | 1.18 | 1.10 |
| RMSE | 3.71 | 1.06 | 2.54 | 1.04 | 1.09 | 1.05 |
| MAE | 3.54 | 0.83 | 2.11 | 0.83 | 0.85 | 0.83 |

Çizelge 4-4’de ikinci eğitim ve test seti kullanılmıştır. Birinci eğitim ve test setindeki sonuçlara yakın ve paralel sonuçlar çıkmıştır.

Çizelge 4-5 Eğitim seti u3.base test seti u3.test için tahmin sonuçları.

| | Pearson | K - Means | ACO | ACO + K - Means | Yaşa Göre | Cinsiyete Göre |
|------|---------|-----------|------|-----------------|--------------|-------------------|
| MSE | 13.64 | 1.11 | 6.84 | 1.06 | 1.16 | 1.07 |
| RMSE | 3.69 | 1.05 | 2.61 | 1.03 | 1.08 | 1.03 |
| MAE | 3.52 | 0.84 | 2.21 | 0.82 | 0.84 | 0.83 |

Çizelge 4-5’de üçüncü eğitim ve test seti kullanılmıştır. Sonuçlar ilk iki eğitim ve test setindeki sonuçlara yakın ve paralel sonuçlar içermektedir.

Çizelge 4-6 Eğitim seti u4.base test seti u4.test için tahmin sonuçları.

| | Pearson | K - Means | ACO | ACO + K - Means | Yaşa Göre | Cinsiyete Göre |
|------|---------|-----------|------|-----------------|--------------|-------------------|
| MSE | 13.65 | 1.09 | 7.22 | 1.06 | 1.17 | 1.09 |
| RMSE | 3.69 | 1.04 | 2.68 | 1.03 | 1.08 | 1.04 |
| MAE | 3.52 | 0.83 | 2.32 | 0.82 | 0.85 | 0.83 |

Çizelge 4-6’da dördüncü eğitim ve test seti kullanılmıştır. Sonuçlar ilk üç eğitim ve test setindeki sonuçlar ile yakın ve paralel sonuçlar içermektedir.

Çizelge 4-7 Eğitim seti u5.base; test seti u5.test için tahmin sonuçları.

| | Pearson | K - Means | ACO | ACO + K - Means | Yaşa Göre | Cinsiyete Göre |
|------|---------|-----------|------|-----------------|--------------|-------------------|
| MSE | 13.66 | 1.10 | 4.48 | 1.09 | 1.17 | 1.09 |
| RMSE | 3.69 | 1.05 | 2.11 | 1.04 | 1.08 | 1.04 |
| MAE | 3.52 | 0.83 | 1.61 | 0.82 | 0.85 | 0.83 |

Çizelge 4-7’de beşinci eğitim ve test seti kullanılmıştır. Sonuçlar ilk dört eğitim ve test setlerindeki sonuçlar ile yakın ve paralel sonuçlar içermektedir. Farklı farklı veriler ile yapılan testlerin sonuçlarının yakın ve paralel olması, sistemin iyi sonuçlar verdiğinin bir göstergesidir.

Çizelge 4-8 Eğitim seti u1-5.base; test seti u1-5.test için tahmin sonuçlarının ortalaması.

| | Pearson | K - Means | ACO | ACO + K - Means | Yaşa Göre | Cinsiyete Göre |
|------|---------|-----------|------|-----------------|-----------|----------------|
| MSE | 13.72 | 1.11 | 5.77 | 1.08 | 1.18 | 1.09 |
| RMSE | 3.698 | 1.05 | 2.38 | 1.04 | 1.08 | 1.04 |
| MAE | 3.526 | 0.84 | 1.94 | 0.82 | 0.85 | 0.83 |

Çizelge 4-8’de MovieLens tarafından çapraz doğrulama kıstaslarına göre 5 eğitim, 5 test olarak ayrılmış dosyalar ile yapılan testlerin ortalaması verilmiştir.

Çizelge 4-9 Eğitim seti u5.base; test seti u5.test için tahmin sonuçlarının standart sapmalarının ortalaması.

| | Pearson | K - Means | ACO | ACO + K - Means | Yaşa Göre | Cinsiyete Göre |
|------|---------|-----------|-------|-----------------|-----------|----------------|
| MSE | 0.087 | 0.024 | 1.51 | 0.021 | 0.015 | 0.019 |
| RMSE | 0.014 | 0.011 | 0.32 | 0.008 | 0.005 | 0.011 |
| MAE | 0.009 | 0.012 | 0.379 | 0.005 | 0.004 | 0.008 |

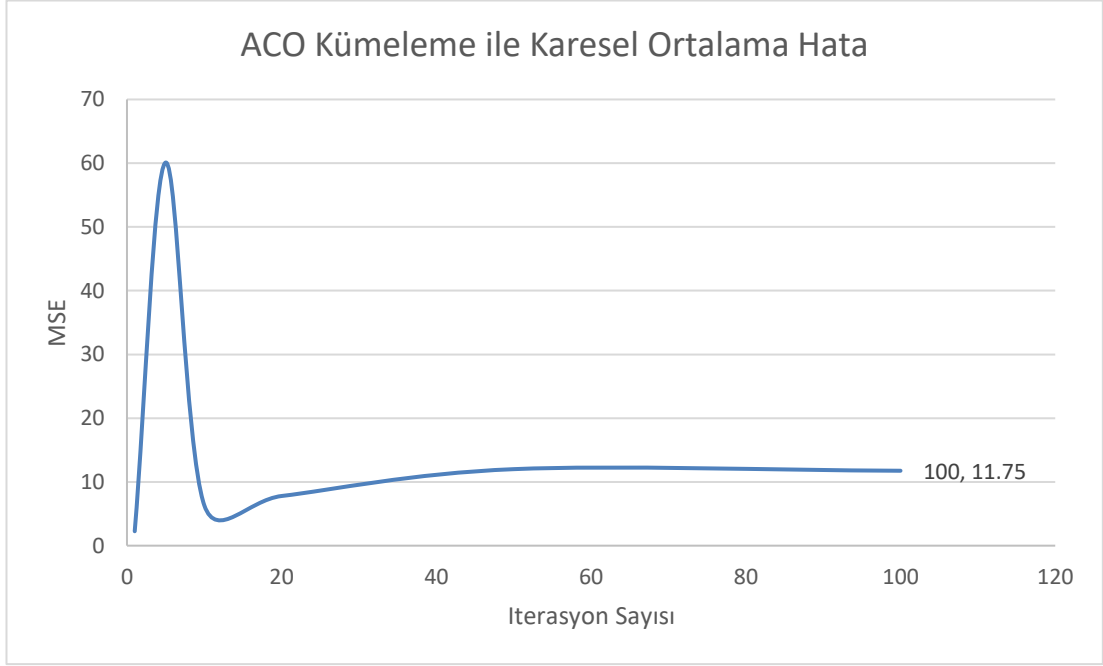
Çizelge 4-9 eğitim ve test setleri ile elde edilen tahmin sonuçlarının ortalamalarının standart sapmalarını göstermektedir. Sadece ACO ile yapılan testlerde standart sapmanın yüksek olduğu göze çarpmaktadır. ACO + K – Means ile yapılan testlerin sonuçları istikrarlı olduğu için standart sapmaları iyi çıkmıştır.

Pearson Korelasyon Katsayısı ile derecelendirme tahmini yapma: kullanıcıların pearson hesabı ile birbirlerine benzerlikleri hesaplanmıştır. Benzer kullanıcıların aynı oyu vereceği tahmin edilerek oy verilmemiş öğelere kullanıcılar için oy tahmini yapılmıştır. Kullanıcının kendisine en yakın beş kullanıcı seçilmiştir.

Bu kullanıcıların tahmin edilecek ögeye verdikleri oyların aritmetik ortalaması alınıp bu kullanıcı için tahmin olarak konulmuştur. Yapılan testlerde görülmüştür ki sadece klasik pearson ile yapılan oy tahminlerinde bile yaklaşık 15 hata payı değeri ile doğru tahminleme yapılabilmektedir.

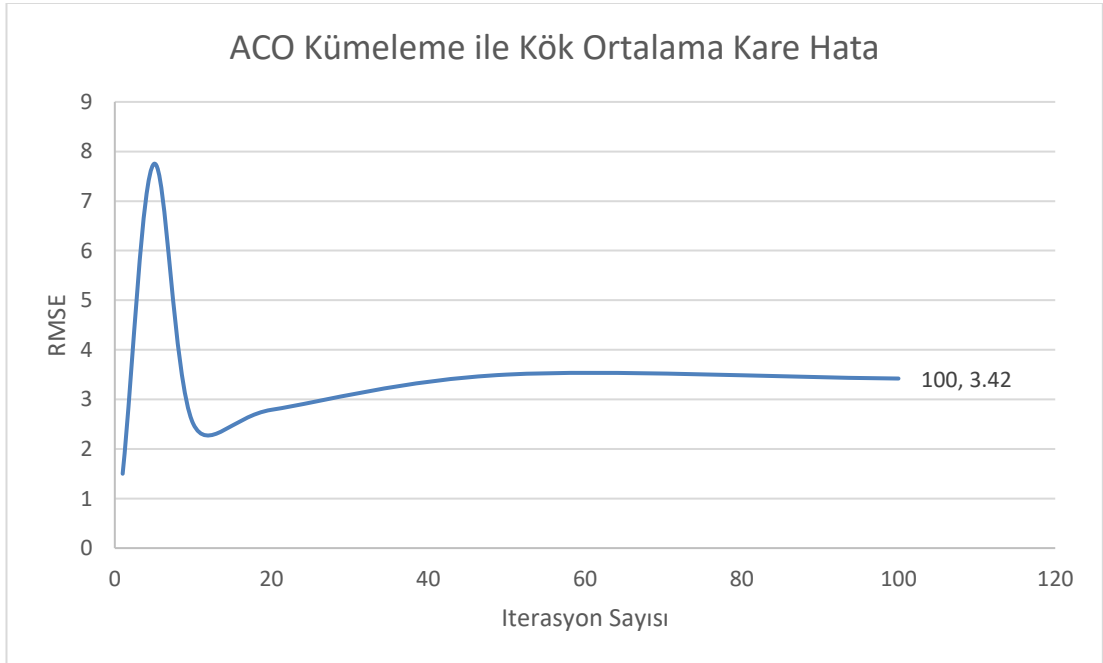
K – Ortalama Kümeleme ile derecelendirme tahmini yapma: kullanıcıların pearson hesabı ile birbirlerine benzerlikleri hesaplanmıştır. Hesaplanan benzerlikler üzerinden beş merkezli bir kümeleme yapılmıştır. Her kullanıcının ait olduğu küme belirlenmiştir. Bu kümelerin her öge için ortalamaları hesaplanmıştır. Daha önce seçilerek silinen derecelendirmeleri tahmin etmek için kullanıcının kendi ortalaması ve K – ortalama kümeleme ile bulunan küme ortalamasının aritmetik ortalaması alınıp ögeye oy olarak verilmiştir. Sonuçlar karesel ortalama hata, kök ortalama hatası ve ortalama mutlak hata metrikleri ile değerlendirilmiştir. 1'in altında bir hata değeri ile doğru tahminleme yapılabilmektedir.

Karınca Kolonisi Optimizasyon Algoritmasını kümeleme için kullanarak derecelendirme tahmini yapma: kullanıcıların pearson hesabı ile birbirlerine benzerlikleri hesaplanmıştır. Hesaplanan benzerlikler ACO algoritmasındaki yol uzunluğu gibi kullanılarak bu yollar üzerinde karıncaların koku biriktirmeleri sağlanmıştır. Güzergâh üzerinde en fazla kokunun biriktiği yol korunup diğer yollar silinmiştir. Karıncaların yüksek kokunun biriktiği yolları takip etmesi istenmiştir. Bu yöntem ile K – ortalama kümeleme algoritmasında önceden atanan atanmış kümeleme problemi giderilmiştir. K – ortalama kümelemede beş olarak atadığımız küme sayısını burada biriken kokular belirlemiştir. ACO sezgisel bir yöntem olduğu için 4 – 8 arasında değişen rasgele sayıda küme oluşmuştur. Bu yöntem K ortalama kümeleme yöntemine göre küme sayısının otomatik belirleme yönünden daha avantajlıdır. Seçilerek silinen oylar bu kümeleme yöntemi ile tahmin edilmiştir. Her bir öge için kümenin ortalaması ve kendi oy ortalamasının aritmetik ortalaması tahmin olarak kaydedilmiştir. Sonuçlar karesel ortalama hata, kök ortalama hatası ve ortalama mutlak hata metrikleri ile değerlendirilmiştir. Testler 10 karınca ile yapılmıştır. Aşağıdaki grafik karınca kolonisi optimizasyon algoritmasındaki yineleme sayısını arttırarak aldığımız karesel ortalama hata bilgilerini yansıtmaktadır.



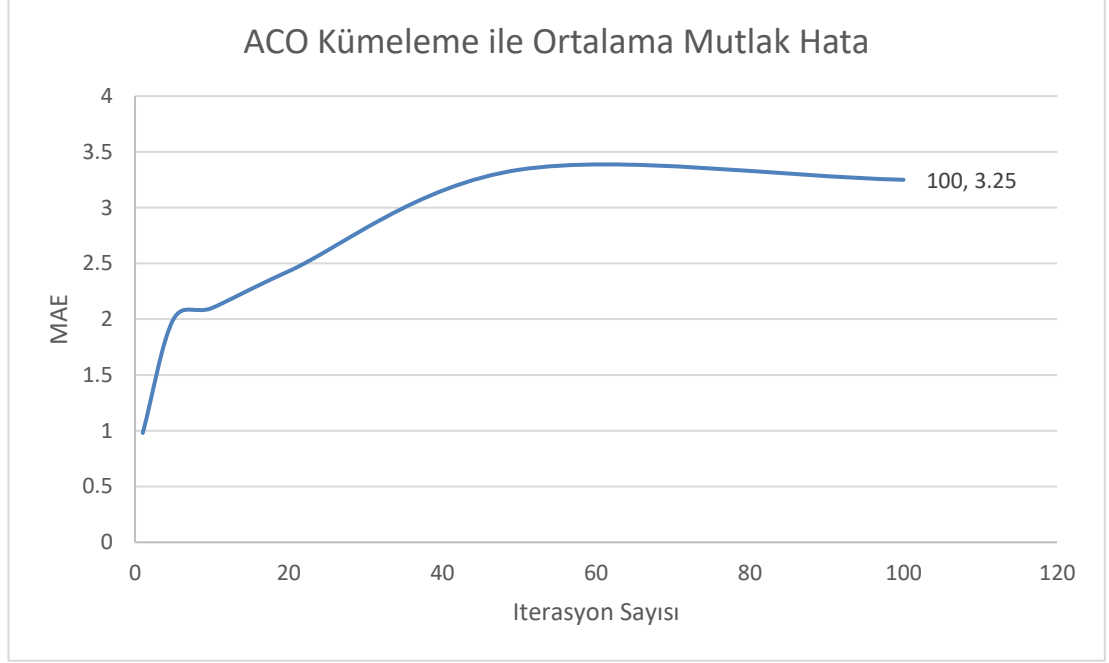
Şekil 4-1 ACO kümeleme ile karesel ortalama hata.

Şekil 4-1’de görüldüğü üzere ACO iterasyon sayısı dolayısıyla koku miktarı artması ile oluşan kümeler ve MSE arasında paralel bir korelasyon yoktur.



Şekil 4-2 ACO kümeleme ile kök ortalama kare hata.

Şekil 4-2’de görüldüğü üzere “ACO Kümeleme ile Kök Ortalama Kare Hata” hesaplarında, ACO iterasyon sayısı ve koku miktarı artması ile oluşan kümeler ve RMSE arasında paralel bir korelasyon yoktur.

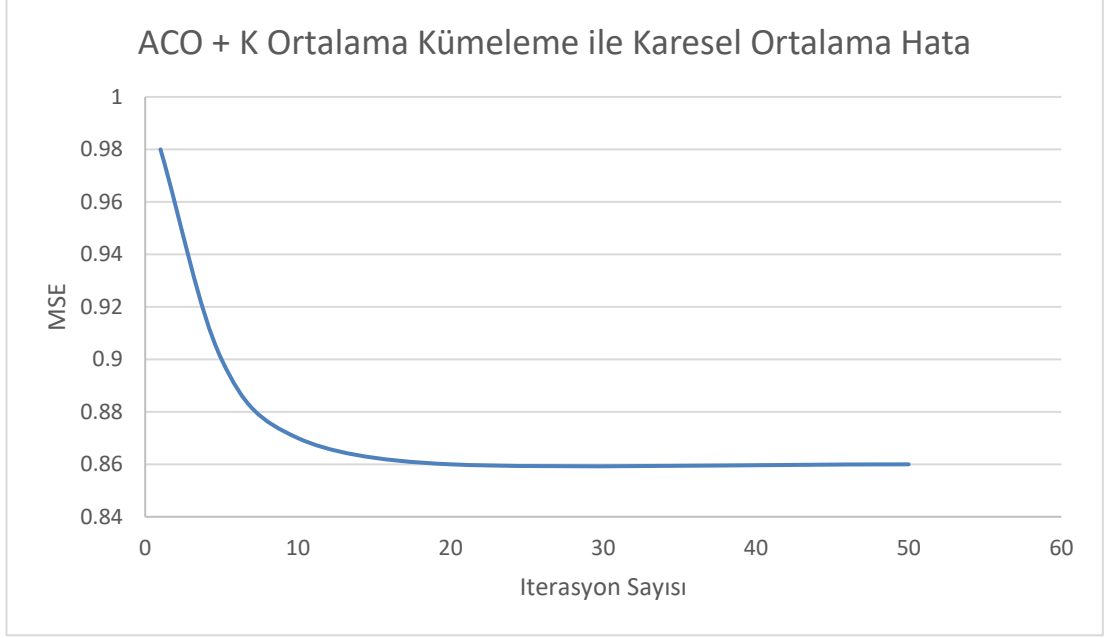


Şekil 4-3 ACO kümeleme ile ortalama mutlak hata.

Şekil 4-3’de görüldüğü üzere “ACO Kümeleme ile Ortalama Mutlak Hata” hesaplarında, ACO iterasyon sayısı ve koku miktarı artması ile oluşan kümeler ve MAE arasında paralel bir korelasyon vardır.

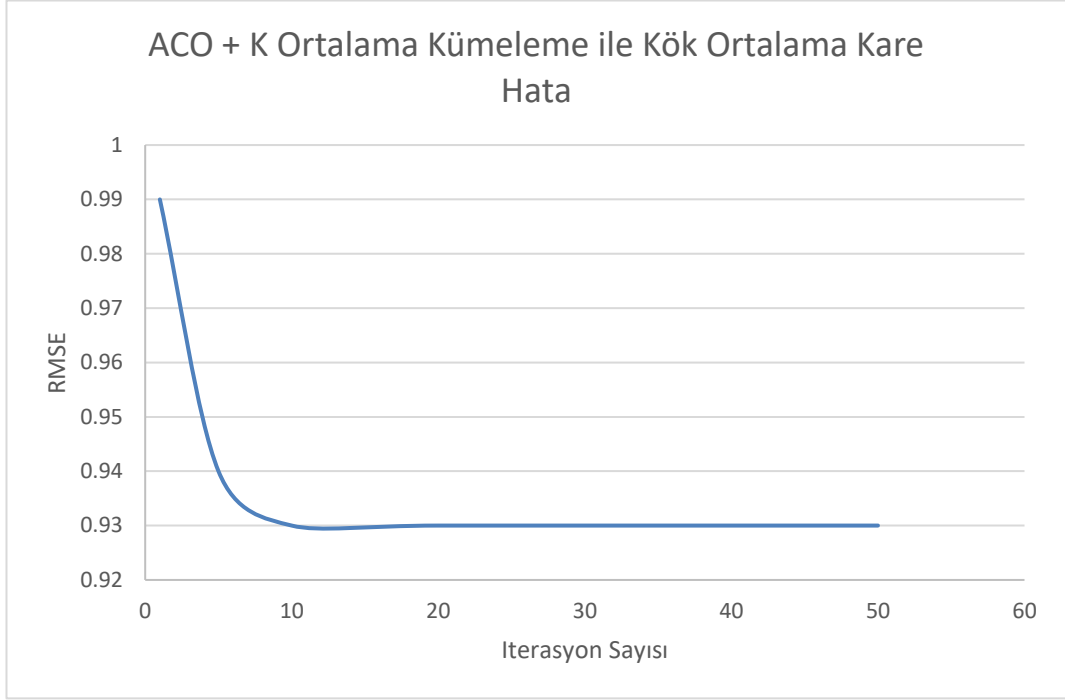
Karınca Kolonisi Optimizasyon Algoritması ve K ortalama kümeleme birlikte kullanılarak derecelendirme tahmini yapma: yapılan çalışmalarda kümeleme için karınca kolonisi optimizasyon algoritması ile yapılan testlerde istenilen sonuçlar elde edilememiştir. Bunu üzerine başka bir hibrit sistem oluşturulmuştur. Karıncaların bıraktıkları kokular K ortalama kümeleme yöntemi üzerinden kümelendi. Kullanıcıların bir filme vereceği derecelendirme bu kümeleme üzerinden tahmin edilmiştir. Sonuçlar aşağıdaki grafikteki gibidir. Test 10 karınca ile yapılmıştır. On karınca bir yineleme yaptığında alınan sonuç 5 yinelemeye göre daha az iyi sonuç vermiştir. Yineleme sayısı yirmiye ulaştığında tahmin doğruluğu oranının optimuma ulaştığı grafikte görülmektedir. Yirmiden sonraki yineleme sayıları sonuç üzerinde önemli bir değişikliğe sebep olmamaktadır. Yineleme sayısının yüz ve

üzerine çıkması halinde koku dağılımında bozulmalar görüldüğü için bu değerler dikkate alınmamıştır.



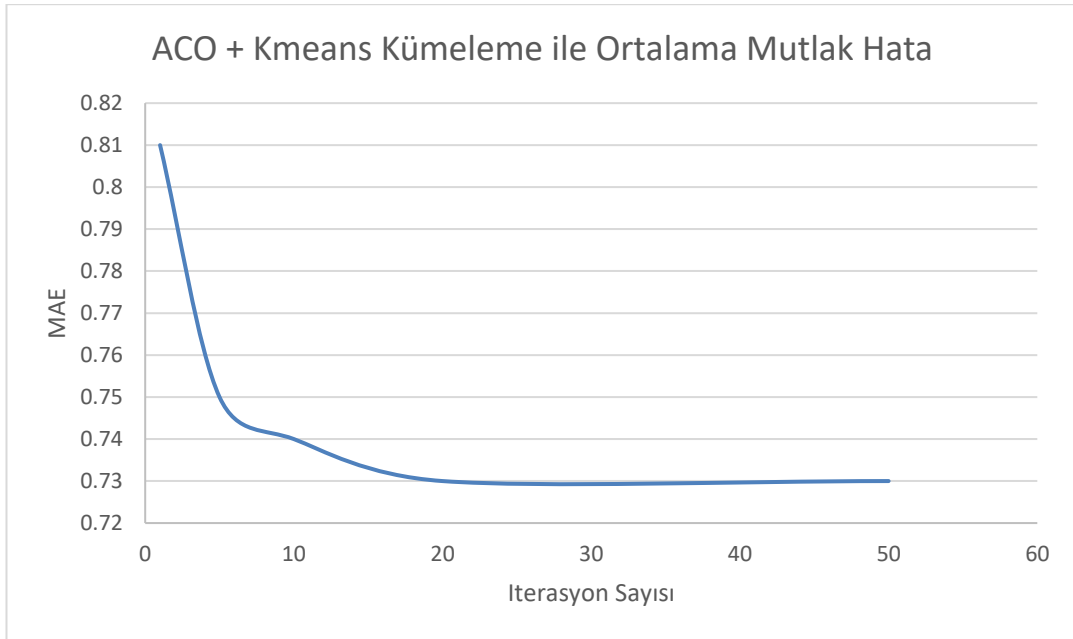
Şekil 4-4 ACO + K ortalama kümeleme ve karesel ortalama hata.

Şekil 4-4’de görüldüğü üzere “ACO + K Ortalama Kümeleme ve Karesel Ortalama Hata” hesaplarında, iterasyon sayısı ve koku miktarı artması ile oluşan kümeler ve MSE arasında ters bir korelasyon vardır.



Şekil 4-5 ACO + K ortalama kümeleme ve kök ortalama kare hata.

Şekil 4-5’de görüldüğü üzere “ACO + K Ortalama Kümeleme ve Kök Ortalama Kare Hata” hesaplarında, iterasyon sayısı ve koku miktarı artması ile oluşan kümeler ve RMSE arasında ters bir korelasyon vardır.



Şekil 4-6 ACO + K ortalama kümeleme ve ortalama mutlak hata.

Şekil 4-6’da görüldüğü üzere “ACO + K Ortalama Kümeleme ve Ortalama Mutlak Hata” hesaplarında, iterasyon sayısı ve koku miktarı artması ile oluşan kümeler ve MAE arasında ters bir korelasyon vardır.

Yaşa göre kümeleme kullanarak derecelendirme tahmini yapma: kullanıcılar yaşı 30’dan küçük olanlar birinci grup, 30 – 40 arası olanlar ikinci grup, 40 – 50 arası olanlar üçüncü grup, 50 yaşından büyükler dördüncü grup olarak kümelendi. Her kümenin öğeler için verdiği derecelendirmelerin aritmetik ortalaması hesaplanarak ortalaması bulunmuştur. Silinen oyların tahmininde bu ortalamalar atanmıştır. Tahmin edilecek ögeye verilecek derecelendirme bu ögenin yaş grubuna göre belirlenen kümenin ortalamasına eşitlenmiştir. Sonuçlar karesel ortalama hata, kök ortalama hatası ve ortalama mutlak hata metrikleri ile değerlendirilmiştir. Yaşa göre yapılan kümeleme başarıyı arttırmıştır.

Cinsiyete göre kümeleme kullanarak derecelendirme tahmini yapma: kullanıcılar cinsiyete göre kadın ve erkek olarak iki grup olarak kümelendi. Her kümenin öğelere verdiği oyların aritmetik ortalaması grup ortalaması olarak kaydedilmiştir. Tahmin edilecek oyu verecek kişi kadın ise kadın kümesinin ortalaması atanmıştır. Tahmin edilecek oyu verecek kişi bir erkek ise erkek kümesinin ortalaması atanmıştır. Sonuçlar karesel ortalama hata, kök ortalama hatası ve ortalama mutlak hata metrikleri ile değerlendirilmiştir. Cinsiyete göre yapılan kümeleme başarılı sonuçlar vermiştir.

4. TARTIŞMA VE SONUÇ

Öneri sistemleri internet çağını yaşadığımız son yirmi yılda hayatın vazgeçilmez bir parçası oldu. Kişiyi en iyi tanıyan yakın aile bireylerinin yaptığı tavsiyelerden daha iyi doğruluk oranıyla tavsiyeler veren sistemler üretildi. Verilerin hızla çoğalması ve büyümesi ile büyük verinin ortasında bir şeyler aramak samanlıkta iğne aramaya benzedi. Tam da bu noktada işbirlikçi filtreleme hayatımıza girdi. İşbirlikçi filtreleme ile artık herhangi bir niteliğe özgü filtreleme yapıp, büyük verinin sadece ilgilenilen kısmı üzerinde işlem yapılabilirdi. İşbirlikçi filtrelemede bir öz niteliğe özgü filtreleme yapılmasının birçok avantajının yanı sıra bazı dezavantajları da vardı. Avantajı samanlıkta iğne aramaktan bizi kurtarması iken, dezavantajı tek bir bakış açısını yansıtması olmuştur.

Hibrit modelde işbirlikçi filtrelemenin bu dezavantajları giderilmiştir. Önerilen yeni hibrit sistemin hali hazırda süregelen klasik yapıya oranla daha iyi sonuçlar aldığı çizelge verileri ile “doğrulanmış model sonuç karşılaştırmaları” bölümünde açıklanmıştır.

Yeni önerilen sistem üzerine akademik dünyada daha önce çalışma yapılmamıştır. İleriki zamanlarda bu konu üzerine çok fazla çalışma yapılacağı ön görülmektedir. Hali hazırda filtreleme, demografik bilgiler, kişisel veya ögesel öznitelikler olarak tamamen birbirinden ayrı olarak çalışılmaktadır. Yapay zekânın hızla ilerlediği, yeni algoritmaların ve yeni bilginin üstel olarak arttığı günümüz dünyasında, farklı paradigmalardan bir arada çalışılması ile yeni hibrit model önerileri kaçınılmaz olacaktır.

KAYNAKÇA

- [1] Sachi Nandan Mohanty (editor) and Jyotir Moy Chatter, "Recommender System with Machine Learning and Artificial Intelligence Practical Tools and Applications in Medical, Agricultural and Other Industries," 2020.
- [2] V. Bhatnagar, *Collaborative filtering using data mining and analysis*. 2016.
- [3] Jannach Dietmar, "Recommender Systems An Introduction," 2010.
- [4] Rounak Banik, "Hands-On Recommendation Systems with Python," 2018.
- [5] Sherif. El-Tawil and Structural Engineering Institute. Technical Committee on Composite Construction., *Recommendations for seismic design of hybrid coupled wall systems*. SEI/America Society of Civil Engineers, 2010.
- [6] F. Ricci, · Lior, R. Bracha, and S. Editors, "Recommender Systems Handbook Second Edition," New York, 2015.
- [7] F. Ricci, L. Rokach, and B. Shapira, "Introduction to Recommender Systems Handbook," in *Recommender Systems Handbook*, Springer US, 2011, pp. 1–35.
- [8] F. Ricci, B. Shapira, and L. Rokach, *Recommender systems handbook, Second edition*. 2015.
- [9] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Transactions on Information Systems*, vol. 22, no. 1. 2004, doi: 10.1145/963770.963772.
- [10] A. Jøsang, R. Ismail, and C. Boyd, "A survey of trust and reputation systems for online service provision," *Decision Support Systems*, vol. 43, no. 2, 2007, doi: 10.1016/j.dss.2005.05.019.
- [11] J. S. Breese, D. Heckerman, and C. Kadie, "Empirical Analysis of Predictive Algorithms for Collaborative Filtering," 1998.
- [12] K. Falk, "Practical Recommender Systems," 2019.
- [13] S. Berkovsky, T. Kuflik, and F. Ricci, "Mediation of user models for enhanced personalization in recommender systems," *User Modeling and User-Adapted Interaction*, vol. 18, no. 3, 2008, doi: 10.1007/s11257-007-9042-9.
- [14] F. Ricci, B. Shapira, and L. Rokach, *Recommender Systems Handbook*. 2011.

- [15] R. Burke, "Hybrid web recommender systems," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2007, vol. 4321 LNCS, doi: 10.1007/978-3-540-72079-9_12.
- [16] A. Dubey, A. Gupta, N. Raturi, and P. Saxena, "Item-based collaborative filtering using sentiment analysis of user reviews," *Communications in Computer and Information Science*, vol. 899, 2019, doi: 10.1007/978-981-13-2035-4_8.
- [17] F. Ricci, B. Shapira, and L. Rokach, "Recommender systems: Introduction and challenges," in *Recommender Systems Handbook, Second Edition*, 2015.
- [18] M. Montaner, B. López, and J. L. de La Rosa, "A taxonomy of recommender agents on the internet," *Artificial Intelligence Review*, vol. 19, no. 4, 2003, doi: 10.1023/A:1022850703159.
- [19] K. Verbert *et al.*, "Context-aware recommender systems for learning: A survey and future challenges," *IEEE Transactions on Learning Technologies*, vol. 5, no. 4. 2012, doi: 10.1109/TLT.2012.11.
- [20] E. Cho, S. A. Myers, and J. Leskovec, "Friendship and mobility: User movement in location-based social networks," 2011, doi: 10.1145/2020408.2020579.
- [21] Y. S. Kim *et al.*, "People recommendation based on aggregated bidirectional intentions in social network site," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2010, vol. 6232 LNAI, doi: 10.1007/978-3-642-15037-1_21.
- [22] M. Coyle and B. Smyth, "Information recovery and discovery in collaborative Web search," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2007, vol. 4425 LNCS, doi: 10.1007/978-3-540-71496-5_33.
- [23] K. Swearingen and R. Sinha, "Beyond Algorithms : An HCI Perspective on Recommender Systems," *ACM SIGIR 2001 Workshop on Recommender Systems (2001)*, 2001.
- [24] N. F. Awad and M. S. Krishnan, "The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization," *MIS Quarterly: Management Information Systems*, vol. 30, no. 1, 2006, doi: 10.2307/25148715.
- [25] L. Chen, M. de Gemmis, A. Felfernig, P. Lops, F. Ricci, and G. Semeraro, "Human decision making and recommender systems," *ACM Transactions on Interactive Intelligent Systems*, vol. 3, no. 3, 2013, doi: 10.1145/2533670.2533675.

- [26] R. Hu and P. Pu, "A study on user perception of personality-based recommender systems," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2010, vol. 6075 LNCS, doi: 10.1007/978-3-642-13470-8_27.
- [27] D. Edelmann, T. F. Móri, and G. J. Székely, "On relationships between the Pearson and the distance correlation coefficients," *Statistics and Probability Letters*, vol. 169, 2021, doi: 10.1016/j.spl.2020.108960.
- [28] G. C. Cardarilli, L. di Nunzio, R. Fazzolari, A. Nannarelli, M. Re, and S. Spano, "N-Dimensional Approximation of Euclidean Distance," *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 67, no. 3, 2020, doi: 10.1109/TCSII.2019.2919545.
- [29] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *American Statistician*, vol. 46, no. 3, 1992, doi: 10.1080/00031305.1992.10475879.
- [30] M. Ahmed, R. Seraj, and S. M. S. Islam, "The k-means algorithm: A comprehensive survey and performance evaluation," *Electronics (Switzerland)*, vol. 9, no. 8, 2020, doi: 10.3390/electronics9081295.
- [31] F. Spagnolo, F. Frustaci, S. Perri, and P. Corsonello, "An efficient connected component labeling architecture for embedded systems," *Journal of Low Power Electronics and Applications*, vol. 8, no. 1, 2018, doi: 10.3390/jlpea8010007.
- [32] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm," *Journal of Global Optimization*, vol. 39, no. 3, 2007, doi: 10.1007/s10898-007-9149-x.
- [33] M. Dorigo and C. Blum, "Ant colony optimization theory: A survey," *Theoretical Computer Science*, vol. 344, no. 2–3, 2005, doi: 10.1016/j.tcs.2005.05.020.
- [34] M. Dorigo, V. Maniezzo, and A. Colorni, "Ant system: Optimization by a colony of cooperating agents," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 26, no. 1, 1996, doi: 10.1109/3477.484436.
- [35] D. Whitley, "A genetic algorithm tutorial," *Statistics and Computing*, vol. 4, no. 2, 1994, doi: 10.1007/BF00175354.
- [36] V. Tereshko, "Reaction-diffusion model of a honeybee colony's foraging behaviour," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2000, vol. 1917, doi: 10.1007/3-540-45356-3_79.

- [37] D. Teodorović, "Transport modeling by multi-agent systems: A swarm intelligence approach," *Transportation Planning and Technology*, vol. 26, no. 4, 2003, doi: 10.1080/0308106032000154593.
- [38] H. Drias, S. Sadeg, and S. Yahi, "Cooperative bees Swarm for solving the maximum weighted satisfiability problem," in *Lecture Notes in Computer Science*, 2005, vol. 3512, doi: 10.1007/11494669_39.
- [39] K. Benatchba, L. Admane, and M. Koudil, "Using bees to solve a data-mining problem expressed as a Max-Sat one," in *Lecture Notes in Computer Science*, 2005, vol. 3562, no. PART II, doi: 10.1007/11499305_22.
- [40] H. F. Wedde, M. Farooq, and Y. Zhang, "BeeHive: An efficient fault-tolerant routing algorithm inspired by honey bee behavior," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2004, vol. 3172 LNCS, doi: 10.1007/978-3-540-28646-2_8.
- [41] X. S. Yang, "Engineering optimizations via nature-inspired virtual bee algorithms," in *Lecture Notes in Computer Science*, 2005, vol. 3562, no. PART II, doi: 10.1007/11499305_33.
- [42] D. Karaboga and B. Basturk, "An Artificial Bee Colony (ABC) Algorithm for Numeric function Optimization," *IEEE Swarm Intelligence Symposium 2006. Indianapolis, Indiana, USA*, 2006.
- [43] J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," 1999, doi: 10.1145/312624.312682.
- [44] C. J. Willmott and K. Matsuura, "Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance," *Climate Research*, vol. 30, no. 1, 2005, doi: 10.3354/cr030079.
- [45] Y. Ephraim and D. Malah, "Speech Enhancement Using a Minimum Mean-Square Error Short-Time Spectral Amplitude Estimator," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 32, no. 6, 1984, doi: 10.1109/TASSP.1984.1164453.