ÇOKLU-ÖLÇÜTLÜ ORTAK FİLTRELEME SİSTEMLERİNDE KABA KÜMELER KULLANIMI

Emin Talip DEMİRKIRAN

Doktora Tezi

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Donanımı Bilim Dalı

Danışman: Prof. Dr. Yaşar HOŞCAN

Eskişehir

Eskişehir Teknik Üniversitesi

Lisansüstü Eğitim Enstitüsü

Şubat 2021

ÖZET

ÇOKLU-ÖLÇÜTLÜ ORTAK FİLTRELEME SİSTEMLERİNDE KABA KÜMELER KULLANIMI

Emin Talip DEMİRKIRAN

Bilgisayar Mühendisliği Anabilim Dalı

Bilgisayar Donanımı Bilim Dalı

Eskişehir Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Şubat 2021

Danışman: Prof. Dr. Yaşar HOŞCAN

Teknolojinin gelişmesi ve Internet'in yaygınlaşması ile birlikte, bilgiye ulaşmak oldukça kolaylaşmıştır. Milyarlarca bilgiye saniyeler içerisinde kolaylıkla ulaşabilme imkânı, beraberinde bilgi bombardımanı denilen problemi de getirmiştir. Bu problemin üstesinden gelmek adına, bilgi yığınlarının içerisinde ihtiyaç olanın hangisi olabileceğini tahmin etmek üzere Öneri Sistemleri oluşturulmuştur. Birbirinden farklı yaklaşımları bulunan onlarca öneri sistemi içerisinde en yaygın olarak kullanılan tekniklerden biri Ortak Filtreleme sistemleridir.

Ortak Filtreleme sistemlerinde kullanıcılar bilgi sahibi oldukları ürünlere oy vererek karakterlerini ortaya koyarlar ve sistem de karşılığında bilgi sahibi olmadıkları ürünleri beğenip beğenmeyeceklerini söyler. Ortak filtreleme sistemlerinde ürünler genel anlamda değerlendirilir ancak bazen; bir ürüne tek bir açıdan yaklaşmaktansa birden fazla perspektiften yaklaşmak, sistemin daha başarılı olmasını sağlar. Bu bağlamda, çoklu-ölçütlü öneri sistemleri geliştirilmiştir. Bu doktora tezinde, çoklu-ölçütlü ortak filtreleme sistemleri derinlemesine incelenmiş ve kaba kümeler teorisi kullanılarak oluşturulmuş toplama fonksiyonu tabanlı özgün bir yaklaşım sunulmuştur.

Geleneksel ortak filtreleme sistemlerinin ölçeklenebilirlik, seyreklik gibi problemleri mevcuttur ve çoklu-ölçütlü ortak filtreleme sistemleri de ortak filtreleme sistemlerinin çoklu uzayı olarak değerlendirildiğinden, bu problemler çok daha büyük sorun teşkil etmektedir. Ayrıca, kriterler arasındaki ilişkinin ortaya çıkarılması da bu sistemler için bir başka optimizasyon problemidir. Dolayısıyla sistemin tahmin üretmedeki doğruluğun artırılması bir zorluk olarak karşımıza çıkmaktadır. Bu doktora tezinde, çoklu-ölçütlü ortak filtreleme sistemlerinin tahmin doğruluğunun artırılması adına birtakım yaklaşımlarda bulunulmuştur.

Bu sistemlerin içerdiği problemlerin üstesinden gelebilmek adına Kaba Kümeler Teorisinden faydalanılan özgün bir yaklaşım önerilmiştir.

Anahtar Sözcükler: Veri Madenciliği, Öneri Sistemleri, Ortak Filtreleme, Çoklu-Ölçütlü Ortak Filtreleme, Kaba Kümeler Teorisi.

ABSTRACT

THE USE OF ROUGH SETS THEORY IN MULTI-CRITERIA RECOMMENDER SYSTEMS

Emin Talip DEMİRKIRAN

Department of Computer Engineering

Programme in Computer Hardware

Eskişehir Technical University, Institute of Graduate Programs, February 2021

Supervisor: Prof. Dr. Yaşar HOŞCAN

Accessing the information has become quite easier with the development of technology and the discovery of the Internet but the opportunity to get billions of data in seconds brought the problem called Information Overload with it. In order to overcome this problem, recommender systems that predicts the necessary information from tons of data has been proposed. Within the dozens of different approaches for recommender systems, collaborative filtering is one of the most commonly used systems for making predictions.

In collaborative filtering recommender systems, the users put forth their characteristics by giving rating to the products they know and the system produces prediction to the products they do not have any idea in return. In traditional collaborative filtering, the products are considered as a whole, but sometimes approaching to a product from multiple aspects makes the system more successful. In this context, multi-criteria recommender systems are developed. In this PhD dissertation, in-depth study of multi-criteria recommender systems is performed and an aggregation function-based multi-criteria collaborative filtering using Rough Sets Theory is proposed as a novel approach.

The traditional collaborative filtering systems have problems such as scalability and sparsity. Since the multi-criteria collaborative filtering systems are considered as the extended version of the traditional single-rating collaborative filtering systems, these problems pose a much grater problem. Additionally, revealing the relation between criteria is yet another optimization problem. Hence, increasing the accuracy in prediction is a challenge. In this dissertation, a number of approaches have been made to improve the predictive accuracy of multi-criteria recommender systems. A novel approach which uses Rough Sets Theory has been proposed to overcome the challenges of these systems.

Keywords: Data Mining, Recommender Systems, Collaborative Filtering, Multi-Criteria Collaborative Filtering, Rough Sets Theory.

TEŞEKKÜR

Bu çalışmada danışmanlığımı yürüten, önerileri ve fikirleriyle beni yönlendiren Prof. Dr. Yaşar HOŞCAN hocama en içten saygı ve şükranlarımı sunuyorum. Başarılı, sofistike ve anlayışlı bir insan olarak, örnek bir akademisyenin nasıl olması gerektiğini gösterdiği için kendilerine ayrıca teşekkür ederim.

Danışman hocamın yanı sıra tez dönemim boyunca bana yol gösteren, eleştirileriyle katkıda bulunan Prof. Dr. Bekir Taner DİNÇER'e, Dr. Öğr. Üyesi İsmail SAN'a ve Dr. Öğr. Üyesi Mehmet KOÇ'a sonsuz teşekkürlerimi sunarım.

Tez konumu belirlememde, bu alanda çalışmamda bana yol gösteren ve bu tezin oluşmasında büyük katkıları olan değerli hocalarım Doç. Dr. Alper BİLGE'ye ve Doç. Dr. Cihan KALELİ'ye minnet ve şükranlarımı belirtmek isterim.

Son olarak, tezimi tamamlayabilmek için bana her zaman manevi destek olan aileme, ilham verici ve motive edici konuşmaları için babam Ecir DEMİRKIRAN'a ve her koşulda yanımda olan değerli eşim Burçin DEMİRKIRAN'a teşekkürlerimi sunarım.

Emin Talip DEMİRKIRAN

12/02/2021

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Eskişehir Teknik Üniversitesi tarafından kullanılan "bilimsel intihal tespit programı"yla tarandığını ve hiçbir şekilde "intihal içermediğini" beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

Emin Talip DEMİRKIRAN

İÇİNDEKİLER

Sayfa

TABLOLAR DİZİNİ

Sayfa

GÖRSELLER DİZİNİ

Sayfa

SİMGELER VE KISALTMALAR DİZİNİ

OF	:	Ortak Filtreleme
PCC	:	Pearson Correlation Coefficient (Pearson Korelasyon Katsayısı)
***k*-NN**	:	*k*-Nearest-Neighbor (*k* En Yakın Komşu)
PCA	:	Principal Component Analysis (Temel Bileşen Analizi)
POI	:	Point of Interests
ÇÖOF	:	Çoklu-Ölçütlü Ortak Filtreleme
ÇÖKV	:	Çoklu-Ölçütlü Karar Verme
ÇÖTS	:	Çoklu-Ölçütlü Tavsiye Sistemleri
AHP	:	Analytic Hierarchy Process
ANP	:	Analytic Network Process
TOPSIS	:	Technique for Order of Preference by Similarity to Ideal Solution
ELECTRE	:	ELimination Et Choix Traduisant la REalité
FMM	:	Flexible Mixture Model
HOSVD	:	Higher-Order Singular Value Decomposition
ANFIS	:	Adaptive Neuro-Fuzzy Inference System
SOM	:	Self-Organizing Map
KKT	:	Kaba Kümeler Teorisi
OMH	:	Ortalama Mutlak Hata (Mean Absolute Error)
KOKH	:	Kare Ortalamalarının Karekökü Hatası (Root Mean Squared Error)
DKB	:	Düzeltilmiş Kosinüs Benzerliği (Adjusted Cosine Similarity)
σ	:	Standart Sapma
$\vee$	:	Ayrılma İşlemi
$\wedge$	:	Birleşme İşlemi

1. GİRİŞ

Günümüzde, İnternet'in hızlı gelişimiyle birlikte, ihtiyaç olunan her türlü bilgiye erişmek mümkün olmuştur. İnsanlar alışveriş yaparken, seyahat ederken, gazete veya kitap okurken İnternet'in kendilerine sunduğu hizmetleri kullanmaktadırlar. Bankacılık işlemlerini gerçekleştirmek için bankaya gitmek, bir kitap, film, müzik albümü, kıyafet vs. satın almak için mağazaları dolaşmak veya yeni seyahat ettiği bir şehirde saatlerce kalacak yer aramak zorunda kalmaktan önemli ölçüde kurtulmuşlardır. Günlük hayatta pek çok ihtiyacımızı karşılayan İnternet ortamında kullanıcılar, ihtiyaçları doğrultusunda aradıklarına kolaylıkla erişebilseler de verinin fazlalığı ve çeşitliliği sebebiyle ürün seçimi konusunda sorunlar yaşayabilmektedirler. Örneğin, dünyanın en büyük e-ticaret sitelerinden biri olan Amazon.com[1] tek başına milyonlarca ürünü oturduğumuz yerden satın alabilmemiz için ürünleri tek bir çatı altında toplamayı başarabilmiştir. Ancak, İnternet'e erişen kullanıcı sayısının ve çevrimiçi ürün sayısının artmasıyla birlikte bilgi bombardımanı (Lu vd., 2015) adı verilen ciddi bir problem oluşmuş, binlerce ürün arasından müşterinin hangi ürünü seçmesi gerektiğini bilmesi neredeyse olanaksız hale gelmiştir.

Günlük hayatta karşımıza çıkan seçenekler arasında seçim yapmak birçoğumuz için önemli bir problem olmuştur. Kitap satın almak için gittiğimiz bir kitabevinde onlarca raf içerisinde bulunan binlerce kitaptan hangisini seçmemiz gerektiği, ilk kez seyahat ettiğimiz bir şehirde hangi restoranları tercih etmemiz gerektiği, bir konuyu anlamak için internette bulduğumuz onlarca kaynak içerisinden okumaya hangisinden başlamamız gerektiği hep bir soru işaretidir. Hepimiz bir film izlemeden ya da bir aktiviteye katılmadan önce yapmak istediğimiz şey hakkında bilgi sahibi olmak isteyen birilerinden yardım almak isteriz. Bu davranışımızın ardında yalnızca kararsızlık değil, zamandan kazanmak, gereksiz yere para kaybetmemek gibi birçok sebep yatar. Müşterilerin bu problemini ortadan kaldırabilmek için, kullanıcılara kişisel öneriler sunan sistemler kullanmaya başlamışlardır. (Goldberg vd., 1992; Resnick ve Varian, 1997; Han ve Kamber, 2006).

Tavsiye sistemleri, müşterilerine hakkında fikir sahibi olmak istedikleri ürünlere tahmin üreten sistemlerdir (Yirgu ve Meier, 2016). Sistemin tahmin üretebilmesi için, genellikle müşterilerin karakteristiği hakkında bilgi sahibi olması gerekir. Bu amaçla,

[1] https://www.amazon.com

müşterilerin ürünler hakkında vermiş oldukları tercihleri toplar. Tercihler iki yolla toplanabilir (Yi vd., 2014): Direkt olarak – *bir diğer deyişle üstü açık* – ; ürünler hakkında yorum yapmasını veya oy vermesini isteyerek veya dolaylı olarak – *bir diğer deyişle üstü kapalı* – ; müşterilerin sıklıkla ziyaret ettiği web siteleri, çerezler adı verilen; dil tercihleri, konum bilgileri gibi faydalı bilgiler içeren metin parçaları, bir ürünü incelerken geçirdikleri zaman, ürünü inceleme sıklığı gibi bilgiler kullanılarak (Oard ve Kim, 1998; Park ve Pennock, 2007). Sistem, bu tercihler yardımıyla, kullanıcının ürünleri beğenip beğenmediğini anlamaya çalışır.

Tavsiye sistemlerinin tahmin üretme yaklaşımlarına göre de farklı türleri mevcuttur: İçerik tabanlı filtreleme, ortak filtreleme (OF), demografik filtreleme, olay tabanlı muhakeme ve melez filtreleme yaklaşımları. Bu tez çalışmasında, tavsiye sistemlerinin en yaygın örneklerinden biri olan OF ve OF sistemlerinin, nesneleri birden fazla kıstas üzerinden incelemeye olanak tanıyan çeşidi olan çoklu-ölçütlü ortak filtreleme (ÇÖOF) sistemleri üzerinde durulmuştur.

Geleneksel OF sistemleri bir kullanıcı veya bir ürün için değerlendirme yaparken, ilgili nesneyi tek bir açıdan değerlendirmektedir. Oysa bir ürünü tek bir açıdan değerlendirmek, o ürün ile ilgili yeterli bilgiye erişememe neden olabilir. Bu nedenle, ilgili ürünü önceden belirlenmiş birkaç kıstas üzerinden değerlendirmek sistemin daha başarılı olmasını sağlayabilir. Örneğin, aynı otele giden ve o otele sayısal olarak aynı değerlendirmeyi yapan iki kullanıcıdan biri oteli konumundan ötürü beğenirken, diğeri ise temizliğinden ötürü beğenmiş olabilir. Aynı değerlendirmeyi yapmış olmaları, bu iki kullanıcının o otel ile ilgili aynı görüşlere sahip olduğu anlamına gelmeyebilir. Bu durumun araştırmacılar tarafından fark edilmesi üzerine, geleneksel OF sistemlerinin genişletilmiş hali olan ÇÖOF sistemleri literatüre dahil olmuştur.

OF sistemlerinin kıtlık, ölçeklenebilirlik, veri seyrekliği, soğuk başlangıç, gizlilik gibi birçok problemleri mevcuttur ve ÇÖOF sistemleri OF sistemlerinden çok daha fazla veri içerdiğinden bu problemler daha büyük ve daha karmaşık sorunlar teşkil etmektedir. Ayrıca, kriterler arasındaki ilişkinin ortaya çıkarılması da bir başka optimizasyon problemidir. Bu tez çalışmasında, ÇÖOF sistemlerinin özellikle doğruluk ve seyreklik gibi problemlerinin üstesinden gelinmeye çalışılmış, kriterler arasındaki ilişkilerin ortaya dökülmesinin sistem performansını ne derece etkilediği anlaşılmaya çalışılmıştır. Bu problemlerin üstesinden gelebilmek adına, veri kümelerindeki gizli örüntüleri başarılı bir

şekilde ortaya çıkaran Kaba Kümeler Teorisi (KKT)'nden faydalanılmıştır. KKT'nin eksik veya keşfedilmemiş örüntüleri ortaya çıkarmadaki başarısı KKT'yi ÇÖOF sistemlerine entegre etmemizdeki başlıca sebeplerden bir tanesidir. Ayrıca, Bulanık Kümeler (Aydemir ve Gündüz, 2020) veya Nötrosofik Kümeler (Dat vd., 2019) gibi yaklaşımların aksine, KKT'nin çıkarım yapmak için ek bir bilgiye veya üyelik fonksiyonuna ihtiyaç duymaması bu çalışmada KKT'nin tercih edilmesindeki etkenlerden bir diğeridir. Son olarak, henüz literatürde, KKT'nin ÇÖOF sistemlerine dahil edildiği herhangi bir sistemin olmayışı da bu yöntemi geliştirmemizdeki motivasyonlarımızdan biridir.

Tezin sonraki bölümleri sırasıyla şu şekildedir: İkinci bölümde tavsiye sistemleri ve çeşitlerinden kısaca bahsedilmiş, üçüncü bölümde ise tezin ana odak noktası olan OF sistemlerinin çeşitlerinden, bu sistemlerde yaşanan problemlerden ve sistemin çalışma yapısından bahsedilmiştir. Dördüncü bölümde ÇÖOF sistemleri anlatılmış ve beşinci bölümde ÇÖOF sistemlerinde KKT kullanımının nasıl gerçekleştirildiği üzerinde durulmuştur. Altıncı bölümde, önerilen sistemin başarısı deneysel bulgular ile test edilmiş ve son olarak, sekizinci bölümde tez çalışması sonuçlandırılmıştır.

2. TAVSİYE SİSTEMLERİ

Tavsiye sistemleri, diğer bir deyişle Öneri Sistemleri, literatürde dört ana kategoride toplanır (Takacs, 2009). Bu dört kategori şu şekilde sıralanabilir: Metinlerden çıkarımların yapıldığı içerik tabanlı filtreleme, ürün veya kullanıcıların profillerinin çıkarıldığı demografik filtreleme, sayısal oylardan çıkarımların yapıldığı ortak filtreleme ve bahsi geçen filtreleme sistemlerinin avantajlı yanlarını bir araya getirerek, doğruluk, performans gibi sorunlarını ortadan kaldırmayı hedefleyen melez yaklaşımlar.

2.1.İçerik Tabanlı Filtreleme

İçerik tabanlı filtreleme sistemlerinde kullanıcılar, hakkında bilgi sahibi oldukları ürünlere metinsel yorumlarda bulunurlar. Sistem kullanıcı metninin içeriğini inceleyerek ürün hakkındaki tercihini ortaya çıkarır. Bu yaklaşımda tahmin kalitesini artırmak için kullanıcıların karakteristiği hakkında bilgiler veren meta datalar, ürünleri kategorize etmeye yarayan etiketler (Tag'ler) kullanılabilir.

2.2.Demografik Filtreleme

Bu yaklaşımda müşterilerin Stereotipi adı verilen yaşı, cinsiyeti, uzmanlık alanı, mesleki bilgileri gibi kalıplaşmış bilgiler temel alınarak demografik kurallar oluşturulur ve bu kurallar temel alınarak tavsiye üretilir (Vozalis ve Margaritis, 2007). Örneğin bir otel rezervasyon sisteminde kullanıcıların yaşı, cinsiyeti gibi temel bilgiler kullanılarak müşterilerin tercihleri hakkında genel bilgiler edinilebilir. Beel vd., (2013) çalışmalarında yaşlı müşterilerin gençlere oranla daha fazla tavsiye istediğini ortaya koymuştur.

2.3.Ortak Filtreleme

Ortak filtreleme (OF), birçok çevrimiçi satıcı tarafından tavsiye amaçlı kullanılan en popüler ve başarılı bilgi filtreleme sistemidir. İlk olarak 1992'de Tapestry Projesi'nin tasarımcıları tarafından ortaya atılan (Goldberg vd., 1992) OF sistemleri, günümüzde e-ticaret uygulamalarından arama motorlarına kadar birçok önemli alanda yer bulmuş sistemlerdir. Pratikte oldukça hızlı ve başarılı bir sistem olan OF sistemleri en popüler tavsiye sistemlerindendir ancak bu sistemin başarısı yeterli sayıda kullanıcı oyları olmasına bağlıdır (Q. Li ve Kim, 2003; Han ve Kamber, 2006; Ahn, 2008).

2.4.Melez Yaklaşımlar

Her tavsiye sisteminin bir diğerine göre avantajlı ve dezavantajlı yanları vardır. Örneğin OF sistemleri içerik tabanlı filtreleme sistemlerine göre çok daha hızlı çalışır.

Milyonlarca nesnenin yer aldığı bir veri kümesi ile çalışmak OF için oldukça kolay iken içerik tabanlı filtreleme sistemleri alan (domain) modelleme ve meta data çıkarımı gibi adımlara ihtiyaç duyar. Bunun yanı sıra OF sistemlerinde "Gray Sheep" adı verilen, kullanıcıya hiçbir gruba dâhil olamamasından ötürü tavsiye üretememek gibi bir problem mevcutken, içerik tabanlı filtrelemede böyle bir sorun yoktur. Dolayısıyla her filtreleme sisteminin kendine has avantajları ve dezavantajları vardır ve bazı problemleri ortadan kaldırmak için bu farklı sistemleri bir araya getirmek sorunların önemli ölçüde ortadan kalkmasını sağlar. Bu tür sistemlere melez sistemler denir. Kısaca ifade etmek gerekirse melez yaklaşımlar, filtreleme sistemlerinin problemlerinin üstesinden gelinmesi amacıyla bir araya getirilen sistemlerden oluşurlar (Jung vd., 2004; Mathew vd., 2016; Yang vd., 2017).

3. GELENEKSEL ORTAK FİLTRELEME SİSTEMLERİ

OF sistemlerinin temelinde yatan düşünce; hakkında bilgi sahibi oldukları ürünlere benzer tercihlerde bulunan kullanıcıların, yeni bir ürüne olan tercihlerinin de benzer yönde olacağı şeklindedir. Bir diğer deyişle, geçmişte benzer tercihlerde bulunan kullanıcıların gelecekte de benzer eğilimlerde bulunacağı düşüncesidir (Goldberg vd., 2001; O'Mahony, 2004; Ricci vd., 2011). Örneğin aynı müziği dinleyip benzer tercihlerde bulunan kullanıcılar gelecekte dinleyecekleri müzikler hakkında tercih verirken ürünler hakkında yine benzer yaklaşımlarda bulunacaklardır.

Bu yaklaşıma göre sistem ilk olarak kullanıcı tercihlerini göz önünde bulundurarak hedef kullanıcıya benzeyen kullanıcıları – bir diğer deyişle komşuları – bulur, ardından tahminde bulunulacak hedef ürünle ilgili komşuların tercihlerini kullanarak ilgili kullanıcının o ürünü beğenip beğenmeyeceğini tahmin eder.

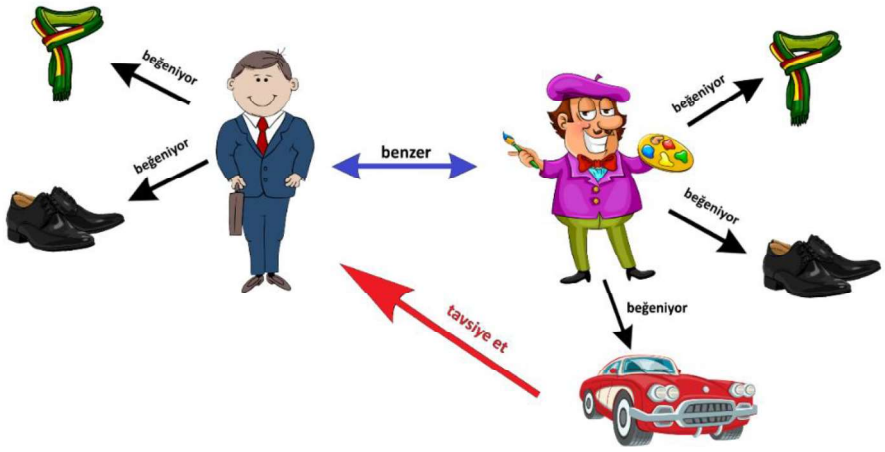
OF sistemlerinin kullanıcıları tanıyabilmesi ve tahmin üretebilmesi için, sistem kullanıcılardan hakkında fikir sahibi oldukları ürünlerle ilgili tercih belirtmesini ister. Bu tercihler genellikle ikili beğeni değerleri (Goldberg vd., 1992; Mild ve Reutterer, 2003) ve sayısal değer yelpazesi (Herlocker vd., 1999; Sarwar vd., 2001) şeklinde iki çeşittir. İkili beğeni değerleri ‘0’ ve ‘1’ şeklindedir. ‘0’kullanıcının ürünü beğenmediğini, ‘1’ ise beğendiğini ifade eder. Sayısal değer yelpazesi ise kullanıcının ürünü sevme derecesine olanak veren çoklu beğeni değerlerinden oluşur. Örneğin yelpazenin 1’den 10’a sayılarla ifade edildiği bir sistemde, en kötü değer olan 1’den en iyi değer olan 10’a doğru, 1-2-3 kullanıcının ürünü beğenmediğini, 4-5-6 ürün hakkında tarafsız kaldığını, 7 ve yukarısı değerler ise kullanıcının o ürünü beğendiğini temsil eder. Bazı sistemler ise tercihler için harf notasyonunu tercih eder (Koren ve Sill, 2013). Örneğin Yahoo!Movies veri tabanı filmler için tercih olarak A’dan F’ye harf derecelendirme sistemi kullanır. En iyi tercih A^+, en kötü tercih ise F olacak şekilde 13 adet tercih değeri vardır. Sistem kullanıcıların tercihlerini $N \times M$’lik bir kullanıcı-ürün matrisinde toplar. N sistemde kayıtlı kullanıcı sayısını temsil ederken, M ise ürün sayısını temsil eder. Matrisin hücrelerinde ilgili kullanıcının ilgili ürüne denk gelen tercihini belirten sayısal değerler bulunur. Toplanan kullanıcı tercihlerinin bir araya getirildiği bu matris yardımıyla sistem kullanıcılar arasındaki benzerlikleri ortaya çıkararak, kullanıcının hakkında fikir sahibi olmadığı bir ürün hakkında benzerlerinin verdikleri tercihler üzerinden bir tahmin değeri üretmeye çalışır.

OF sistemlerinin, kullanıcılardan bilgi sahibi oldukları ürünler hakkında bilgi istemesi işlemi direkt bilgi toplama yöntemidir. Bunun dışında, OF sistemleri dolaylı veri toplama yöntemine de başvurabilir. Bu yöntem sayesinde sistem, daha önce de ifade edildiği üzere, kullanıcıların bir ürünü incelerken harcadıkları zaman, bir ürüne tıklama sayısı, ziyaret ettikleri web siteleri tarafından oluşturulan çerezler gibi bilgiler vasıtasıyla kullanıcıları daha iyi analiz etmeye çalışır. Bir kullanıcının bir ürünü defalarca ziyaret etmesi, o ürünü sevdiğini veya satın almak istediğini gösterir. Dolayısıyla sistem, kullanıcı o ürün hakkında fikir belirtmemiş olsa bile ürünü sevdiğini varsayarak tavsiye üretebilir.

Tablo 3.1. *Geleneksel OF sistemlerinde örnek bir kullanıcı-ürün matrisi*

Kullanıcı	Ürün 1	Ürün 2	Ürün 3	Ürün 4	…	Ürün M
Kullanıcı 1	5	3	8	6	…	?
Kullanıcı 2	--	6	5	1	…	--
Kullanıcı 3	3	--	5	--	…	4
Kullanıcı 4	6	3	8	6	…	3
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Kullanıcı N	5	3	8	6	…	8

Tablo 3.1'de bir OF sisteminde toplanan verilerin bir arada bulunduğu örnek bir kullanıcı-ürün matrisi gösterilmektedir. Kullanıcıların tercihleri 1 ile 10 arasında sayılarla ifade edilmiştir. Matriste '—' ile ifade edilen hücreler kullanıcının ilgili ürünle ilgili bir tercihinin olmadığını ifade etmektedir. '?' işareti ile gösterilen hücre hedef kullanıcı olan birinci kullanıcının hakkında tahmin istediği hedef ürünü ifade eder. Birinci kullanıcının verdiği tercihlere bakılarak sistem ürünlere birinci kullanıcıyla birebir aynı tercihlerde bulunan *n*'inci kullanıcıyı komşu olarak belirler. Birinci kullanıcıya hedef ürünle ilgili tahminde bulunurken bu kullanıcının o ürüne vermiş olduğu oy '8' olduğundan sistem hedef ürün için '8' tahmininde bulunacaktır. Örneğimizde komşuluk seçimi için sistem tek bir kullanıcıyı değerlendirmiş olsa da gerçek sistemlerde komşu sayısı 1'den fazladır ve sistemin en kaliteli tahmini üretebilmesi için seçilen komşuluk sayısı deneysel sonuçlarla belirlenir.



Görsel 3.1. *OF Mantığına Genel Bakış*

OF sistemlerinin, talep edilen nesneye tavsiye üretmenin yanı sıra, *top-n* listesi olarak adlandırılan bir özelliği mevcuttur (Sarwar vd., 2001). Sistem, kullanıcılarından tahmin istemelerini beklemeksizin, kullanıcılarına hakkında bilgi sahibi olmadığı tüm ürünlerle ilgili tavsiye üretir. Bu tavsiye üretimi sonucu hesaplanan değerleri; en büyükten en küçüğe sıralayarak, kullanıcıya en çok beğenebileceği *N* adet ürünü sunar. Bu sayede, kullanıcıların sistemi daha etkin kullanabilmesine olanak sağlar. Özetle, OF sistemlerinin tahmin üretimi ve *top-n* listesi tavsiyesi olarak iki temel fonksiyonelliği bulunmaktadır.

3.1.Ortak Filtreleme Çeşitleri

OF sistemleri algoritmik yaklaşımlarına göre iki temel kategoriye ayrılırlar: Model tabanlı yaklaşım ve Hafıza tabanlı yaklaşım. Model tabanlı yaklaşımda tekil değer ayrıştırma, gizli anlamsal modelleme, bayes ağları gibi veri madenciliği veya makine öğrenimi algoritmaları kullanılarak modeller oluşturulur ve bu modeller aracılığıyla kullanıcıya tavsiye üretilir. Hafıza tabanlı yaklaşım ise bu tez çalışmasının çoğunluğunda anlatıldığı üzere kullanıcı oyları kullanılarak tavsiye üretme yaklaşımıdır. OF sistemlerinin ayrıca, kullanıcı tabanlı ve ürün tabanlı yaklaşımları da mevcuttur.

Her iki yaklaşımın da kendine özgü avantajları ve dezavantajları vardır. Örneğin; hafıza tabanlı yaklaşımlar model tabanlı yaklaşımlara göre daha yüksek doğruluk başarımı sağlarken, model tabanlı yaklaşımın ise çevrimiçi performansı hafıza tabanlı sistemlere göre daha başarılıdır. OF sistemi tasarlanırken genellikle model veya hafıza tabanlı yaklaşımlardan biri kullanılırken, bu iki yöntemin kaynaştırıldığı melez yaklaşımlar da literatürde yer almaktadır (Pennock vd., 2000; Lemire ve Maclachlan,

2008). Su ve Khoshgoftaar (2009), çalışmasında hafıza tabanlı, model tabanlı ve melez yaklaşımları tek bir tabloda karşılaştırmıştır. Tablo 3.2'de OF sistemleri çeşitleri, kullanılan teknikleri, avantajları ve dezavantajları ile irdelenmiştir.

Tablo 3.2. *OF Tekniklerine Genel Bir Bakış (Su ve Khoshgoftaar, 2009)*

OF Kategorisi	Örnek Teknikler	Avantajları	Dezavantajları
Hafıza-tabanlı OF	* Kullanıcı/Ürün Tabanlı OF * Kullanıcı/Ürün Tabanlı top-*n* Tavsiyelendirme	* Uygulaması kolay * Yeni veri kolaylıkla eklenebilir. * Tavsiye edilecek ürünün içeriği önemli değildir. * Eşdeğer ürünlerin ölçeklenmesi basittir.	* Sistem performansı kullanıcıların verdiği oylara bağlıdır. * Veri ne kadar seyrekse performans o kadar düşer. * Yeni kullanıcılara tavsiye üretilemez. * Büyük veritabanlarında ölçeklenebirlik sınırlıdır.
Model-tabanlı OF	* Bayes Ağları (Bayesian Belief Network) * Kümeleme (Clustering) * Markov Karar Süreçleri (The Markov Decision Process) * Gizli Anlamsal Yaklaşım (Latent-Semantic Approach) * Boyut İndirgeme Teknikleri	* Seyreklik, ölçeklenebilirlik problemlerinin üstesinden gelinebilir. * Tavsiye üretimi için sezgisel bağıntılar bulunabilir.	* Model oluşturmanın sisteme yükü vardır. * Ölçeklenebilirlik ile tahmin performansı arasında bir denge oluşturmak gerekir. * Özellikle boyut indirgeme tekniklerinde faydalı bilgi kaybı olur.
Melez yaklaşımlar	* İçerik Tabanlı OF * İçerikle Zenginleştirilmiş OF * Hafıza ve Model Tabanlı OF Sistemleri	* OF sistemlerinin problemlerinin üstesinden gelinebilir. * Tahmin performansı arttırılır. * Seyreklik ve gri koyun gibi problemlerin üstesinden kolaylıkla gelinebilir.	* Sistemin uygulanması daha masraflı ve zorludur. * Daha iyi performans için daha çok harici bilgiye ihtiyaç duyulur.

3.2.Ortak Filtrelemenin Zorlukları

OF sistemleri bilgi filtrelemede ve e-ticaret uygulamalarında oldukça başarılı olsa da bu sistemlerin bazı kısıtlamaları mevcuttur. Bu bölümde OF sistemlerinin en önemli problemleri açıklanmıştır.

3.2.1. Doğruluk

Kullanıcıların beklentilerini karşılamakta yetersiz kalan kalitesiz tavsiye sistemleri çevrimiçi satıcıların satış oranlarını olumsuz yönde etkiler. Kullanıcılar tarafından daha çok tercih edilebilmek ve rakiplerinin önüne geçebilmek için çevrimiçi hizmet sağlayıcıları doğru ve kaliteli tahminler üretebilen tavsiye sistemleri kullanmalıdır. Tavsiye sisteminin bir ürün hakkında tahmin üretirken kullanıcının beğenebileceği bir ürün için olumsuz tavsiye üretmesi sistemin saygınlığını önemli ölçüde düşürmese de beğenmeyeceği bir ürünü tavsiye etmesi kullanıcının sisteme olan güvenini düşürür ve satıcının müşteri kaybetmesine neden olur. Bu bağlamda sistemin doğruluğundan ve kaliteli tavsiyeler ürettiğinden emin olunmalıdır.

3.2.2. Ölçeklenebilirlik

OF sistemlerinin, kullanıcılara tavsiye üretebilmek için, veri tabanında yer alan varlıkları tanımaya ve birbirleri ile olan benzerlikleri öğrenmeye ihtiyacı vardır. Bu görevi yerine getirebilmesi için birtakım karmaşık matematiksel ve istatistiki hesaplamalara ihtiyaç duyar. Onlarca veya yüzlerce varlık için bu tür hesaplamalar yapmak basit olsa da, binlerce hatta milyonlarca kullanıcı ve ürünün yer aldığı bir veri tabanında sistemin hızlı ve efektif çalışması oldukça zordur. OF sistemlerinin kullanıcılar tarafından kabul görebilmesi için sistemin çevrimiçi tepki süresinin kabul edilebilir bir sürede gerçekleşmesi gerekmektedir. Bu problem, literatürde, ölçeklenebilirlik problemi olarak adlandırılır.

3.2.3. Veri seyrekliği ve soğuk başlangıç problemi

OF sistemine yeni dâhil olan bir kullanıcı ilk aşamada sistemdeki bir ürün hakkında tercih belirtmediğinden yeni kullanıcıya benzer kullanıcıların bulunabilmesi mümkün değildir. Benzer bir şekilde sisteme yeni bir ürün eklendiğinde, henüz o ürünle ilgili hiçbir kullanıcı tercih belirtmediğinden, sistemin ürünü bir kullanıcıya tavsiye edebilmesi mümkün değildir. Bu durum soğuk başlangıç problemi olarak adlandırılır (Wei vd., 2017).

OF sisteminin güvenilir sonuçlar üretebilmesi için kullanıcıların yeterli sayıda ürünü oylaması, ürünlerin de yeterli sayıda kullanıcı tarafından oylanmış olması gerekmektedir. Kullanıcı birkaç ürüne oy vermedikçe veya bir ürüne yeterli sayıda tercih belirtilmedikçe sistemin özellikle yeni kullanıcılar veya ürünler arasındaki ilişkiyi kurabilmesi mümkün değildir ve bu durum sistemin kesin ve güvenilir tavsiyeler oluşturmasını zorlaştırmaktadır.

3.2.4. Anlamdaşlılık

Veri tabanında yer alan bir ürün dikkat edilmeden farklı isimle yeniden veri tabanına eklenebilir. Örneğin, gıda ürünleri satan bir e-ticaret sitesi veri tabanında, önceden '86916440' barkod numarası ile eklenmiş bir ürünün barkod numarasının üretici firma tarafından '86916235' olarak değiştirildiğini varsayalım. Eski ürün sistemden çıkarılmadan aynı ürün farklı barkod numarası ile eklendiği takdirde aynı ürün hem '86916440' barkod numarası ile hem de '86916235' barkod numarası ile veri tabanında yer alacaktır. Bu durumda kullanıcıların bir kısmı eski ürüne oy vermeye devam ederken bir kısmı da yeni ürüne oy verecektir. Bu durum, sistemin o ürün için sağlıklı tavsiye üretebilmesini engellerken aynı zamanda seyreklik problemini de arttırmaktadır. E-ticaret sitelerinde bu ve benzeri durumlar sıkça yaşandığından bu problem dikkate alınması gereken bir problemdir.

3.2.5. Gri koyun (Gray Sheep) problemi

Sisteme yeni dâhil edilmiş bir kullanıcının veya ürünün ilk etapta soğuk başlangıç problemi mevcut iken oy verilmeye başlandıktan sonra da gri koyun (gray sheep) adı verilen problem ortaya çıkabilir. Sistemin, entelektüel olarak tanımladığı bir grup kullanıcı veya ürün vardır. Bu gruba dâhil olan ürünler sistemin geneli tarafından oy almayan, çok az insan tarafından oy verilmiş ürünlerdir. Benzer bir şekilde yine bu gruba dahil olan kullanıcılar, kullanıcıların genelinin oy verdiği ürünlerden ziyade başka ürünlere oy veren kişilerdir. Sistem, bu kişiler veya ürünler için verilen oylardan mantıklı sonuçlar elde edemeyebilir. Gri koyun problemi bu tür durumlar için ifade edilir. Özetle, bir varlığın, verilen oylardan sonra ortaya çıkan karakteristiğinin hiçbir grupla benzerlik gösterememesi durumunda, sistemin hedef varlığa tavsiye üretememesi problemidir.

3.2.6. Gizlilik

OF sistemlerinin veri tabanları müşterilerin alışveriş alışkanlıkları ve karakteristikleri ile ilgili detaylı bilgiler içermektedir (Elahi vd., 2016). Müşteriler

verilerini çevrimiçi satıcılarla paylaşırken aynı zamanda mahremiyetlerini de paylaşmış olurlar. Bu durum, bazı kötü niyetli kişi veya şirketler tarafından kullanılabilir. Bu durumda kullanıcılar istenmeyen pazarlama stratejileri, şantaj, fiyat ayrımcılığı, devlet gözetimine tabi tutulmak gibi durumlarla karşılaşabilirler (Zhang vd., 2014). Örneğin, kişi siyasi içerikli veya yasak bir ürünü satın almışsa, satıcı veya sisteme sızan birileri tarafından şantaja maruz kalabilir veya devamlı pahalı ürünler satın alan müşteriye satıcı, alım gücünün yüksek olduğunun farkında olarak bir ürünü diğer kullanıcılara sattığı fiyatın daha üstüne satmak isteyebilir. Bu tip tehlikelerin farkında olan müşteriler ürünler hakkındaki tercihlerinin açığa çıkmasını istemediklerinden OF sistemlerinin gizlilik problemlerini çözmesi ve güvenilir olduğunun taahhüdünü verebilmesi gerekir.

3.2.7. Shilling Ataklar

Kötü niyetli kişiler veya şirketler bazı ürünlerin popülaritesini artırmak veya düşürmek amacıyla veri tabanına sahte profiller yerleştirerek sistemi manipüle etmeye çalışabilir. OF sistemleri genel olarak halka açık sistemler olduklarından, sisteme sahte profillerin girmesine engel olamaz fakat birtakım algoritmalarla saldırı profillerini keşfedebilir ve sistemi kötü niyetli kişilerden koruyabilir fakat bu saldırılara karşı tam anlamıyla korunabilmek mümkün değildir ve dolayısıyla bu durum OF sistemlerinin zorlukları arasındadır.

3.3.OF Sistemi Uygulaması: Bir Örnek

Bu bölümde, OF sistemlerinin tavsiye üretimi aşamalarının nasıl gerçekleştirildiği, örnek bir veri matrisi üzerinde gösterilmiştir. Elimizde 6 adet kullanıcı ve 6 adet ürün içeren basit bir veri matrisi olsun. Bu veri kümesinde kullanıcılar, bazı ürünlere 1 ile 5 arasında oy vermiştir. Veri kümesinde, kullanıcıların oy belirtmediği hücreler '--' ile ifade edilmiştir. Hedef kullanıcı olan *Ferhat*, 6. ürün olan *Fular* için sistemden tavsiye istemektedir. Dolayısıyla, ilgili kullanıcının hedef ürününe ait hücre '?' ile gösterilmektedir. Veri matrisinin anlık görüntüsü Tablo 3.3'te yer almaktadır.

Tablo 3.3. *OF sistemi tavsiye üretimi için oluşturulmuş örnek bir kullanıcı-ürün matrisi*

	Atkı	**Bere**	**Cüzdan**	**Dizlik**	**Eldiven**	**Fular**
Ahmet	--	3	--	4	1	2
Burçin	2	4	1	--	3	1
Ceyda	--	5	3	2	4	--
Defne	5	--	2	1	--	3

Emin	3	5	--	1	2	4
Ferhat	4	4	--	2	3	**?**

OF Sistemi hedef ürüne tavsiye üretimi için şu aşamaları uygulayacaktır:

- Benzerlik Hesabı
- Komşuluk Seçimi
- Tahmin Üretimi

3.3.1. Ön aşama: bilgi toplama

OF sistemlerinin tavsiye üretimi için gerekli aşamalarının uygulanabilmesi için sistemin birtakım önbilgilere ihtiyacı vardır. Bu bilgiler benzerlik hesabında kullanılacak ölçütün ihtiyacı olan veriler, tahmin üretiminde kullanılacak yöntemin gereksinimleri gibi bilgilerdir. Örneğin, benzerlik hesaplamasında korelasyon tabanlı bir yaklaşım uygulanacaksa ve tahmin üretiminde Pearson Correlation Coefficient (PCC) kullanılacaksa, sistemin kullanıcıların oy ortalamaları, standart sapmaları ve z-skor verileri gibi verilere ihtiyacı vardır. Benzerlik hesaplamasında uzaklık metrikleri kullanılan bir sistemde ise bu tip verilere ihtiyaç olmaz. Bu örneğimizde OF sistemi tavsiye üretimi için PCC metriğini kullanacağı için, öncelikle ihtiyacımız olan verileri derleyeceğiz.

Ön aşamada sistem, veri kümesinde yer alan kullanıcı oylarının aritmetik ortalamalarını, standart sapmalarını ve z-skor değerlerini hesaplayıp kaydetmektedir. Kullanıcı oylarının aritmetik ortalaması hesabı denklem 3.1'de verilmiştir (Medhi, 1992).

$$\bar{r}_a = \frac{1}{n}\sum_{i=1}^{n} a_i = \frac{a_1 + a_2 + \cdots + a_n}{n}, \tag{3.1}$$

$\bar{r}_a$, *a* kullanıcısının vermiş olduğu oyların aritmetik ortalamasını, *n,* kullanıcının ürünlere vermiş olduğu oy sayısını, a_i ise *a* kullanıcısının *i*'inci verdiği sayısal oy değerini ifade eder. Bu durumda, kullanıcıların vermiş olduğu oylara göre aritmetik ortalamaları şu şekilde hesaplanır:

$\bar{r}_{Ahmet} = \frac{3+4+1+2}{4} = 2,5$

$\bar{r}_{Burçin} = \frac{2+4+1+3+1}{5} = 2,2$

$\bar{r}_{Ceyda} = \frac{5+3+2+4}{4} = 3,5$

$\bar{r}_{Defne} = \frac{5+2+1+3}{4} = 2,75$

$\overline{r}_{Emin} = \frac{3+5+1+2+4}{5} = 3$

$\overline{r}_{Ferhat} = \frac{4+4+2+3}{5} = 3,25.$

Kullanıcıların standart sapmaları, vermiş oldukları oylar ve aritmetik ortalamaları yardımıyla bulunur. Standart sapma formülü denklem 3.2'de verilmiştir (Medhi, 1992).

$$\sigma_a = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(a_i - \overline{a}_i)^2}, \tag{3.2}$$

Verilen formülde σ_a, *a* kullanıcısının standart sapmasını, *n,* kullanıcının ürünlere vermiş olduğu oy sayısını, $\overline{a}_i$ ise *a* kullanıcısının aritmetik ortalamasını ifade eder. Bu durumda, kullanıcıların standart sapmaları şu şekilde hesaplanır:

$\sigma_{Ahmet} = \sqrt{\frac{1}{3}\sum_{i=1}^{4}[(3-2,5)^2 + (4-2,5)^2 + (1-2,5)^2 + (2-2,5)^2]} = 1,29$

$\sigma_{Burçin} = \sqrt{\frac{1}{4}\sum_{i=1}^{5}[(2-2,2)^2 + (4-2,2)^2 + (1-2,2)^2 + (3-2,2)^2 + (1-2,2)^2]} = 1,30$

$\sigma_{Ceyda} = \sqrt{\frac{1}{3}\sum_{i=1}^{4}[(5-3,5)^2 + (3-3,5)^2 + (2-3,5)^2 + (4-3,5)^2]} = 1,29$

$\sigma_{Defne} = \sqrt{\frac{1}{3}\sum_{i=1}^{4}[(5-2,75)^2 + (2-2,75)^2 + (1-2,75)^2 + (3-2,75)^2]} = 1,70$

$\sigma_{Emin} = \sqrt{\frac{1}{4}\sum_{i=1}^{5}[(3-3)^2 + (5-3)^2 + (1-3)^2 + (2-3)^2 + (4-3)^2]} = 1,58$

$\sigma_{Ferhat} = \sqrt{\frac{1}{3}\sum_{i=1}^{4}[(4-3,25)^2 + (4-3,25)^2 + (2-3,25)^2 + (3-3,25)^2]} = 0,95.$

Standart sapması ve aritmetik ortalaması bilinen bir kullanıcının vermiş olduğu oyların z-skor değeri denklem 3.3'te verilen formül ile hesaplanır (Gravetter vd., 2020).

$$z_{a,i} = \frac{r_{a,i} - \overline{r}_a}{\sigma_a}, \tag{3.3}$$

$z_{a,i}$, *a* kullanıcısının *i*'inci ürününün z-skorunu, $r_{a,i}$, *a* kullanıcısının *i*'inci ürüne verdiği sayısal oy değerini, $\overline{r}_a$, *a* kullanıcısının vermiş olduğu oyların aritmetik ortalamasını ve σ_a ise *a* kullanıcısının standart sapmasını ifade eder. Bu bilgilere göre z-skor değerleri hesaplanarak Tablo 3.4'te verilen z-skoru matrisi elde edilir.

Tablo 3.4. *Örnek kullanıcı-ürün matrisinin z-skoru verileri ile doldurulmuş hali*

	Atkı	**Bere**	**Cüzdan**	**Dizlik**	**Eldiven**	**Fular**
Ahmet	--	0,38	--	1,16	-1,16	-0,38
Burçin	-0,15	1,38	-0,92	--	0,61	-0,92
Ceyda	--	1,16	-0,38	-1,16	0,38	--

Defne	1,31	--	-0,43	-1,02	--	0,14

3.3.2. Aşama 1: benzerlik hesabı

OF sisteminin, ön aşamada ihtiyacı olan bilgiler hesaplandıktan sonra 1. aşama olan benzerlik hesabı yapılmaktadır. Bu aşamada hedef ürün veya kullanıcının, diğer ürün veya kullanıcılar ile olan benzerlikleri, birtakım metrikler aracılığıyla bulunur. Bu aşamada, daha önce de belirtildiği üzere PCC metriği kullanılmıştır. PCC ölçütü ile benzerlik hesabı formülü denklem 3.4'te verilmiştir (Herlocker vd., 1999).

$$w_{a,u} = \frac{\sum_{i=1}^{m}\left(r_{a,i} - \bar{r}_i\right) \times \left(r_{u,i} - \bar{r}_i\right)}{\sigma_a \times \sigma_u}, \tag{3.4}$$

$w_{a,u}$ hedef kullanıcı a ile komşu kullanıcı u arasındaki benzerliği, m hedef kullanıcı a ile komşu kullanıcı u'nun ortak olarak oyladığı ürün sayısını, $r_{a,i}$ hedef kullanıcı a'nın hedef ürüne vermiş olduğu oyu, $r_{u,i}$ komşu kullanıcı u'nun hedef ürüne vermiş olduğu oyu, $\bar{r}_i$ hedef ürüne verilen oyların ortalamasını, σ_a ve σ_u ise sırasıyla, hedef kullanıcı a ile komşu kullanıcı u'nun standart sapmalarını ifade eder. Belirtilen denklem, kullanıcı tabanlı sistemlerde kullanıcılar arasındaki benzerliği ortaya çıkarmakta kullanılır. Aynı yaklaşımla, iki ürün arasındaki benzerliği de hesaplamak mümkündür. Denklem 3.4'te verilen formül, denklem 3.5'te verilen formülle özetlenebilir.

$$w_{au} = \sum_{i=1}^{m} z_a \times z_u \tag{3.5}$$

Bu durumda, hedef kullanıcı *Ferhat*'ın diğer kullanıcılarla olan benzerliği şu şekilde hesaplanır:

$$\boldsymbol{w_{Ferhat,Ahmet} = (0,78 \times 0,38) + (-1,30 \times 1,16) + (-0,26 \times -1,16) = -0,91}$$
$$\boldsymbol{w_{Ferhat,Burçin} = (0,78 \times -0,15) + (0,78 \times 1,38) + (-0,26 \times 0,61) = 0,80}$$
$$\boldsymbol{w_{Ferhat,Ceyda} = (0,78 \times 1,16) + (-1,30 \times -1,16) + (-0,26 \times 0,38) = 2,32}$$
$$\boldsymbol{w_{Ferhat,Defne} = (0,78 \times 1,31) + (-1,30 \times -1,02) = 2,36}$$
$$\boldsymbol{w_{Ferhat,Emin} = (0,78 \times 0) + (0,78 \times 1,26) + (-1,30 \times -1,26) + (-0,26 \times -0,63) = 2,80.}$$

3.3.3. Aşama 2: komşuluk seçimi

Bu aşamada OF sistemi, 1.aşamada bulunan benzerlik değerleri aracılığıyla komşulukları belirler. Komşuluk belirleme aşamasında k en yakın komşu (k-NN), eşik (threshold) ve mutlak eşik (absolute threshold) gibi yaklaşımlar mevcuttur. Bu

örneğimizde *k*-NN yaklaşımı kullanılmıştır. *K* değeri 4 olarak belirlenmiş ve buna göre hedef kullanıcı *Ferhat* için bir önceki aşamadaki verilere göre şu şekilde seçilmiştir:

$$w_{Ferhat,Emin} > w_{Ferhat,Defne} > w_{Ferhat,Ceyda} > w_{Ferhat,Burçin} > w_{Ferhat,Ahmet}$$

K = {Emin, Defne, Ceyda, Burçin}.

3.3.4. Aşama 3: tahmin üretimi

OF sistemi son aşamada, hedef ürüne tahmin üretimi için, belirlenmiş olan 4 adet komşunun oyları yardımıyla tahmin üretimi gerçekleştirmektedir. Bu aşamada OF sisteminin, hedef kullanıcının komşularının ağırlıkları ve z-skor değerleri yardımıyla oluşturacağı tahmin üretimi formülü denklem 3.6'da verilmiştir.

$$p_{aq} = \bar{r}_a + \sigma_a \times \frac{\sum_{i=1}^{m} w_{au} \times z_{uq}}{\sum_{i=1}^{m} w_{au}} \tag{3.6}$$

Sonuç olarak, hedef kullanıcı *Ferhat*'ın hedef ürün *Fular*'ı beğenip beğenmeyeceğini belirten OF tahmini şu şekilde hesaplanarak işlem tamamlanmış olur.

$$p_{aq} = 3,25 + 0,95\, x\, \frac{(2,80 \times 0,63) + (2,36 \times 0,14) + (0,80 \times -0,92)}{(2,80 + 2,36 + 0,80)} = 3,47.$$

Bu durumda OF sisteminin *Ferhat* adlı kullanıcı için *Fular* adlı ürüne üretmiş olduğu tahmin 5 üzerinden 3,47'dir. Bu da *Ferhat*'ın daha önce diğer ürünlere belirtmiş olduğu oylara bakılarak, komşularının verdiği oylar yardımıyla *Fular* adlı ürünü beğeneceği anlamına gelmektedir. Örnek kullanıcı-ürün matrisinin son hali Tablo 3.5'te verilmiştir.

Tablo 3.5. *Tavsiye üretimi sonrası kullanıcı-ürün matrisinin görüntüsü*

	Atkı	Bere	Cüzdan	Dizlik	Eldiven	Fular
Ahmet	--	3	--	4	1	2
Burçin	2	4	1	--	3	1
Ceyda	--	5	3	2	4	--
Defne	5	--	2	1	--	3
Emin	3	5	--	1	2	4
Ferhat	4	4	--	2	3	**3,47**

3.4.OF Sistemleri için Yapılan Çalışmalar

Literatürde OF sistemleri üzerine yapılmış bazı çalışmalar özet olarak aşağıda belirtilmiştir.

İlk ortaya atıldığı, 1992 yılından bu yana OF sistemleri ile ilgili yüzlerce çalışma yapılmıştır. Tapestry projesi'nde (Goldberg vd., 1992) ilk kez kullanıldığında, haber grupları için oluşturulmuş çok sayıda elektronik posta ve mesajın filtrelenmesi için yardımına başvurulan OF sistemleri, daha sonra makale, kitap, film, şehir, otel, restoran gibi birçok alanda tavsiye üretimi için kullanılmıştır. Chen vd. (2005) OF sistemlerini kitap tavsiye etmesi için kullanmıştır. Cleger-Tamayo vd. (2011) kullanıcıların ilgisini çekebilecek haberleri top-*n* listesi olarak sunan bir OF sistemi geliştirmiştir. Sun vd. (2014) kullanıcılara tavsiye üretirken, onların sosyal ilişkilerinden faydalanan, sosyal ağlar için geliştirilmiş bir tavsiye sistemi oluşturmuştur. Benzer bir şekilde, Zheng ve Li (2011), sosyal ağlarda sıkça kullanılan Etiket (Tag) ve Zaman bilgilerini kullanan bir OF sistemi geliştirmiştir. Jia vd. (2015), OF sistemlerini, turistler için kişiye özgü cazibe merkezleri tavsiyesinde kullanmıştır. Li vd. (2007) OF sistemlerini müzik tavsiye üretimi için kullanmıştır. Müzik alanında tavsiye sistemi kullanımına bir başka örnek ise Tan vd. (2011) tarafından önerilen çalışmadır. Bunların yanı sıra OF sistemlerinin en yaygın örnekleri, film tavsiye üretiminde karşımıza çıkmaktadır. Ungar ve Foster (1998), Shan ve Banerjee (2010), Chen vd. (2015) bunlardan yalnızca birkaçıdır.

İlk önerilmeye başlandığı dönemlerde OF sistemlerinde tahmin üretimi yapılırken, yalnızca kullanıcıların ürünler için belirtmiş oldukları sayısal oylar hesaba katılmıştır. Hafıza tabanlı olarak adlandırılan bu sistemler, uygulanması kolay (easy-to-implement), hızlı ve efektif sonuçlar veren sistemlerdir. Kullanıcı-tabanlı ve ürün-tabanlı olmak üzere ikiye ayrılır. Hafıza tabanlı sistemlerin ortaya çıkışı her ne kadar 90'lı yıllara dayansa da, hızlı ve efektif oluşu dolayısıyla bugün de araştırmacılar tarafından kullanılmaktadır. Sarwar vd. (2001), Linden vd. (2003), Gao vd. (2011), Li vd. (2016) çalışmalarında ürün-tabanlı OF sistemlerini analiz etmiştir. Zhao ve Shang (2010), Jia vd. (2015) ve Koohi ve Kiani (2016) ise çalışmalarında kullanıcı-tabanlı OF sistemlerini analiz etmeyi tercih etmişlerdir. Daha iyi sonuçlar üretmek adına kullanıcı ve ürün tabanlı sistemleri bir araya getiren çalışmalar da yapılmıştır (Wang vd., 2006; Thakkar vd., 2019).

Hafıza tabanlı OF sistemleri her ne kadar başlangıçta uygulanması basit ve hızlı sistemler olsa da sisteme yeni ürün ve kullanıcıların eklenmeye devam etmesi ile birlikte, önceki bölümde de ifade edilen ölçeklenebilirlik problemi ciddi boyutlara ulaşmaya başlar. Bu sorunun üstesinden gelmek isteyen araştırmacılar, model tabanlı yaklaşımlar adı verilen yaklaşımlar önermiştir. Bu yaklaşım ile sistem, dâhil edilen her yeni varlık

için benzerlik ve komşuluk hesaplamalarını tekrar tekrar yapmak yerine; hesaplamalarını belirli aralıklarla yaparak sistemin çok daha hızlı çalışmasını sağlar. Model oluşturmak için lineer cebir, yapay zekâ, gizli sınıflama (latent class), sinir ağları veya kümeleme gibi yöntemler kullanılır. Goldberg vd. (2001), PCA (Principal Component Analysis - Temel Bileşen Analizi) adı verilen yöntem aracılığıyla kullanıcıları gruplayan bir model tabanlı yaklaşım önermiştir. Chee vd. (2001), kümeleme yöntemi ile benzer kullanıcıları gruplayarak tavsiye üretimi yapan ve bu sayede hafıza tabanlı sistemlere göre daha hızlı sonuç veren bir OF sistemi geliştirmiştir. Jiang vd. (2015), sosyal kullanıcılara POI (Point of Interests) adı verilen ihtiyaç noktaları için tavsiye üreten model tabanlı OF yaklaşımı önermiştir. Tavsiye üretirken yalnızca geo-etiketleri kullanmakla kalmamış, aynı zamanda ihtiyaç noktaları için kültürel, mimarı, kent simgesi gibi çıkarımlarda da bulunarak algoritmik performansı artırmayı hedeflemişlerdir. Ortega vd. (2016), matris faktörizasyonu yardımıyla grup tavsiyeleri üreten bir OF sistemi yaklaşımında bulunmuştur.

OF sistemleri literatürde yalnızca algoritmik yaklaşımlarına göre incelenmemiştir. OF sistemlerinin zorluklarının üstesinden gelmek için de çok sayıda spesifik çalışma yapılmıştır. Bu çalışmaların en yoğun olarak yapıldığı problemler veri seyrekliği (data sparsity), ölçeklenebilirlik (scalability), anlamdaşlılık (synonymy) ve shilling ataklardır. Bobadilla vd. (2012), sinirsel öğrenme temelli optimizasyon kullanılarak oluşturulmuş bir benzerlik ölçütü önererek soğuk başlangıç probleminin üstesinden gelmeye çalışmıştır. Fernández-Tobías vd. (2016) de kullanıcıların kişiliklerini ortaya çıkaran birtakım harici bilgiler yardımıyla soğuk başlangıç probleminin üstesinden gelmeye çalışmıştır. Zhang vd. (2005) tekil değer ayrışımı kullanılarak oluşturulan bir yöntem ile sistemin, seyreklik ve ortalama hesap yükünü düşürmeyi amaçlamıştır. Polat ve Du (2003), rasgele hale getirilmiş sarsım teknikleri yardımıyla gizliliği koruyan bir OF sistemi geliştirmiştir. Bu sayede verdikleri oyların başkaları tarafından öğrenilmesini istemeyen kullanıcıların oyları veri toplayıcılarının erişimine engellenmiştir. Chirita vd. (2005), shilling atak profillerini tespit eden bir algoritma yardımıyla shilling atak problemini ortadan kaldırmayı amaçlamıştır.

4. ÇOKLU-ÖLÇÜTLÜ ORTAK FİLTRELEME SİSTEMLERİ

Geleneksel OF sistemleri kullanıcıların ürünlere verdikleri oyları kullanarak tavsiye üretmektedir. Bu tür sistemlerde kullanıcılar, ürünler hakkındaki genel görüşlerini sayısal rakamlar (örneğin; 10 üzerinden 8) veya dereceler (örneğin; 5 yıldız üzerinden 4 yıldız) ile ifade ederler. Ancak bazı durumlarda ürünün geneline oy vermek o ürün hakkındaki tercihlerimizi tam anlamıyla ortaya koyamayabilir. Örneğin, insanlar bir film hakkında tercihlerini belirtirken izledikleri filmi farklı açılardan değerlendirir ve filmin yönetmeni, kurgusu, oyunculukları, görsel efektleri vb. unsurlarını ayrı ayrı değerlendirirler. Filmin genel olarak beğenilip beğenilmediğinin anlaşılmasında, bu değerlendirmeler belirleyici olur. Aynı filmi izleyen iki farklı kullanıcıdan biri filmi oyuncu kadrosundan dolayı beğenirken, diğeri kurgusundan dolayı beğenmiş olabilir. Ancak, ikisi de filmin genel anlamda değerlendirdiğinde, aynı oyları vermiş olabilirler. Benzer bir şekilde bir otelde konaklayan iki müşteriden biri otelin servisini, diğeri ise oteli bulunduğu konumdan dolayı tercih etmiş olabilir. Bu durumda geleneksel filtreleme sistemleri, kullanıcılar aslında farklı tercihlere sahipken bu kullanıcıları komşu olarak belirleyebilir ve bu durum sistemin kaliteli tahminde bulunmasına engel olabilir. Bu problemin ortadan kaldırılması için tavsiye sistemleri, çoklu-ölçütlü oylarla birleştirilmiştir (Adomavicius ve Tuzhilin, 2005; Kermany ve Alizadeh, 2017).

Çoklu-Ölçütlü Ortak Filtreleme (ÇÖOF) sistemleri, kullanıcıların ürünleri farklı açılardan değerlendirmesine imkân sağlayarak geleneksel OF sistemlerine göre daha kişiselleştirilmiş ve daha kaliteli tavsiyeler sunar. Bu sistemlerde geleneksel OF sistemlerinde kullanılan Kullanıcı – Ürün matrisine ek olarak Ölçüt boyutu da bulunmaktadır. Sistem tarafından ürünler için k adet ölçüt belirlenir ve kullanıcılardan bu ölçütler için tercih belirtmesi istenir. Kullanıcıların sistem tarafından belirtilen ölçütlerin dışında, ürünlerin bütününe tercihlerini belirtebilmeleri için ayrıca bir ölçüt daha bulunur. Bu ölçüt, *genel değerlendirme* ölçütü veya *ortalama değerlendirme* ölçütü şeklinde tanımlanabilir. Genel tercih ölçütü ile birlikte kullanıcılar bir ürünün k+1 ölçütü ile ilgili tercihlerini belirtirler. Sistem bu tercihleri üç boyutlu *Kullanıcı* × *Ürün* × *Ölçüt* matrisinde toplar ve bu verileri kullanarak tavsiye üretir. Tablo 4.1'de ÇÖOF sistemlerinde örnek bir matrisin iki boyutlu gösterimi bulunmaktadır.

Tablo 4.1. *ÇÖOF sistemlerinde örnek bir kullanıcı-ürün-kriter matrisi*

	Ürün 1	**Ürün 2**	**Ürün 3**	**Ürün 4**	**…**	**Ürün M**
Kullanıcı 1	$5_{4,4,6,6}$	$3_{2,2,4,4}$	$8_{7,7,9,9}$	$6_{4,4,8,8}$	…	?
Kullanıcı 2	$-_{-,-,-,-}$	$6_{4,4,8,8}$	$5_{5,5,5,5}$	$1_{1,1,1,1}$	…	$-_{-,-,-,-}$
Kullanıcı 3	$3_{3,3,3,3}$	$-_{-,-,-,-}$	$5_{4,4,6,6}$	$-_{-,-,-,-}$	…	$4_{3,3,4,4}$
Kullanıcı 4	$6_{5,5,7,7}$	$3_{2,2,4,4}$	$8_{7,7,9,9}$	$6_{4,4,8,8}$	…	$3_{4,4,2,2}$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Kullanıcı N	$5_{6,6,4,4}$	$3_{4,4,2,2}$	$8_{9,9,7,7}$	$6_{8,8,4,4}$	…	$8_{7,7,9,9}$

Kullanıcılar, ürünlerin dört adet alt ölçütü ve geneli için tercihlerini belirtmişlerdir. Tercihlerde ilk sırada ve büyük punto ile gösterilen değerler genel tercihleri, diğer dört tercih ürünün ölçütleri için sırasıyla verilen tercihleri temsil etmektedir. Bu matrisin Tablo 3.1'de belirtilen geleneksel OF sistemine ait kullanıcı-ürün matrisinden farkı, ölçütler için verilen tercihlerin de yer almasıdır. Tıpkı diğer sistemde olduğu gibi bu sistemde de komşu sayısının 1 olduğunu ve birinci kullanıcının *m*'inci ürünü için tahmin istendiğini varsayalım. Geleneksel OF sisteminde tekil tercihleri kullanan sistem *n*'inci kullanıcıyı komşu olarak belirlemiş ve komşunun *m*'inci ürüne vermiş olduğu '8' değerini tahmin olarak belirtmişti. Ancak bu örnekte sistem, kullanıcıların ölçütlere vermiş oldukları oylara bakarak aslında birinci kullanıcı ile *n*'inci kullanıcının çok da benzer kullanıcılar olmadığını görebilmektedir. Ölçütlere verilen tercih değerleri gözlemlendiğinde hedef kullanıcımız olan birinci kullanıcı ile *n*'inci kullanıcı farklı tercihler belirtmişler ve ürünlerin geneline tamamen aynı tercihlerde bulunmuş olsalar da farklı tercihlere sahip kullanıcılardır. Bu örnekte dördüncü kullanıcımız birinci kullanıcıya en yakın tercihlerde bulunan kullanıcı olduğundan komşu olarak dördüncü kullanıcı seçilir ve dördüncü kullanıcı *m*'inci ürüne '3' değeri verdiğinden sistem hedef kullanıcımıza tahmin olarak '3' değeri sunacaktır.

ÇÖOF sistemlerinin çalışma prensibi genel olarak geleneksel tekil ölçütlü OF sistemlerine benzerdir. İki sistem arasındaki tek fark genel tercihler için tahmin üretirken yapılan yaklaşımlardır. Ürünlerin ölçütleri için yapılan komşuluk hesabı ve tahmin üretme işlemi geleneksel OF ile yapılır. Genel tercihler için ise genellikle iki farklı yaklaşım uygulanır: Benzerlik tabanlı yaklaşım ve toplama fonksiyonu tabanlı yaklaşımlar (Adomavicius ve Kwon, 2007).

4.1. Benzerlik Tabanlı ÇÖOF Sistemleri

Benzerlik tabanlı ÇÖOF sistemleri, ölçütlerin her biri için bağımsız olarak geleneksel OF sistemlerinde uygulanan prosedürleri uygulayarak ölçüt bazında benzerlikleri hesaplar ve genel tercihler için kullanılacak benzerlik hesabı için de ölçütler için üretilen benzerlikleri kullanarak tahmin üretir. Adomavicius ve Kwon (2007) genel tercihlerde kullanıcılar arasındaki benzerliğin hesaplamasında iki farklı yaklaşımda bulunmuşlardır: Ölçütler için üretilen benzerlik değerlerinin ortalaması veya minimum benzerliğin seçilmesi. Minimum benzerliğin seçilmesindeki amaç, en kötü senaryoda sistemin performansını tespit etmek iken, benzerliklerin ortalamasının alınmasındaki amaç ise sistemin ortalama performansını tespit etmektir. Bu yaklaşıma göre iki kullanıcı u ve u' arasındaki ortalama benzerlik hesabı denklem 4.1'de verilmiştir.

$$\text{Bnz}_{\text{ort}}(u, u') = \frac{1}{k+1} \sum_{i=0}^{k} \text{Bnz}_i(u, u') \tag{4.1}$$

Formülde k ile belirtilen değer kriter sayısını, $\text{Bnz}_i(u, u')$ ile ifade edilen değer i'inci kriter için kullanıcı u ve u' arasındaki benzerliği ifade eder. Kullanıcı u ve u' arasındaki minimum benzerlik hesabı ise denklem 4.2'de verilmiştir.

$$\text{Bnz}_{\text{min}}(u, u') = \min_{i=0,\ldots,k} \text{Bnz}_i(u, u') \tag{4.2}$$

Bilindiği üzere geleneksel OF sistemlerinde benzerlik hesabı yapılırken uzaklık metrikleri (distance metrics) adı verilen birtakım istatistiki hesaplamalar yapılır. Adomavicius ve Kwon (2007) çoklu boyutlu uzaklık metrikleri yardımıyla benzerlik hesaplaması yapılabileceğini önermiştir. Bunun için, çalışmalarında, Manhattan, Öklit veya Chebyshev çoklu boyutlu uzaklık metriklerinden herhangi birinin kullanılabileceğini önermişlerdir. Çalışmalarında, kullanıcı u ve u' arasındaki uzaklık denklem 4.3'te özetlenmiştir.

$$d_{\text{kullanıcı}}(u, u') = \frac{1}{|I(u,u')|} \sum_{i \in I(u,u')} d_{\text{oy}}(R(u,i)R(u', i)) \tag{4.3}$$

Formülde d harfi ile ifade edilen değer İngilizce tanımıyla *distance* olan uzaklık ifadesinin kısaltmasını, $I(u, u')$ değeri ile ifade edilen değer kullanıcı u ve u''nun ortak olarak oy verdiği ürün sayısını, $R(u, i)$ ve $R(u', i)$ ifadeleri ise sırasıyla kullanıcı u ve u''nun belirli bir ürünün her bir kriterine verdiği oyları içeren veri kümelerini ifade eder. Belirtilen formül yardımıyla hesaplanan, kullanıcı u ve u' arasındaki uzaklık son olarak denklem 4.4'te verilen formül aracılığıyla benzerlik değerine dönüştürülür.

$$\text{Bnz}(u, u') = \frac{1}{1+d_{\text{kullanıcı}(u,u')}} \tag{4.4}$$

Bu noktada bilinmesi gereken bir ince ayrım bulunmaktadır: Uzaklık ölçütü ile benzerlik ölçütü birbirleri ile zıt anlamlı olarak ilişkilidir. İki kullanıcı arasındaki uzaklık ne kadar küçükse aralarındaki benzerlik o kadar büyüktür. Benzer bir şekilde iki kullanıcı arasındaki uzaklık ne kadar büyükse aralarındaki benzerlik te bir o kadar küçüktür. İki kullanıcı birbirleri ile her anlamda özdeş ise aralarındaki uzaklık değeri 0 ve benzerlik değeri 1'dir. İki kullanıcı arasında hiçbir benzerlik yok ise bu durumda bu iki kullanıcının uzaklık değerleri 1, benzerlik değerleri ise 0 olacaktır.

4.2. Toplama Fonksiyonu Tabanlı ÇÖOF Sistemleri

Bu sistemlerin temelinde yatan düşünce ürünün ölçütleri ile geneline verilen tercihler arasında bir ilişki olduğu varsayımıdır. Her bir ölçütün genel tercihler üzerinde belirli oranlarda etkisinin bulunduğu düşünülür. Bu yaklaşımda, genel tercihlerin belirtildiği kriter, bir başka bağımsız kriter gibi düşünülmez; diğer kriterlerin *f* ile ifade edilen bir toplama fonksiyonu olarak değerlendirilirler:

$$r_0 = f(r_1, r_2, \dots, r_k) \tag{4.5}$$

Toplama fonksiyonu tabanlı ÇÖOF sistemleri üç aşamalıdır:

1. Herhangi bir tavsiye tekniği kullanılarak *k* adet kriter için tavsiye üretimi
2. Alan uzmanı, istatistiki yöntemler veya makine öğrenimi gibi teknikler aracılığıyla bir toplama fonksiyonu çıkarımı
3. Üretilen tahminler ve çıkarımı yapılan toplama fonksiyonu aracılığıyla genel tercih kriteri için tahmin üretimi

Birinci aşamada, ortam *k* adet tek boyutlu sistem olarak düşünülüp her bir kriter için ayrı ayrı tavsiye sistemi oluşturulmaktadır. Bu noktada dikkat edilmesi gereken nokta, tavsiye üretimi aşamasında yalnızca OF sistemlerinin kullanılmasının zorunlu olmadığı, içerik tabanlı filtreleme veya melez yaklaşımlarla da sistemin tavsiye üretebilmesi özgürlüğüne sahip olunduğudur.

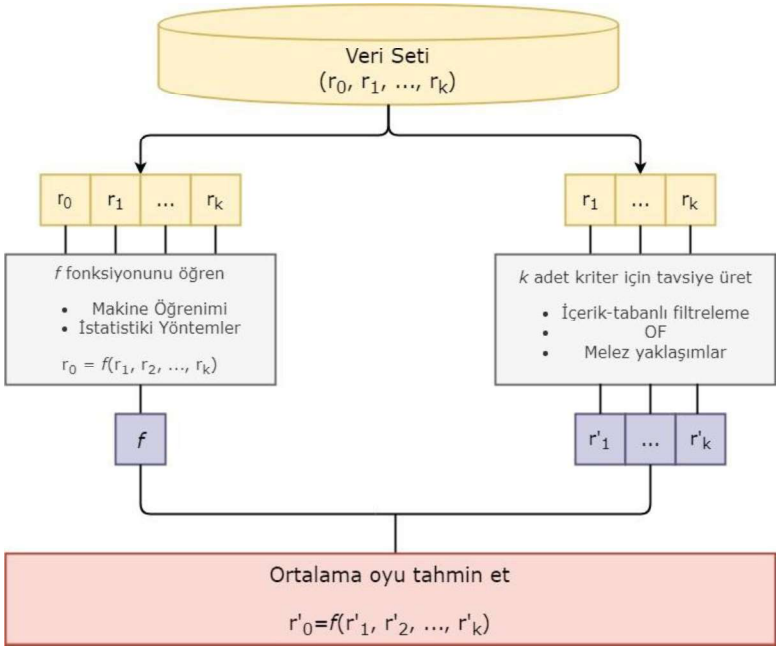
İkinci aşama toplama fonksiyonunun çıkarımı işlemidir. Bu aşamada, her bir kriterin genel tercih kriteri üzerinde nasıl bir öneminin olduğu ortaya çıkarılır. Bu bağlamda en temel yaklaşım, *Alan Uzmanı* olarak ifade edilen ve sistem tarafından ön tanımlanmış sabit oran yaklaşımıdır. Tüm kriterlere eşit ağırlık vermek ve kriterlerin

ortalamasını almak gibi yaklaşımlar bu yaklaşıma bir örnektir (Adomavicius ve Kwon, 2007). Bir diğer ve oldukça etkili bir yaklaşım ise, regresyon analizi gibi bazı istatistiksel yöntemlerin kullanımıdır (Jannach vd., 2012; Bilge ve Kaleli, 2014; Nilashi vd., 2016). Son olarak, yapay sinir ağları gibi makine öğrenimi teknikleri ile de toplama fonksiyonu çıkarımı yapmak mümkündür (Nilashi vd., 2014).

Bu tez çalışmasında önerilen toplama fonksiyonlu ÇÖOF sistemi, istatistiki yöntemle oluşturulmuş sistemlere bir örnektir. Bu sisteme göre geleneksel OF sistemi aracılığıyla kriterler için üretilen tahminler, Kaba Kümeler Teorisi (KKT) aracılığıyla ortaya çıkarılan kriterlerin ağırlıkları, bir lineer regresyon yaklaşımı ile denklem 4.6'da verilen formülde bir araya getirilmiş ve genel tercihler için tahmin üretimi gerçekleştirilmiştir:

$$r_0 = w_1 r_1 + w_2 r_2 + \cdots + w_k r_k + c, \tag{4.6}$$

Formülde *w* ile ifade edilen ve 1'den *k*'ya kadar olan değerler kriterler için sistem tarafından bulunan ağırlıkları, *r* ile ifade edilen değerler her bir kriter için üretilen tahminleri, *c* ise verilerden elde edilen sabit bir değeri ifade eder. *K* adet ölçüt için her bir ölçütün oranı, o ölçüt için verilen tercihle çarpılır ve toplanarak o kullanıcı için genel tercih değeri hesaplanır. Görsel 4.1'de toplama fonksiyonu tabanlı ÇÖOF sistemlerinin genel konsepti yer almaktadır.



Görsel 4.1. *Toplama fonksiyonu tabanlı ÇÖOF yaklaşımına genel bakış*

4.3. ÇÖOF Sistemleri için Yapılan Çalışmalar

"Çoklu-ölçütlü" ifadesi literatürde yalnızca ÇÖOF sistemleri için kullanılmamıştır. Bu ifadeye literatürde, Çoklu-Ölçütlü Karar Verme (ÇÖKV) mekanizmaları ve Çoklu-Ölçütlü Tavsiye Sistemleri (ÇÖTS) adı altında pek çok kez başvurulmaktadır. Bilindiği üzere OF sistemleri tavsiye sistemlerinin, yalnızca sayısal oylarla ilgilenen bir alt dalıdır ve ÇÖOF sistemleri dışında da çoklu-ölçütlü verilerden istifade eden sistemler mevcuttur. Bunun yanı sıra, birtakım problemlerin üstesinden gelebilmek için ÇÖKV mekanizmalarından yararlanan ÇÖOF sistemleri de mevcuttur. Sonuç olarak, öncelikle bu kavramları ilgilendiren çalışmalardan bahsetmek ÇÖOF sistemlerini anlamamızda bize fayda sağlayacaktır.

ÇÖKV, özetle, belirli sayıda ölçüt içerisinden en faydalı ölçütlerin belirlenmesi işlemidir. Literatürde, bu bağlamda için önerilmiş çok sayıda ÇÖKV mekanizması mevcuttur ve en popüler olanları AHP - Analytic Hierarchy Process (Saaty, 1986; Hinduja ve Pandey, 2018), ANP - Analytic Network Process (Saaty, 1996; Pamučar vd., 2017), TOPSIS – Technique for Order of Preference by Similarity to Ideal Solution (Hwang vd., 1993; Joshi ve Kumar, 2016) ve ELECTRE - ELimination Et Choix Traduisant la REalité'dir (Benayoun vd., 1966; Rashid vd., 2018). Adomavicius vd. (2011) çalışmalarında, ÇÖKV mekanizmalarını kullanarak tavsiye üreten ÇÖTS sistemlerini incelemiş ve bu çalışmaları kategorik olarak bir araya getirmiştir. Çalışmalarını, Roy (1996)'un çalışmasında belirttiği adımlar ışığında gerçekleştirmiştir.

Tavsiye sistemlerinin, gerçek hayat uygulamalarında, çoklu-ölçütlü verilerle çalışan örnekleri her geçen gün daha fazla karşımıza çıkmaktadır. Özellikle e-ticaret sitelerinde örneklerine sıklıkla rastladığımız çoklu-ölçütlü veri tabanları, araştırmacıların ilgi odağı olmuştur. Araştırmacıların sıklıkla çalıştığı veri tabanlarından biri, TripAdvisor, HRS.com gibi otel değerlendirme sitelerinin verileridir (Fuchs ve Zanker, 2012; Jannach vd., 2012; Nilashi vd., 2015). Bir diğer çoklu-ölçütlü veri tabanı ise Yahoo Movies! veri kümesidir (Jannach vd., 2012; Nilashi ve Ithin, 2014; Kermany ve Alizadeh, 2017). Çalışmalarda kullanılan bir diğer çoklu-ölçütlü veri kümesi ise Zagat's Guide adı verilen restoran değerlendirme veri kümesidir (Teng ve Lee, 2007). Bunların yanı sıra, farklı senaryolarda kullanılmak üzere kendi çoklu-ölçütlü veri kümesini oluşturan araştırmacılar da mevcuttur (Hu, 2014).

Hafıza ve model tabanlı geleneksel OF sistemlerinin genişletilmesi ile oluşturulan benzerlik tabanlı ÇÖOF sistemleri üzerine yapılan çalışmalar şunlardır: Manouselis ve Costopoulou (2007), hafıza tabanlı OF sistemleri yardımıyla çevrimiçi kullanıcılara e-market tavsiyesi veren bir ÇÖOF sistemi önermiştir. Tang ve McCalla (2009) da benzer bir yaklaşımla, makale tavsiyesi üreten bir ÇÖOF sistemi önermiştir. Bilge ve Kaleli (2014), ürün odaklı hafıza tabanlı OF sistemlerinin farklı uzaklık ölçütleri kullanılarak davranışlarının karşılaştırıldığı bir ÇÖOF sistemi çalışması sunmuştur. Sahoo vd. (2006) çalışmalarında, Si ve Jin (2003) tarafından önerilen FMM (Flexible Mixture Model) modelini genişleterek çoklu-ölçütlü model tabanlı bir yaklaşım sunmuştur. Li vd. (2008), OF sistemlerinde kullanılan Tekil Değer Ayrışımı'nı çoklu lineer yapıda kullanarak (Multi-linear Singular Value Decomposition) model tabanlı bir ÇÖOF sistemi önermiştir. Liu vd. (2011) de kullanıcıların kümelenmesi ile oluşturulan modellerden istifade eden bir OF yaklaşımı ile ÇÖOF sistemi sunmuştur. Hu (2014) te genel tercih kriteri için değişken olmayan indeksleri kullanan kullanıcı tabanlı ÇÖOF sistemi geliştirmiştir.

ÇÖOF sistemlerinin toplama fonksiyonu temelli örneklerine literatürde sıkça rastlamaktayız. Bu yöntemlerin önemli bir çoğunluğu regresyon analizi ve makine öğrenimi yöntemleri etrafında toplanmıştır. Toplama fonksiyonlu yaklaşım ilk olarak Adomavicius ve Kwon (2007) tarafından önerilmiştir. Liu vd. (2011), Least Square Regresyon yöntemi ile oluşturulmuş toplama fonksiyonu tabanlı ÇÖOF sistemi önermiştir. Jannach vd. (2012) ise, çalışmalarında hem kullanıcı hem de ürün temelli Support Vector Regresyonu'nu ağırlıklı olarak bir araya getirdiği WeightedSVM metodunu önermiştir. Bilge ve Kaleli (2014), benzerlik-tabanlı ÇÖOF sisteminin yanı sıra lineer regresyon tabanlı toplama fonksiyonlu ÇÖOF sistemi üzerine de çalışmıştır. Nilashi vd. (2015), önceki toplama fonksiyonu tabanlı sistemlerden daha doğru sonuçlar alabilmek adına, kullanıcı segmentlerinin kümeleme yöntemi ile otomatik olarak belirlendiği bir geliştirilmiş regresyon tabanlı ÇÖOF sistemi önermiştir.

Bahsi geçen tüm bu yaklaşımların yanı sıra, geleneksel OF ve ÇÖOF sistemlerinin problemlerinin üstesinden gelebilmek için önerilen melez yaklaşımlar da literatürde geniş bir yer almaktadır. Shambour ve Lu (2011) ürün tabanlı OF sistemlerinin problemlerini yenebilmek adına ürün odaklı hafıza tabanlı ÇÖOF sistemi ile anlamsal filtrelemeyi bir araya getirdiği Multi-Criteria Semantic-enhanced CF (MC-SeCF) yaklaşımını önermiştir. Zarrinkalam ve Kahani (2012) makale alıntılama üzerine yaptığı çalışmasında, ÇÖOF ile

içerik-tabanlı filtreleme sistemlerini bir araya getirmiştir. Nilashi vd. (2014a), ÇÖOF sistemlerinin problemleri olan Seyreklik ve Ölçeklenebilirlik problemlerinin üstesinden gelebilmek için, HOSVD (Higher-Order Singular Value Decomposition: Yüksek Mertebeden Tekil Değer Ayrıştırma) ve ANFIS (Adaptive Neuro-Fuzy Inference System: Sinirsel-Bulanık Çıkarım Sistemi) yöntemlerini biraraya getirmiştir. HOSVD ölçeklenebilirlik problemi için boyut indirgemede kullanılırken, ANFIS seyreklik problem için bulanık kurallar çıkarımı için kullanılmıştır. Yine Nilashi vd. (2014b), ÇÖOF sistemlerinin tahmin doğruluğunu artırmak adına, ANFIS ve SOM (Self-Organizing Map: Kendini Düzenleyen Harita) Kümelemesi yöntemlerini biraraya getirerek melez bir ÇÖOF yaklaşımında bulunmuştur. Kermany ve Alizadeh (2017), kriterler arasındaki ilişkiyi ortaya çıkarmak için adaptif bir sinirsel-bulanık çıkarım sistemi önermiştir. Çalışmalarında benzerlik hesabı için, bulanık Cosine ve Jaccard ölçütlerini bir araya getirmiş ve sistemin güvenilirlik ve gürbüzlük gibi problemlerinin üstesinden gelebilmek için de ürün ve kullanıcı tabanlı benzerlikleri konveks bir kombinasyonda bir araya getirmiştir.

5. ÇOKLU-ÖLÇÜTLÜ ORTAK FİLTRELEME SİSTEMLERİNDE KABA KÜMELER KULLANIMI

Daha önceki bölümlerde bahsedildiği gibi ÇÖOF sistemleri, benzerlik tabanlı ve toplama fonksiyonu tabanlı olmak üzere iki ana kategoriye ayrılır. Benzerlik tabanlı ÇÖOF sistemleri k-boyutlu bir geleneksel OF sistemi olarak düşünülür ve her kriter için tavsiye üretildikten sonra komşuluk seçimi için gerekli yaklaşımlar uygulanır ve bu sayede genel tercih kriterine tavsiye üretilir. Toplama fonksiyonu tabanlı ÇÖOF sistemlerinde ise kriterler ve genel tercihlerin yer aldığı kriter arasındaki ilişkiler irdelenerek, istatistiki, makine öğrenimi gibi yöntemler aracılığıyla kriter ağırlıklandırması yapılır. Kriterlerin ağırlıklandırılması aşamasında devreye KKT girer.

5.1. Kaba Kümeler Teorisi

Pawlak (1982) tarafından geliştirilen KKT eksik, tutarsız ve gereksiz veri içeren veri kümelerinde belirsizliklerin ve kuşkuların üstesinden gelen bir matematiksel metottur (Bonikowski ve Bryniarski, 1998). KKT tutarsız veri içindeki gizli örüntüleri ortaya çıkarmaya olanak sağlar. Ayrıca KKT öznitelik seçme, öznitelik çıkarma veya elde etme, karar kuralları elde etme ve örüntü elde etme gibi birçok işlem için de kullanılır. KKT'nin bu başarısı, onun birçok alanda kullanılmasına neden olmuştur. KKT'nin ilk aşamasını oluşturan bilgi sistemi, satırları nesneleri (örnekleri, numuneleri) ve sütunları öznitelikleri gösteren geniş ve kapsamlı bilgi içeren bir matristir. Bir bilgi sistemi, $S = (U, A)$ çifti ile tanımlanır. Bilgi sisteminde bahsi geçen U, boş olmayan sonlu nesne kümesi ve A, sonlu öznitelik kümesi ise $a: U \rightarrow V_a$ her $a \in A$ için bir V_a öznitelik değer kümesi mevcuttur. Burada, A öznitelik kümesine, koşul nitelikler kümesi de denilir. Eğer bir bilgi sisteminde bir karar niteliği de varsa bu bilgi sistemine karar sistemi veya tablosu denir.

Bir karar sistemi, var olan problem veya oluşturulmuş olan model hakkında geniş ve kapsamlı bir şekilde bilgileri barındırmaktadır. Buna bağlı olarak, bu bilgi sistemi gereksiz yere çok büyümüş, genişlemiş olabilir. Herhangi bir $B \subseteq A$ olduğunda, B özniteliklerinin her alt kümesi ayırt edilemezlik bağıntısı adı verilen, U evreninin bir denklik bağıntısını tanımlar. (x, y) ikilisi, U'dan durum, nitelik çiftleri olmak üzere, denklem 5.1'de tanımlanan bir $IND_S(B)$ denklik bağıntısına B-ayırt edilemezlik bağıntısı denir.

$$IND_s(B) = \{(x,y) \in U^2 \mid \forall a \in B\ a(x) = a(y)\} \tag{5.1}$$

KKT'de temel yaklaşım şudur: Eğer X nesne kümesi, öznitelik kümesinin özelliklerini kullanarak kesin bir biçimde tanımlanamıyorsa alt ve üst yaklaşımlar ile gösterilebilir.

- Alt yaklaşım kümesi $\underline{R}(X)$ ile gösterilir ve kümeye kesin olarak ait olan nesneler kümesidir. Yani her X_i nesnesinin X'e kesin olarak ait olduğunu ifade eder. Alt yaklaşım kümesi denklemi, denklem 5.2'de gösterilmektedir.

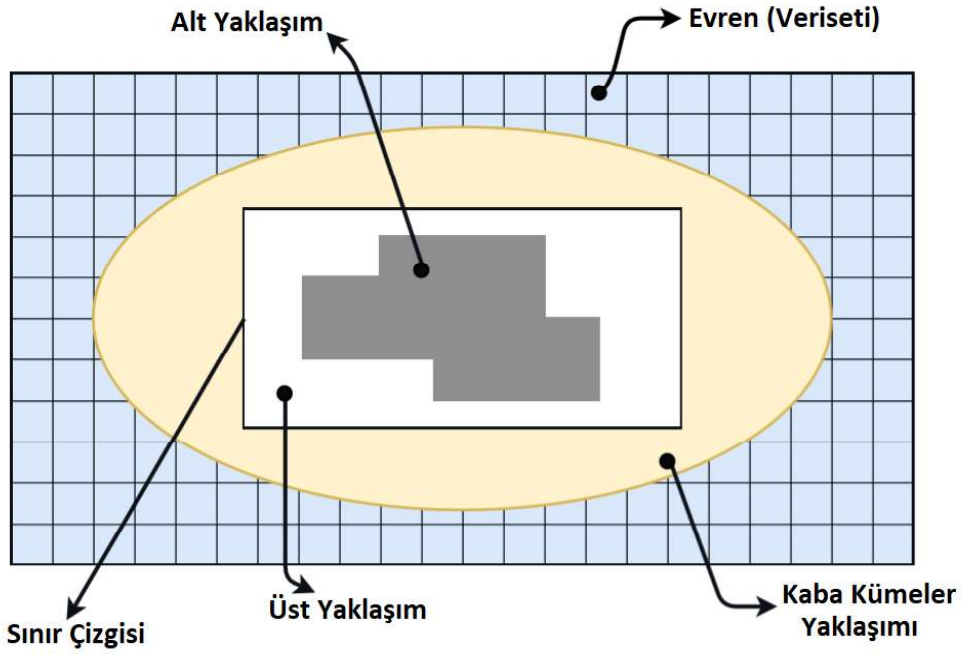
$$\underline{R}(X) = \{X_i \in U \mid [X_i]_{INDs(R)} \subset X\} \tag{5.2}$$

- Üst yaklaşım kümesi ise kümeye ait olma olasılığı olan nesneler kümesidir. Yani her X_i nesnesinin X'e ait olabileceğini ifade eder. Üst yaklaşım kümesi denklemi, denklem 5.3'te gösterilmektedir.

$$\bar{R}(X) = \{X_i \in U \mid [X_i]_{INDs(R)} \cap X \neq 0\} \tag{5.3}$$

- Sınır bölgesi ise, ne kesin olarak bir kümeye ait olan ne de olmayan nesnelerin kümesini ifade eder. Sınır bölgesi denklemi, denklem 5.4'te gösterilmektedir.

$$RN(X) = \bar{R}(X) - \underline{R}(X) \tag{5.4}$$



Görsel 5.1. *KKT konsepti ve yaklaşımları*

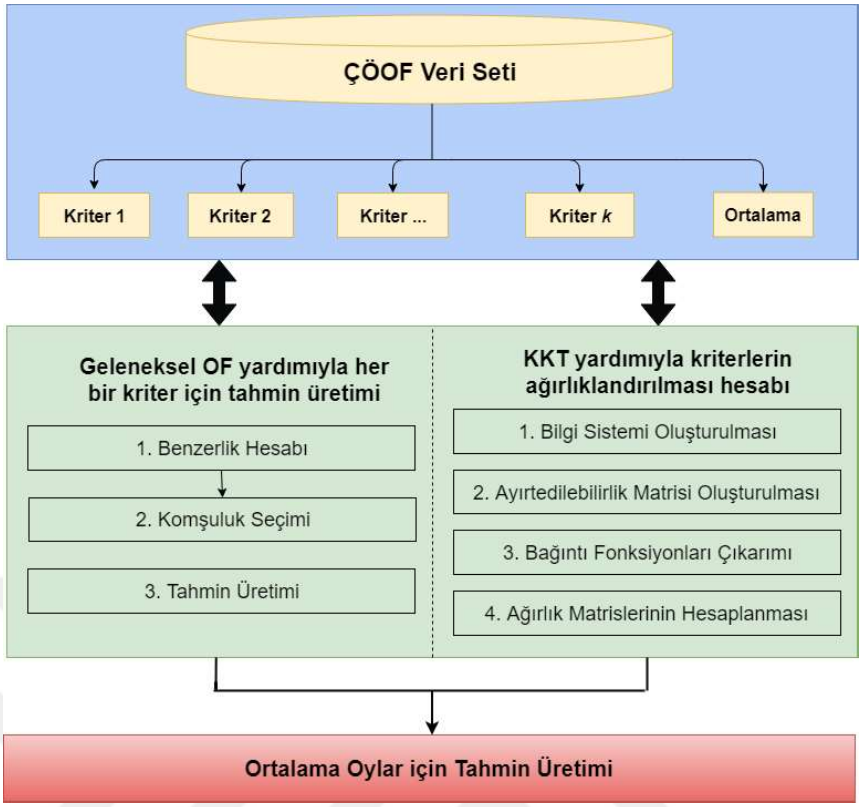
Kullanıcıların; bir ürünün kriterlerine verdiği oyların, birbirleriyle ilişkili olup olmadığını ve o ürünün genel tercih kriteri üzerinde nasıl bir etkisinin olduğunu öğrenebilmek için KKT kullanılabilir. KKT'nin ÇÖOF sistemlerine nasıl dâhil edilebileceği konusu bir sonraki bölümde ayrıntılı olarak ifade edilmiştir.

KKT'de iki objenin birbirinden ayırt edilebilmesi neredeyse imkânsız ise bu iki obje arasındaki ilişki karakteristik ilişki ile tanımlanır. İki objenin, özellikler kümesinde yer alan elemanlarının birbirlerinden ayırt edilebilmesi mümkün ise bu iki obje arasındaki ilişki ayırt edilebilirlik veya ayırt edilemezlik ilişkisi ile tanımlanır. Bu tez çalışmasında, bilgi sistemlerinde karşılaştırılan hücreler dolu olduğundan, tablolarda ayırt edilebilirlik ilişkisi irdelenmiştir.

5.2. KKT ile Toplama Fonksiyonu Tabanlı ÇÖOF Sistemi Yaklaşımı

Sistemin genel çalışma prensibinin özeti şu şekildedir: k adet kriter için tahmin üretiminde herhangi bir tavsiyelendirme algoritması (geleneksel OF, içerik-tabanlı filtreleme veya melez yaklaşımlar) kullanılabilir. Genel tercih kriteri için tahmin üretimi işleminde ise, k adet kriter için üretilen tahminler ile, KKT yardımıyla elde edilen kriter ağırlıkları bir araya getirilerek sonuca varılır.

Kriterlerin ağırlıklandırılması işlemi adımları ise şu şekildedir; kullanıcıların ürünlere verdikleri oylar ayrıştırılarak her kullanıcının kendi bilgi sistemi oluşturulur, bu bilgi sistemlerinden her bir ürün veya kullanıcı için ayırt edilemezlik matrisleri hesaplanır ve bu matrislerden elde edilen ilişki fonksiyonları irdelenerek kriterlerin yine ürün veya kullanıcı bazında ağırlıkları bulunur. Bu ağırlıklar her bir kriter için OF sistemi yardımıyla bulunan tahmin değerleri ile çarpılarak genel tercih kriteri için bir tahmin üretilir. Sonraki alt bölümlerde ürün odaklı ağırlıklandırma için sistemin uygulama aşamaları ayrıntılı olarak anlatılmıştır.



Görsel 5.2. *KKT kullanan ÇÖOF sistemin uygulama adımlarına genel bir bakış*

5.2.1. Geleneksel OF yardımıyla her bir kriter için tahmin üretimi

Önerilen sistemin ilk aşaması veri kümesinde yer alan her bir kriter için sistem tarafından tahmin üretilmesi aşamasıdır. Bu aşamada geleneksel OF sisteminden yararlanılmaktadır. Benzerlik hesabı aşamasında sistem, hedef varlığın diğer varlıklarla olan benzerliğini sayısal bir rakama döker. Hedef varlığın diğer varlıklarla olan benzerliğini en yüksekten en düşüğe doğru sıralayan sistem bir sonraki aşamada hesaba katılmak üzere seçilecek k adet komşu varlığı belirler. Son olarak sistem, belirlemiş olduğu komşuların verileri yardımıyla hedef varlığa tahmin üretiminde bulunur. Aşağıdaki bölümlerde, önerilmiş olan sistemin daha önce de bahsi geçen üç adımı uygulaması işlemi ayrıntılı olarak anlatılmıştır.

5.2.1.1. Benzerlik hesabı

OF sistemlerinin tahmin üretiminde gerçekleştirdiği ilk aşama hedef varlık için benzerlik hesaplanması aşamasıdır. Bu aşamada birtakım istatistiki yöntem ve metriğe başvurulur. Literatürde, benzerlik hesaplanması için önerilmiş birçok metrik bulunmaktadır. Bunlardan en yaygın olarak kullanılanları Cosine, Manhattan, Öklit, Jaccard, Chebyshev ve PCC'dir. Bu metrikler yardımıyla sistem iki varlık arasındaki

benzerliği 0 ile 1 arasında veya -1 ile 1 arasında bir sayı ile ifade edebilir. Örneğin, 0 ile 1 arasında benzerlik değerleri üreten bir algoritmada iki varlık arasında 0,40 benzerlik değeri hesaplanıyorsa, bu iki varlığın birbirlerine %40'lık bir oranla benzediği sonucu ortaya çıkar. -1 ile 1 arasında benzerlik değerleri bulan bir algoritmaya göre -1 negatif yönde %100 uyumluluk, 1 pozitif yönde %100 uyumluluk, 0 ise hiçbir bağıntının olmadığını ifade eder.

Bu tez çalışmasında, yukarıda bahsedilen benzerlik metriklerinden Öklit, Manhattan, Chebyshev ve PCC metriklerinin geleneksel OF ve ÇÖOF sistemlerine etkisi yapılan deneylerle irdelenmiştir. Dolayısıyla bu metriklerin formüllerini belirtmekte fayda vardır.

Öklit mesafesi'ne göre iki kullanıcı arasındaki mesafe, onları birbirine bağlayan segmentlerin uzunluğuna eşittir ve denklem 5.5'te verilen formül ile hesaplanır (Guo, 2014).

$$d_{A,B} = \sqrt{\sum_{i=1}^{k} \left| r_{A,i} - r_{B,i} \right|^2}, \tag{5.5}$$

$d_{A,B}$, A ve B kullanıcıları arasındaki mesafeyi, k kullanıcıların oy matrisinde yer alan oy sayısını, $r_{A,i}$ ve $r_{B,i}$ sırasıyla belirli ürün i için A ve B kullanıcılarının verdiği oyu ifade eder.

Manhattan mesafesi bir şehirde tek yön olup olmadığına bakmaksızın kare şeklindeki bloklar arasında yol alarak gidilebilecek mesafeyi verir ve denklem 5.6'da verilen formül ile hesaplanır (Guo, 2014).

$$d_{A,B} = \sum_{i=1}^{k} \left| r_{A,i} - r_{B,i} \right| \tag{5.6}$$

Chebyshev veya bir diğer deyişle maksimum değer mesafesi formülü denklem 5.7'de verilmiştir (Adomavicius ve Kwon, 2007).

$$d_{A,B} = \max_{i=1,2,\ldots,k} \left| r_{A,i} - r_{B,i} \right| \tag{5.7}$$

Yapılan deneysel çalışmaların önemli bir bölümünde ise korelasyon tabanlı benzerlik hesabı yapılmaktadır ve bu hesap için en popüler olarak kullanılan PCC metriği kullanılmaktadır. PCC metriği formülü, denklem 3.4'te verilmiştir.

5.2.1.2. Komşuluk seçimi

Sistem, hedef varlıkların diğer varlıklarla olan benzerliklerini hesapladıktan sonra bu değerleri en yüksekten en düşüğe doğru sıralar ve bu değerleri matrislere kaydeder. Önceki bölümde de belirtildiği üzere komşuluk seçimi yaklaşımı olan *k*-NN veya eşik değer yaklaşımlarından biri seçilerek hedef varlık için komşular belirlenmiş olur.

Bu tez çalışmasında gerçekleştirilen deneysel çalışmalarda komşuluk seçimi için *k*-NN algoritmasına başvurulmuştur. Komşuluk sayısının seçiminde çevrimiçi performans ile tahmin doğruluğu arasındaki dengenin hesaba katılması gerekir. Az sayıda komşu seçmek sistemin hızlı çalışmasını sağlarken başarılı olmasını zorlaştırır. Çok sayıda komşu seçmek ise sistemin çevrimiçi performansını oldukça düşürebilir. Bu dengenin doğru ayarlanması adına komşuluk sayısının kaç alınması gerektiği ile ilgili bir deney yapılmış ve komşuluk seçiminde seçilecek k değeri bu deneyden çıkan sonuçlara göre belirlenmiştir.

5.2.1.3. Tahmin üretimi

OF sistemlerinin işlem aşamalarının sonuncusu olan tahmin üretimi aşamasında, sistem, hedef varlığa en benzer varlıkların oylarının ağırlıklı kombinasyonu yardımıyla bir tahmin üretimi gerçekleştirir.

Bu tez çalışmasında sistemin tahmin üretimi için Resnick vd. (1994)'nin çalışmalarında önermiş oldukları formül kullanılmıştır. Tahmin üretimi için kullanılan formül denklem 5.8'de verilmiştir.

$$p_{aq} = \bar{r}_a + \sigma_a \times \frac{\sum_{u=1}^{n} \left(\frac{r_{uq} - \bar{r}_u}{\sigma_u} \right) \times w_{au}}{\sum_{u=1}^{n} w_{au}}, \tag{5.8}$$

p_{aq} ilgili ürün q için hedef kullanıcı a'ya sistem tarafından üretilen tahmin değerini, n ise komşu sayısını ifade eder.

5.2.2. KKT kullanılarak kriterlerin ağırlıklandırılması

Önerilen sistemin ikinci aşaması olan KKT kullanılarak kriterlerin ağırlıklandırılması aşaması, birinci aşama olan OF sistemi yardımıyla tahmin üretiminden

bağımsızdır ve işleme başlamak için birinci aşamadan gelecek herhangi bir girdiye ihtiyacı yoktur. Bu aşama, çok kanallı programlama ile birinci aşama ile senkronize bir şekilde gerçekleştirilebilmektedir.

Bu aşamada, genel tercihler için tahmin üretilebilmesi adına, her bir ürün için kriterlerin önem ağırlıklarını ortaya çıkaran bir dizi adım uygulanır. Bu aşamalar sırasıyla şunlardır:

- Bilgi Sistemi Oluşturulması
- Ayırt edilebilirlik Matrisi Oluşturulması
- Bağıntı Fonksiyonları Çıkarımı
- Ağırlık Matrislerinin Hesaplanması

5.2.2.1 Bilgi sistemi oluşturulması

Önerilen metodun ilk aşaması her ürün için bilgi sistemi oluşturulması aşamasıdır. Bu aşamada $S_i(u, c)$ ile ifade edilen matris bilgi sistemi olarak adlandırılır ve i ürünü, u kullanıcıyı ve c ise kriteri temsil eder. Her ürün için oluşturulan bilgi sistemi matrisi, satırlarında kullanıcıları sütunlarında kriterleri barındıran ve hücrelerinde kullanıcının kriterlere vermiş olduğu oyları barındıran bir matristir. Matrisi daha kolay anlamak adına Tablo 5.1'de örnek bir veri kümesi ve Tablo 5.2'de ilk kullanıcı için bu veri kümesinden oluşturulan bilgi sistemi yer almaktadır. Örnek veri kümesi yedi ürün ve altı kullanıcıdan oluşan bir *kullanıcı × ürün* matrisidir ve her bir hücrede, ilgili ürünün dört ayrı kriterine verilmiş oylar yer almaktadır. -,-,-,- ile gösterilen hücreler ilgili kullanıcının ilgili ürün hakkında fikir sahibi olmadığını veya henüz oy vermediğini ifade etmektedir. Bir sonraki tabloda yer alan bilgi sistemi birinci kullanıcı olan *Ahmet* için oluşturulan bilgi sistemidir. Sistem bu aşamayı tüm ürünler için gerçekleştirdikten sonra birinci aşama tamamlanmış olur.

Tablo 5.1. *Örnek bir kullanıcı × ürün × kriter matrisi*

Kullanıcı	Ürün 1	Ürün 2	Ürün 3	Ürün 4	Ürün 5	Ürün 6	Ürün 7
Ahmet	5,5,4,5	4,4,2,2	4,4,4,4	3,3,2,2	--,--,--,--	3,3,3,3	5,5,4,5
Burçin	3,3,3,3	4,1,2,4	3,2,4,5	3,3,3,3	2,5,3,1	2,2,3,4	1,2,2,2
Ceyda	--,--,--,--	--,--,--,--	3,2,5,1	2,3,4,5	2,3,5,2	2,4,5,2	--,--,--,--
Defne	3,3,1,1	2,3,4,1	--,--,--,--	--,--,--,--	2,3,4,5	2,1,2,3	2,3,4,5
Emin	3,3,3,3	4,4,4,4	1,3,5,2	--,--,--,--	1,2,1,3	--,--,--,--	3,3,3,3
Ferhat	--,--,--,--	2,3,4,2	--,--,--,--	3,3,2,1	--,--,--,--	1,1,1,1	--,--,--,--

Tablo 5.2'de, birinci ürün için oluşturulmuş bilgi sistemi görülmektedir. Bu ürün için *Ahmet*, *Burçin*, *Defne* ve *Emin* adlı kullanıcılar oy vermiş olduğundan yalnızca dört satır bulunur. İlgili ürün için henüz oy vermemiş kullanıcıların verileri bilgi sisteminden elenmiş olur. Matrisin sütunlarında ilgili ürünün dört kriteri ve hücrelerde ilgili kullanıcıların kriterlere vermiş olduğu oylar yer almaktadır.

Tablo 5.2. *Birinci ürün için oluşturulmuş bilgi sistemi*

Kullanıcı	Kriter 1	Kriter 2	Kriter 3	Kriter 4
Ahmet	5	5	4	5
Burçin	3	3	3	3
Defne	3	3	1	1
Emin	3	3	3	3

5.2.2.2. Ayırt edilebilirlik matrisi oluşturulması

Bu bölümde bilgi sisteminden ayırt edilebilirlik matrisi oluşturulması işlemi ayrıntılı olarak anlatılmıştır. *U,* evren olarak ta adlandırılan nesnelerin boş olmayan sonlu kümesi ve *R* ise *U*'nun eşitlik bağıntısı olsun. $R \subset U \times U$ ayırt edilebilirlik bağıntısı olarak ifade edilir. Aynı eşitlik sınıfına dâhil edilen elemanlar ayırt edilemez, dâhil edilmeyen elemanlar ise ayırt edilebilir olarak ifade edilir. Bu sistemde nesnelerin sonlu kümesinde aynı eşitlik sınıfına dâhil olmayan elemanların ayırt edici olmasından ötürü sistem ayırt edilebilir matrisine odaklanmıştır.

Her ürün için oy veren kullanıcı sayısı n olsun; sistem her ürün için bir $n \times n$ matris oluşturur ve matrisin satır ve sütunlarına kullanıcıları sırasıyla doldurur. Matrisin alt üçgeni kullanıcıların verdiği oylara göre karşılaştırılarak doldurulur. Matrisin köşegeni ve üst bölümü ise boş bırakılır. Matrisin alt bölümü için; karşılaştırılan iki kullanıcı bir kriter için aynı oyu vermişse göz ardı edilir, farklı oylar verilmişse o kriter hücreye yazılır. İlgili hücreler doldurulduktan sonra ayırt edilebilirlik matrisi oluşmuş olur.

İşlemin daha anlaşılır olması için Tablo *5.2*'de yer alan birinci ürün için oluşturulmuş bilgi sisteminin ayırt edilebilirlik matrisi Tablo 5.3'te verilmiştir. Buna göre; birinci ürün için 4 kullanıcı oy vermiştir: *Ahmet*, *Burçin*, *Defne* ve *Emin*. Bu kullanıcılar matrisin satırlarına ve sütunlarına yerleştirilerek 4×4'lük bir matris oluşturulmuştur. *Ahmet* ve *Burçin* adlı kullanıcılar karşılaştırılırken bu kullanıcıların tüm kriterlere farklı oylar verdiği görüldüğünden *Ahmet* ve *Burçin*'in kesiştiği hücreye dört kriterin de adı yazılmıştır. Bir başka örnekle anlatmak gerekirse *Burçin* ve *Defne* adlı

kullanıcıların oyları incelendiğinde; bu iki kullanıcının ilgili ürünün ilk iki kriterine aynı oyu, diğer iki kritere farklı oylar verdiğini görebiliyoruz. Dolayısıyla bu iki kullanıcının kesiştiği hücrede yalnızca C_3 ve C_4 kriterleri yer almaktadır.

Tablo 5.3. *Birinci ürün için oluşturulan bilgi sistemine göre ayırt edilebilirlik matrisi*

	Ahmet	**Burçin**	**Defne**	**Emin**
Ahmet				
Burçin	C_1,C_2,C_3,C_4			
Defne	C_1,C_2,C_3,C_4	C_3,C_4		
Emin	C_1,C_2,C_3,C_4		C_3,C_4	

5.2.2.3. Bağıntı fonksiyonları çıkarımı

Bir önceki bölümde kullanıcıların birbirleriyle karşılaştırılması ile ortaya çıkarılan ayırt edici matrisi, o ürünün ayırt edici fonksiyonunun bulunmasında bize yardımcı olur. Ayırt edici fonksiyon f_A, matristeki her hücrenin; kendi içerisinde yer alan verilerin ayrılma işlemi ($\vee$) ile, diğer hücrelerle ise birleşme işlemi ($\wedge$) ile bir araya getirilmesi ile oluşur. Ayrılma işlemi ve birleşme işlemi matematikte bileşik önermeler olarak ifade edilir. Ayrılma işlemi, iki veya daha çok önermenin "veya" bağlacı ile birbirine bağlanması işlemi iken; birleşme işlemi ise iki veya daha çok önermenin "ve" bağlacı ile birbirine bağlanması işlemidir. Ayrılma işlemi ile bir araya getirilen p ve q önermelerinden oluşan $p \vee q$ bileşik önermesinde, önermelerden en az biri doğru iken önerme doğru (1), ikisi de yanlışken önerme yanlıştır (0). Birleşme işlemi ile bir araya getirilen p ve q önermelerinden oluşan $p \wedge q$ bileşik önermesinde ise, önermelerden ikisi de doğru iken doğru (1), diğer durumlarda önerme yanlıştır (0). Bu bilgiler ışığında, bir önceki bölümde yer alan örnekte ortaya çıkarılan ayırt edici matrisinden üretilen ayırt edici fonksiyon aşağıda yer almaktadır:

$$f_{ürün1} = (C_1 \vee C_2 \vee C_3 \vee C_4) \wedge (C_1 \vee C_2 \vee C_3 \vee C_4) \wedge (C_3 \vee C_4) \wedge (C_1 \vee C_2 \vee C_3 \vee C_4) \wedge (C_3 \vee C_4)$$

Ortaya çıkan denklem, Boole cebirinin sadeleştirme işlemleri kullanılarak sadeleştirilir. Bu işlemler şu şekildedir (Telçeken ve Doğan, 2004):

$$a \vee (a \wedge b) \equiv a$$

$$a \wedge (a \vee b) \equiv a$$

$$a \vee (b \wedge c) \equiv (a \vee b) \wedge (a \vee c)$$

$$a \wedge (b \vee c) \equiv (a \wedge b) \vee (a \wedge c)$$

Sadeleştirme işlemi sonucunda ayırt edici fonksiyon $f_{\text{ürün}_1}$ nihai halini alır:

$f_{ürün1} = (C_1 \vee C_2 \vee C_3 \vee C_4) \wedge (C_1 \vee C_2 \vee C_3 \vee C_4) \wedge (C_3 \vee C_4) \wedge (C_1 \vee C_2 \vee C_3 \vee C_4) \wedge (C_3 \vee C_4)$

Ayırt edici fonksiyonda yer alan $(C_1 \vee C_2 \vee C_3 \vee C_4)$ önermesi *a*, $(C_3 \vee C_4)$ önermesi ise *b* ile ifade edilecek olunursa, fonksiyon şu şekilde ifade edilebilir:

$f_{\text{ürün}_1} = a \wedge a \wedge b \wedge a \wedge b.$

Boole cebiri teoremlerinden; aynı kuvvet (özdeşlik) kuralına göre $a \wedge a \equiv a$'dır. Dolayısıyla önermenin yeni hali şu şekilde olur:

$f_{\text{ürün}_1} = a \wedge b = (C_1 \vee C_2 \vee C_3 \vee C_4) \wedge (C_3 \vee C_4).$

Ayırt edici fonksiyonun yeni durumunda yer alan $(C_3 \vee C_4)$ önermesi *a,* $(C_1 \vee C_2)$ önermesi ise *b* ile ifade edilecek olunursa, fonksiyon şu şekilde ifade edilebilir:

$f_{\text{ürün}_1} = a \wedge (a \vee b) \equiv a = (C_3 \vee C_4).$

KKT uygulanması ile ortaya çıkan bu fonksiyondan çıkarılacak sonuç şudur: Birinci ürün için üçüncü ve dördüncü kriterler eşit öneme sahiptir, birinci ve ikinci kriterlerin ise dikkate alınmaması gerekir. Kriterlerin önem dereceleri her zaman burada verilen örnekteki gibi kolaylıkla fark edilmez. Gerçek veri kümelerinde ortaya çıkan senaryolar çok daha karmaşıktır. Bu aşamada ortaya çıkarılan ayırt edici fonksiyonlar sayesinde, kriterlerin önem derecesinin ortaya çıkarıldığı ve akabinde ağırlık matrislerinin bulunduğu yoğunluk hesaplama işlemi için gerekli zemin hazırlanmış olur.

5.2.2.4. Ağırlık matrislerinin hesaplanması

Bağıntı fonksiyonları çıkarımı yöntemi ile ortaya çıkarılan ayırt edici fonksiyonlar ağırlık matrisi adı verilen bir matriste toplanır. Ayırt edici fonksiyonların, kriterlerin ağırlıkları hakkında bilgi verebilmesi için basit bir dönüşüm işlemine tabii tutulması gerekir. *f(i)*, *i* ürününe ait ayırt edilebilirlik fonksiyonu ve *Sayı*(f_i, c_j) c_j kriterinin f_i ayırt edici fonksiyonu içerisindeki frekansı olsun. Her bir kriterin, fonksiyon içerisindeki yoğunlukları denklem 5.9 ve denklem 5.10'da verilen formüller yardımıyla bulunur (Çekik ve Telçeken, 2016).

$$N = Sayı(f_i, \vee) + 1 \tag{5.9}$$

$$d_{i,j} = \frac{Sayı(f_i, c_j)}{N} \tag{5.10}$$

Formülde yer alan, *Sayı*(f_i, ∨) değeri bize fonksiyonda yer alan ayrılma operatörü sayısını vermektedir. Her bir kriterin fonksiyonda geçme sayısının ayrılma operatörü sayısı *N*'e bölünmesi sonucunda kriterlerin yoğunlukları bulunur. Her bir ürün için ortaya çıkarılan kriterler yoğunlukları 1'e oranlanarak kriter ağırlıkları hesaplanmış olur. Bir örnekle ifade etmek gerekirse, önceki bölümlerde 1.ürün için ortaya çıkarılan $f_{ürün1}$ ayırt edici fonksiyonu üzerinde bahsi geçen formül uygulandığında;

$f_{ürün1} = (C_3 \vee C_4)$

$$N = Sayı(f_i, \vee) + 1 = 2$$

$$d_{\text{ürün1, } c_1} = \frac{Sayı(f_{\text{ürün1}}, c_1)}{N} = \frac{0}{2} = 0, \qquad d_{\text{ürün1, } c_2} = \frac{Sayı(f_{\text{ürün1}}, c_2)}{N} = \frac{0}{2} = 0,$$

$$d_{\text{ürün1, } c_3} = \frac{Sayı(f_{\text{ürün1}}, c_3)}{N} = \frac{1}{2} = 0.5, \qquad d_{\text{ürün1, } c_4} = \frac{Sayı(f_{\text{ürün1}}, c_4)}{N} = \frac{1}{2} = 0.5$$

olarak hesaplanmış olur. Sistem, bu formül sayesinde tüm veri kümesi için ağırlık matrisini ortaya çıkarmış olur. Sonuçta elde edilen ağırlık matrisi, genel tercih kriteri için tahmin üretiminde kullanılır. KKT kullanılarak kriter ağırlıklandırılması işlem adımları başarılı bir şekilde uygulandıktan sonra Tablo 5.4'te gösterilen örnek ağırlık matrisi bulunmuş olur.

Tablo 5.4. *Veri kümesi için hesaplanan örnek bir ağırlık matrisi*

Ürünler	Kriter 1	Kriter 2	Kriter 3	Kriter 4
Ürün 1	0	0	0,5	0,5
Ürün 2	0,25	0,25	0,25	0,25
…	…	…	…	…
Ürün *n*	0,40	0	0,20	0,40

5.2.3. Genel tercih oyları için tahmin üretimi

Bu aşamada kriterler için OF sistemi aracılığıyla üretilmiş tahminler ile KKT yardımıyla kriterler için bulunan ağırlıklar bir araya getirilerek, her bir ürün için genel tercih kriteri oy tahmini üretimi işlemi gerçekleşir. Kullanıcı bir ürünün kriterleri için tahmin istediğinde, sistem OF sistemi aracılığıyla tahmin üreterek bir sonuca varır. Kullanıcı genel tercih için bir tahmin istediğinde ise KKT aracılığıyla ortaya çıkarılan ağırlıklar hesaba katılarak özgün bir tahmin yürütme algoritması çalıştırılır. Bu aşamayı

daha anlaşılır bir şekilde ifade etmek istersek, yine önceki bölümlerde aşamalı olarak ilerlediğimiz örneği kullanabiliriz.

Örneğin; oy aralığı 1-5 arası olan bir veri kümesinde *Ahmet* adlı kullanıcı birinci ürünün kriterlerine {5, 5, 4, 5} oylarını vermiştir. Kullanıcı, sistemden genel tercih için bir oy tahmini yapmasını ister. Sistem öncelikle *Ahmet* için OF sistemi aracılığıyla dört ayrı kriter için tahmin hesabı yapar. Sistemin *Ahmet* için ürettiği tahminler {(4,10), (3,72), (3,98), (4,47)} olsun. KKT yardımıyla birinci ürün için bulunan ağırlıklar yukarıdaki bölümlerde de anlatıldığı üzere {(0), (0), (0,5), (0,5)}'tir; yani birinci ve ikinci kriter önemsizken, üçüncü ve dördüncü kriter eşit öneme sahiptir. Bu durumda sistemin *Ahmet* için üreteceği genel oy tahmini şu şekilde olur:

$$P_{ürün1,\ Ahmet} = (3{,}98 \times 0{,}5) + (4{,}47 \times 0{,}5) = 4{,}23.$$

Bir başka örnekte, bir ürün hakkında hiçbir bilgisi olmayan bir kullanıcıyı ele alalım. *Defne* isimli bir kullanıcı '*Esaretin Bedeli*' adlı film için tavsiye istemektedir. Bu kullanıcının o film hakkında daha önce bir bilgisi olmasa da her bir kriter için sistem OF sistemi yardımıyla tahmin yürütür. Tahminler {(3,17), (4,21), (4,01), (3,44)} olsun. *Esaretin Bedeli* adlı film için sistemin belirlediği kriter ağırlıkları {(0,25), (0,25), (0), (0,50)} ise, bu ürünün ortalaması *Defne* için şu şekilde olacaktır:

$$P_{EsaretinBedeli,\ Defne} = (3{,}17 \times 0{,}25) + (4{,}21 \times 0{,}25) + (3{,}44 \times 0{,}5) = 3{,}57.$$

6. DENEYLER

Bu bölümde, OF sistemleri, ÇÖOF sistemleri ve önerilen KKT kullanan ÇÖOF sistem uygulamalarının başarımları bir dizi deneysel çalışma ile irdelenmiştir. Önerilen sistemlerin uygulanmasında geleneksel MovieLens veri kümesi ve çoklu-ölçütlü Yahoo!Movies veri kümesi kullanılmıştır. Sistemlerin başarısı istatistiki yöntemlerle analiz edilmiş ve önerilen sistem, literatürde kabul görmüş benzer çalışmalarla kıyaslanmıştır.

6.1.Kullanılan Veri Kümesi

Deneylerimizde veri kümeleri olarak, tek ölçütlü film oylama veri kümesi MovieLens ml-100k[2] ve çoklu-ölçütlü film oylama veri kümesi olan Yahoo!Movies[3] YM5-5, YM10-10 ve YM20-20 veri kümeleri kullanılmıştır. ML-100K veri kümesinin karakteristik özellikleri aşağıda belirtilmiştir:

- Veri kümesinde 943 adet kullanıcı ve 1682 adet ürün vardır. 1.586.126 adet olası kullanıcı-ürün hücresinden yalnızca 100.000 adet hücre oylanmıştır. Dolayısıyla, veri kümesinin doluluk oranı %6,30'dur.
- Veri kümesinin oy yelpazesi 1 – 5 aralığındadır. 1 en düşük oyu temsil ederken 5 en yüksek oyu temsil etmektedir. Her kullanıcı en az 20 ürüne oy vermiştir.

Yahoo!Movies veri kümesinin karakteristik özellikleri ise şu şekildedir:

- Veri kümesi, kullanıcıların filmlere verdiği oy sayıları ve filmlerin kullanıcılar tarafından oylanma sayılarına göre 3 ayrı kümeye bölünmüştür. Bu veri kümeleri YM5-5, YM10-10 ve YM20-20 adıyla belirtilmiştir. Veri kümelerinde yer alan 5-5, 10-10 ve 20-20 ifadeleri her bir kullanıcının vermesi gereken en az oy sayısını ve her filmin sahip olması gereken en az oy sayısını ifade etmektedir. Örneğin YM5-5 veri kümesinde, her kullanıcı en az 5 ürüne oy vermiş ve her ürün en az 5 kullanıcı tarafından oylanmıştır.
- Veri kümelerinin 1-5 ve 1-13 oy yelpazesi şeklinde iki farklı çeşidi mevcuttur. 1-13 oy yelpazesi, birçok ülkenin eğitim sisteminde notlandırma için kullanılan, A+, A, A-, B+, B, B-, C+, C, C-, D+, D, D- ve F derecelerinin sayısallaştırılmış halidir[4]. A+ en yüksek sayı olan 13'ü temsil ederken F en düşük sayı olan 1'i temsil eder. 1-5 oy yelpazesi ise 1-13 oy yelpazesinin gruplanarak 1-5 yelpazesine

[2] https://grouplens.org/datasets/movielens/
[3] https://webscope.sandbox.yahoo.com/catalog.php?datatype=r
[4] https://en.wikipedia.org/wiki/Academic_grading_in_the_United_States

dönüştürülmesi ile oluşturulan yelpazedir. Bu yelpazede, A ile ifade edilen oylar 5, B ile ifade edilen oylar 4, C ile ifade edilen oylar 3, D ile ifade edilen oylar 2 ve son olarak F değeri ile ifade edilen oylar 1 olarak değerlendirilir.

YM5-5 veri kümesi en çok kullanıcıya ve ürüne sahip veri kümesi iken YM20-20 veri kümesi en az kullanıcı ve ürüne sahip veri kümesidir. Veri kümelerinde yer alan kullanıcı ve ürün sayıları Tablo 6.1'de belirtilmiştir.

Tablo 6.1. *Veri kümelerinde yer alan kullanıcı, ürün ve oy sayıları*

Veri Kümesi	Kullanıcı Sayısı	Ürün Sayısı	Oy Sayısı (Kriter Başına)
ml-100k	943	1682	100.000
YM5-5	5978	3079	82.599
YM10-10	1827	1471	48.026
YM20-20	429	491	18.504

6.2.Değerlendirme Metrikleri

Sistemin başarısını ölçmek için literatürde sıklıkla kullanılan istatistiki metrikler kullanılmıştır. Bu metrikler; Ortalama Mutlak Hata (OMH), Kare Ortalamalarının Karekökü Hatası (KOKH), Kapsam (Coverage) ve *N*-Kesinlik (Precision-at-*N*) metrikleridir. Bu metrikler yardımıyla elde edilen sonuçlar diğer çalışmalarla karşılaştırılmış ve sistemin performansı hakkında bilgi edinilmiştir.

OMH hesabı şu şekilde yapılmaktadır: Kullanıcıların daha önce oy verdikleri ürünler boş oylarla değiştirilerek bu ürünlere sistem tarafından öneriler oluşturulur ve mutlak değer içerisinde, her oy için, tahmin edilen değerden gerçek oylar çıkartılarak toplam oy sayısına bölünür. OMH hesabı formülü denklem 6.1'de verilmiştir.

$$OMH = \left(\sum_{i=1}^{N} |p_i - r_i|\right) \Big/ N, \tag{6.1}$$

Formülde yer alan N değeri toplam oy sayısını, p_i ilgili ürün için üretilen oy değerini ve r_i ise o ürün için gerçekte verilen oy değerini temsil eder.

KOKH hesabında ise OMH'den farklı olarak önerilen oylar ile gerçek oylar arasındaki farkın karesi toplam oy sayısına bölünerek çıkan sonucun karekökü alınır. KOKH hesabı formülü denklem 6.2'de verilmiştir.

$$KOKH = \sqrt{\sum_{i=1}^{N} |p_i - r_i|^2 \Big/ N} \tag{6.2}$$

OMH ve KOKH değerleri ne kadar düşük ise sistemin başarısı o kadar iyi demektir. OMH'nin yanı sıra KOKH'nin de irdelenmesi, sistemin büyük hata oranlarına yaklaşımının nasıl olduğunu anlayabilmek içindir.

Kapsam değeri, veri kümesinde yer alan oyların yüzde kaçına oy verildiğini göstermek için kullanılmıştır. Kapsam değerinin hesaba katıldığı deneylerde, tüm veri kümesine tahmin üretimi gerçekleştirildiğinden; bu deneylerde kapsam değeri 1,0'dır. Kapsam formülü kısaca denklem 6.3'te verilen formül ile ifade edilir:

$$Kapsam = \frac{Tahmin\ üretimi\ yapılan\ ürün\ sayısı}{Toplam\ oy\ sayısı} \tag{6.3}$$

N-Kesinlik hesabı şu şekilde yapılır: Öncelikle; içerisinde oy değeri bulunmayan hücreler göz ardı edilir. Ardından, değerlendirmeye alınacak oylar için sistem tarafından tahmin üretilir ve gerçek oylar ile bu oylar için üretilen tahminler birer matriste tutulur. Son olarak, en yüksek değere sahip olan *N* adet gerçek oy ve *N* adet tahmin edilen oylar değerlendirmeye alınarak; kesişen ürünlerin toplam sayıya oranı yüzde olarak *N*-Kesinlik değerini gösterir. *N*-Kesinlik formülü ise denklem 6.4'te verilmiştir.

$$N - Kesinlik = \frac{(N\ \text{Adet Önerilen Ürünlerden İlgili Olanların Sayısı})}{(N\ \text{Adet Önerilen Ürün Sayısı})} \tag{6.4}$$

Sistemin başarımının test edilmesinde değerlendirme metriklerinin kullanımı kadar veri önişleme aşamaları da önemlidir. Veri önişleme aşamasında; veri kümesinin eğitim ve test kümesine ayırılması, veri dönüşümü, veri elenmesi gibi işlemlere başvurulur. Veri dönüşümü işlemi veri kümesinin sayısallaştırılması ve Bölüm 6.1'de de ifade edilen 1-13 yelpazesinin 1-5 yelpazesine dönüştürülmesi gibi işlemleri kapsamaktadır. Veri elenmesi işlemleri, algoritmanın çalışmasında hiçbir katkısı olmayan varlıkların, örneğin tüm ürünlere aynı oyu veren kullanıcılar gibi, elenmesi gibi işlemleri kapsamaktadır. Veri kümesinin eğitim ve test kümesine ayrılması konusunda ise literatürde sıklıkla kullanılan *K*-Katlamalı Çapraz Doğrulama (*K*-Fold Cross Validation) deney metodolojisi kullanılmıştır. *K*-Katlamalı Çapraz Doğrulama metodolojisi, veri kümesini *K* adet parçaya bölerek, her defasında bir parçanın test, geri kalan parçaların eğitim kümesi

olarak seçilmesini ve bu şekilde tüm veri kümesinin irdelenmesini amaçlayan bir yöntemdir.

6.3.Deneysel Bulgular

Bu tez çalışması kapsamında uygulanan ilk deney, farklı mesafe metriklerinin geleneksel OF sistemlerine etkisinin nasıl olduğunun irdelenmesi üzerinedir. OF sisteminin uygulanması sırasında benzerlik hesabı için kullanılabilecek pek çok metrik mevcuttur ve her bir metrik sistemin performansını doğrudan doğruya etkilemektedir. Bu deney ile hangi metriğin tavsiye oluşturmada daha başarılı olduğunun görülmesi amaçlanmıştır. Öklit, Manhattan, Chebyshev ve PCC metrikleri YM5-5, YM10-10 ve YM20-20 veri kümelerinin 1-5 ve 1-13 oy yelpazeli versiyonları üzerinde test edilerek sistemin OMH ve KOKH değerleri çıkarılmıştır. Eldeki veri kümelerinin bütünü kullanılmış ve tüm oy verilmiş ürünler için tavsiye üretilerek sonuçlar değerlendirilmiştir. Komşuluk seçimi için *k* değeri 20 olarak tanımlanmıştır. Buna göre, bahsi geçen metriklerin 1-5 yelpazeli veri kümeleri için OMH ve KOKH değerleri Tablo 6.2 ve 1-13 yelpazeli veri kümeleri için OMH ve KOKH değerleri ise Tablo 6.3'te verilmiştir.

Tablo 6.2. *Yahoo!Movies veri kümeleri üzerinde farklı mesafe metrikleri kullanımının sistem performansına etkisi. Komşuluk = 20, Oy Yelpazesi: 1-5.*

	Öklit		**Manhattan**		**Chebyshev**		**PCC**	
Veri Kümesi	**OMH**	**KOKH**	**OMH**	**KOKH**	**OMH**	**KOKH**	**OMH**	**KOKH**
YM20.Hikâye	0,5021	0,7621	0,5889	0,8545	0,5758	0,8395	**0,4634**	**0,6851**
YM20.Oyunculuk	0,4790	0,7316	0,5506	0,8092	0,5397	0,7963	**0,4377**	**0,6562**
YM20.Görsellik	0,4620	0,7153	0,5327	0,7905	0,5226	0,7786	**0,4288**	**0,6441**
YM20.Yönetmenlik	0,4963	0,7591	0,5801	0,8514	0,5666	0,8359	**0,4672**	**0,6898**
YM10.Hikâye	0,5760	0,8450	0,6600	0,9106	0,6599	0,9105	**0,4783**	**0,7189**
YM10.Oyunculuk	0,5394	0,8004	0,6154	0,8599	0,6154	0,8599	**0,4454**	**0,6787**
YM10.Görsellik	0,5324	0,7916	0,6094	0,8507	0,6091	0,8506	**0,4468**	**0,6800**
YM10.Yönetmenlik	0,5768	0,8471	0,6574	0,9086	0,6569	0,9083	**0,4781**	**0,7179**
YM5.Hikâye	0,6107	0,8834	0,6964	0,9495	0,6965	0,9496	**0,4775**	**0,7231**
YM5.Oyunculuk	0,5703	0,8333	0,6485	0,8952	0,6485	0,8952	**0,4460**	**0,6847**
YM5.Görsellik	0,5651	0,8293	0,6444	0,8892	0,6444	0,8892	**0,4460**	**0,6860**
YM5.Yönetmenlik	0,6069	0,8762	0,6929	0,9449	0,6929	0,9449	**0,4774**	**0,7245**

Tablo 6.3. *Yahoo!Movies veri kümeleri üzerinde farklı mesafe metrikleri kullanımının sistem performansına etkisi. Komşuluk = 20, Oy Yelpazesi: 1-13.*

	Öklit		**Manhattan**		**Chebyshev**		**PCC**	
Veri Kümesi	**OMH**	**KOKH**	**OMH**	**KOKH**	**OMH**	**KOKH**	**OMH**	**KOKH**

YM20.Hikâye	**1,3862**	2,0654	1,6279	2,3799	1,4971	2,2263	1,4025	**1,9628**
YM20.Oyunculuk	1,3811	2,0547	1,5779	2,3181	1,4417	2,1524	**1,3485**	**1,8904**
YM20.Görsellik	1,3487	2,0255	1,5398	2,2598	1,3718	2,0614	**1,3243**	**1,8604**
YM20.Yönetmenlik	**1,4191**	2,1158	1,6435	2,4078	1,4963	2,2273	1,4205	**1,9906**
YM10.Hikâye	1,7513	2,4832	2,0798	2,7798	2,0792	2,7797	**1,4909**	**2,1298**
YM10.Oyunculuk	1,6662	2,3907	1,9553	2,6493	1,9541	2,6484	**1,4129**	**2,0321**
YM10.Görsellik	1,6560	2,3705	1,9379	2,6220	1,9371	2,6219	**1,4192**	**2,0352**
YM10.Yönetmenlik	1,7618	2,4901	2,0738	2,7751	2,0730	2,7745	**1,4982**	**2,1348**
YM5.Hikâye	1,8394	2,5893	2,1050	2,8230	2,1050	2,8230	**1,4303**	**2,0890**
YM5.Oyunculuk	1,7411	2,4808	1,9833	2,6931	1,9833	2,6931	**1,3543**	**1,9988**
YM5.Görsellik	1,7395	2,4704	1,9801	2,6801	1,9801	2,6801	**1,3638**	**2,0046**
YM5.Yönetmenlik	1,8417	2,5926	2,1007	2,8185	2,1007	2,8185	**1,4343**	**2,0920**

Tablo 6.2 ve Tablo 6.3 verileri incelendiğinde mesafe metriklerinin OF sistemlerinin performansını doğrudan etkilediği anlaşılmaktadır. OMH ve KOKH değerleri hücrelerde sırasıyla yazılmıştır. Bu değerler ne kadar küçük ise sistem o kadar az hata ile tahmin üretmiş demektir. Her bir veri kümesinde bulunan en iyi OMH ve KOKH değerleri kalın punto ile gösterilmiştir. Buna göre, en başarılı sonucun PCC tarafından alındığı görülmekte olup, bir sonraki başarılı metriğin Öklit mesafesi metriği olduğu anlaşılmaktadır. Manhattan ve Chebyshev mesafe metriklerinin sistem başarısında PCC ve Öklit mesafesi kadar etkili olmadığı görülmektedir. Bu tez çalışmasında yapılan diğer deneylerde, benzerlik hesabında PCC metriğinin kullanılmasının başlıca sebebi bu deneylerde ortaya çıkan sonuçlardır.

Yapılan bir diğer deney ise, tahmin üretiminde komşuluk seçiminin OF sistemine etkisinin analiz edilmesi üzerinedir. Bu deneyde benzerlik hesabında PCC metriği kullanılmış ve komşu seçiminde artışa gidildiğinde sistemin nasıl etkilendiği gözlemlenmiştir. Buna göre, bahsi geçen komşuluklar ile, 1-5 yelpazeli veri kümeleri için OMH ve KOKH değerleri Tablo 6.4 ve 1-13 yelpazeli veri kümeleri için OMH ve KOKH değerleri ise Tablo 6.5'te verilmiştir.

Tablo 6.4. *Yahoo!Movies veri kümeleri üzerinde komşuluk sayısının sistem performansına etkisi. Benzerlik metriği = PCC, Oy Yelpazesi: 1-5.*

	K = 10		**K = 20**		**K = 40**		**K = 80**	
Veri Kümesi	**OMH**	**KOKH**	**OMH**	**KOKH**	**OMH**	**KOKH**	**OMH**	**KOKH**
YM20.Hikâye	0,4839	0,7217	0,4634	0,6851	0,4484	0,6562	**0,4396**	**0,6400**
YM20.Oyunculuk	0,4568	0,6862	0,4377	0,6562	0,4188	0,6227	**0,4079**	**0,6029**
YM20.Görsellik	0,4503	0,6799	0,4288	0,6441	0,4128	0,6159	**0,4023**	**0,5972**
YM20.Yönetmenlik	0,4844	0,7189	0,4672	0,6898	0,4492	0,6584	**0,4380**	**0,6391**

YM10.Hikâye	0,5029	0,7459	0,4783	0,7189	0,4540	0,6857	**0,4288**	**0,6494**
YM10.Oyunculuk	0,4721	0,7080	0,4454	0,6787	0,4211	0,6478	**0,3986**	**0,6152**
YM10.Görsellik	0,4708	0,7096	0,4468	0,6800	0,4208	0,6453	**0,3969**	**0,6098**
YM10.Yönetmenlik	0,5045	0,7502	0,4781	0,7179	0,4526	0,6841	**0,4295**	**0,6505**
YM5.Hikâye	0,5081	0,7529	0,4775	0,7231	0,4478	0,6907	**0,4204**	**0,6569**
YM5.Oyunculuk	0,4752	0,7130	0,4460	0,6847	0,4160	0,6523	**0,3892**	**0,6179**
YM5.Görsellik	0,4752	0,7171	0,4460	0,6860	0,4166	0,6521	**0,3884**	**0,6154**
YM5.Yönetmenlik	0,5072	0,7557	0,4774	0,7245	0,4488	0,6917	**0,4201**	**0,6558**

Tablo 6.5. *Yahoo!Movies veri kümeleri üzerinde komşuluk sayısının sistem performansına etkisi. Benzerlik metriği = PCC, Oy Yelpazesi: 1-13.*

	K = 10		**K = 20**		**K = 40**		**K = 80**	
Veri Kümesi	**OMH**	**KOKH**	**OMH**	**KOKH**	**OMH**	**KOKH**	**OMH**	**KOKH**
YM20.Hikâye	1,4618	2,0597	1,4025	1,9628	1,3697	1,9136	**1,3554**	**1,8930**
YM20.Oyunculuk	1,4030	1,9886	1,3485	1,8904	1,3103	1,8367	**1,3013**	**1,8209**
YM20.Görsellik	1,3854	1,9641	1,3243	1,8604	1,2872	1,8102	**1,2784**	**1,7960**
YM20.Yönetmenlik	1,4879	2,1006	1,4205	1,9906	1,3767	1,9233	**1,3670**	**1,9036**
YM10.Hikâye	1,5664	2,2277	1,4909	2,1298	1,4251	2,0400	**1,3596**	**1,9428**
YM10.Oyunculuk	1,4901	2,1308	1,4129	2,0321	1,3480	1,9455	**1,2902**	**1,8602**
YM10.Görsellik	1,4936	2,1307	1,4192	2,0352	1,3484	1,9371	**1,2859**	**1,8462**
YM10.Yönetmenlik	1,5689	2,2222	1,4982	2,1348	1,4241	2,0304	**1,3606**	**1,9400**
YM5.Hikâye	1,5093	2,1770	1,4303	2,0890	1,3556	1,9997	**1,2860**	**1,9074**
YM5.Oyunculuk	1,4308	2,0806	1,3543	1,9988	1,2829	1,9108	**1,2158**	**1,8200**
YM5.Görsellik	1,4408	2,0895	1,3638	2,0046	1,2932	1,9130	**1,2254**	**1,8215**
YM5.Yönetmenlik	1,5091	2,1769	1,4343	2,0920	1,3610	2,0054	**1,2922**	**1,9122**

Bu deneyden yorumlanması gereken bilgi şudur: Komşuluk seçiminin yükselmesi sistemin daha başarılı tahmin üretmesini sağlamaktadır, ancak özellikle YM20-20 veri kümesinin verileri de dikkate alındığında; komşuluk seçimi belirli bir sayıya kadar sistemin başarısını önemli ölçüde arttırsa da, bir noktadan sonra önceki değerlere kıyasla daha az arttırdığı anlaşılmaktadır. Komşuluk sayısının artması, sistemin daha uzun sürede tahmin üretmesi anlamına geldiğinden komşuluk sayısının akıllıca seçilmesi, sistemin başarısı ile çevrimiçi performansı arasındaki dengenin sağlanması için önemlidir. Dolayısıyla bundan sonraki deneylerde komşuluk sayısı belirli bir seviyede tutulacaktır.

Bir önceki deneyde yapılan çalışmaya benzer bir şekilde, tek ölçütlü veri kümesi ml-100k üzerinde, mesafe ölçütleri ve komşuluk seçiminin sistem üzerindeki etkisi OMH ve KOKH verileri aracılığıyla gözlemlenmiştir. Sonuçlar, Tablo 6.6'da yer almaktadır. Deney sonuçlarına göre en iyi sonuç PCC metriği ile alınmaktadır. PCC'yi Öklit, Öklit'ı

Chebyshev ve son olarak Chebyshev'i de Manhattan metriği takip etmektedir. Komşuluk seçiminde kullanılan K değerinin sistemi logaritmik olarak etkilediği, başlangıçta hızlı ardından yavaşlayan bir performans artışı sergilediği söylenebilir. Bu durumda komşuluk seçimini önceki deneylerde de ifade edildiği üzere 20 civarında seçmek hem sistemin çevrimiçi performansını arttıracak hem de başarıdan taviz verilmeden sistemin optimum şekilde çalışması sağlanmış olacaktır.

Tablo 6.6. *ML-100K veri kümesinde farklı mesafe metrikleri kullanımının ve komşuluk sayısının sistem performansına etkisi.*

Komşuluk Sayısı	Öklit		Manhattan		Chebyshev		PCC	
	OMH	KOKH	OMH	KOKH	OMH	KOKH	OMH	KOKH
10	0,7807	0,9893	0,8099	1,0189	0,8000	1,0086	**0,6714**	**0,8838**
20	0,7577	0,9681	0,7957	1,0077	0,7772	0,9881	**0,6498**	**0,8514**
30	0,7419	0,9532	0,7851	0,9997	0,7601	0,9717	**0,6429**	**0,8399**
40	0,7289	0,9406	0,7772	0,9933	0,7456	0,9574	**0,6404**	**0,8351**
50	0,7181	0,9299	0,7699	0,9873	0,7348	0,9459	**0,6394**	**0,8326**
100	0,6889	0,8987	0,7484	0,9709	0,6930	0,8974	**0,6397**	**0,8290**

Bu tez çalışmasında hazırlanan deneylerden bir diğeri ise Adomavicius ve Kwon (2007)'nin çalışmalarında önermiş oldukları benzerlik-tabanlı ÇÖOF sisteminin performansının analiz edilmesi üzerinedir. Yahoo!Movies veri kümeleri YM10-10 ve YM20-20 ile yapılan çalışmada, veri kümeleri tek-çıkışlı çapraz doğrulama yöntemi (leave-one-out cross validation) ile her seferinde 1 hedef kullanıcı, kalanlar eğitim verisi olacak şekilde ayrılarak veri kümesinin tümü irdelenmiş, böylece her oy için tavsiye üretildikten sonra üretilen tahmin değerleri gerçek oylarla karşılaştırılarak sistemin performansı değerlendirilmiştir. Genel tercih kriteri haricindeki kriterler, geleneksel OF sistemi ile PCC yardımıyla benzerlikler hesaplanmış, bu benzerliklerden, En Kötü Senaryo (Worst-Case) ve Ortalama Senaryo (Average-Case) yaklaşımları ile genel tercih kriterlerinin tahmini için benzerlikler seçilmiştir. Tahmin üretiminde kullanılan komşuluk sayısı içinse 10, 20, 40 ve 80 olarak belirlenerek sistemin performansı ölçülmüştür. YM20-20 veri kümesi ile yapılan çalışmalardan elde edilen OMH ve KOKH verileri sırasıyla Tablo 6.7 ve Tablo 6.8'de verilmiştir.

Tablo 6.7. *YM20-20 veri kümesinde genel tercih kriteri için benzerlik tabanlı ÇÖOF sisteminin farklı komşuluk değerleri ile OMH verileri*

Oy Yelpazesi	Yaklaşım	k = 10	k = 20	k = 40	k = 80
1-5 Aralığı	En Kötü	0,7099	0,715	0,7066	0,6683
	Ortalama	**0,6991**	**0,699**	**0,6826**	**0,6628**
1-13 Aralığı	En Kötü	**1,977**	**1,9765**	1,9045	1,827
	Ortalama	1,9711	1,9824	**1,8843**	**1,8239**

Tablo 6.8. *YM20-20 veri kümesinde genel tercih kriteri için benzerlik tabanlı ÇÖOF sisteminin farklı komşuluk değerleri ile KOKH verileri*

Oy Yelpazesi	Yaklaşım	k = 10	k = 20	k = 40	k = 80
1-5 Aralığı	En Kötü	**0,9505**	0,9568	0,9438	0,8935
	Ortalama	0,9569	**0,9469**	**0,9241**	**0,8925**
1-13 Aralığı	En Kötü	2,6838	**2,6772**	2,5736	2,4792
	Ortalama	**2,6796**	2,6794	**2,5497**	**2,4788**

Tablo 6.7 ve Tablo 6.8'de açıkça görüldüğü üzere, geleneksel OF sisteminde de olduğu gibi, benzerlik tabanlı ÇÖOF sisteminde de komşuluk sayısı arttıkça sistemin tahmin doğruluğu artmaktadır, ancak komşuluk sayısı arttıkça sistemin tahmin üretme hızının düşeceği de unutulmamalıdır. Grafikler komşuluk seçiminde 10 ile 20 değerlerinin birbirine çok yakın sonuçlar verdiğini göstermektedir. Komşuluk sayısının 40, 80 gibi sayılara çıkarılması sistemin başarısını arttırsa da komşuluk seçiminde yüksek değerlerin kullanımı, sistemin çalışma hızını önemli ölçüde etkileyecektir. Dolayısıyla, çevrimiçi performans ve tahmin doğruluğu gibi parametreler bir arada düşünüldüğünde, komşuluk değeri seçiminin dikkatlice yapılması gerekmektedir. Aynı çalışmanın, YM10-10 veri kümesi ile kullanılmasından elde edilen OMH ve KOKH verileri ise sırasıyla Tablo 6.9 ve Tablo 6.10'da verilmiştir.

Tablo 6.9. *YM10-10 veri kümesinde genel tercih kriteri için benzerlik tabanlı ÇÖOF sisteminin farklı komşuluk değerleri ile OMH verileri*

Oy Yelpazesi	Yaklaşım	K= 10	K = 20	K = 40	K = 80
1-5 Aralığı	En Kötü	0,7510	0,7520	0,7460	0,7321
	Ortalama	**0,7400**	**0,7356**	**0,7308**	**0,7245**
1-13 Aralığı	En Kötü	**1,0072**	1,0088	0,9982	0,9814
	Ortalama	1,0093	**1,0066**	**0,9838**	**0,9725**

Tablo 6.10. *YM10-10 veri kümesinde genel tercih kriteri için benzerlik tabanlı ÇÖOF sisteminin farklı komşuluk değerleri ile KOKH verileri*

Oy Yelpazesi	Yaklaşım	K= 10	K = 20	K = 40	K = 80

1-5 Aralığı	En Kötü	2,1731	2,1711	**2,1503**	2,1457
	Ortalama	**2,1703**	**2,1680**	2,1548	**2,1421**
1-13 Aralığı	En Kötü	2,9219	2,9105	**2,8872**	**2,8678**
	Ortalama	**2,9204**	**2,9084**	2,8913	2,8697

Bir diğer çalışmada; bu tez kapsamında önerilmiş olan, KKT kullanılarak oluşturulan toplama fonksiyonu tabanlı ÇÖOF sisteminin (RoughMCCF), Yahoo!Movies veri kümesi üzerinde genel tercih kriteri için tahmin üretiminde ne derece başarılı olduğu irdelenmiştir. YM5-5, YM10-10 ve YM20-20 veri kümeleri kullanılarak, veri kümelerinde yer alan 4 ayrı kriter kümesinin tamamına (Acting: Oyunculuk, Story: Hikâye, Directory: Yönetmenlik, Visual: Görsellik) geleneksel OF sistemi yardımıyla tahmin üretilmiş, KKT yardımıyla filmlerin her biri için sabit ağırlıklar çıkarılmış ve kriterlere verilen oylar ağırlıkları ile çarpılarak genel tercih oyları hesaplanmıştır. Sistemin bu konudaki başarısı OMH ve KOKH ve *N*-Kesinlik değerleri ile irdelenmiştir. Yapılan deneyde OF tahmin üretimi aşamasında Manhattan, Chebyshev, Öklit ve PCC mesafe metrikleri kullanılmış ve komşuluk seçimi 20 olarak belirlenmiştir. Manhattan, Chebyshev ve Öklit mesafe metrikleri ile elde edilen OMH bulguları Tablo 6.11'de, KOKH bulguları Tablo 6.12'de ve PCC ile bulunan OMH ve KOKH verileri ise Tablo 6.13'te yer verilmiştir. Bu deneyden elde edilen bulgular, sistemin YM20-20 veri kümesinde en iyi performansı sergilediğini ortaya koymuştur.

Tablo 6.11. *Yahoo!Movies veri kümeleri üzerinde KKT kullanılarak oluşturulmuş ÇÖOF sisteminin OMH verileri. 1 – 5 ve 1 – 13 oy yelpazeleri, Metrikler = Manhattan, Chebyshev ve Öklit, Komşuluk = 20*

Veri Kümesi	Oy Aralığı	Manhattan	Chebyshev	Öklit
YM5-5.Overall	1-5 Aralığı	0,7818	0,7815	**0,7114**
	1-13 Aralığı	2,1437	2,1447	**1,8875**
YM10-10.Overall	1-5 Aralığı	0,7464	0,7460	**0,6789**
	1-13 Aralığı	2,0961	2,0967	**1,7824**
YM20-20.Overall	1-5 Aralığı	0,6776	0,6687	**0,6084**
	1-13 Aralığı	1,6533	1,4939	**1,4099**

Tablo 6.12. *Yahoo!Movies veri kümeleri üzerinde KKT kullanılarak oluşturulmuş ÇÖOF sisteminin KOKH verileri. 1 – 5 ve 1 – 13 oy yelpazeleri, Metrikler = Manhattan, Chebyshev ve Öklit, Komşuluk = 20*

Veri Kümesi	Oy Aralığı	Manhattan	Chebyshev	Öklit
YM5-5.Overall	1-5 Aralığı	1,0634	1,0628	**1,0077**
	1-13 Aralığı	2,9037	2,9050	**2,6645**

YM10-10.Overall	1-5 Aralığı	1,0268	1,0256	**0,9732**
	1-13 Aralığı	2,8376	2,8384	**2,5340**
YM20-20.Overall	1-5 Aralığı	0,9537	0,9465	**0,8824**
	1-13 Aralığı	2,3727	2,1886	**2,0658**

Tablo 6.13. *Yahoo!Movies veri kümeleri üzerinde KKT kullanılarak oluşturulmuş ÇÖOF sisteminin OMH ve KOKH verileri. 1 – 5 ve 1 – 13 oy yelpazeleri, Metrik = PCC, Komşuluk = 20*

Veri Kümesi	Oy Aralığı	OMH	KOKH
YM5-5.Overall	1-5 Aralığı	0,4705	0,7729
	1-13 Aralığı	1,3185	1,9027
YM10-10.Overall	1-5 Aralığı	0,4555	0,7639
	1-13 Aralığı	1,3228	1,8982
YM20-20.Overall	1-5 Aralığı	**0,4240**	**0,7372**
	1-13 Aralığı	**1.2891**	**1.8672**

Bu deneyde, 5-Kesinlik ve 7-Kesinlik değerlerinin çıkarımı için uygulanan deney yöntemi şu şekildedir: Veri kümeleri, 5-Katlamalı Çapraz doğrulama ile eğitim ve test verilerine ayrılmıştır. Veri kümesinin %80 oranında eğitim ve %20 oranında test kümesine ayrıldığı 5 varyasyon oluşturulmuştur. Her varyasyonda seçilen %20'lik test kümesindeki veriler, diğer varyasyonlarda yer alan verilerden farklıdır ve çakışan veri (aynı kullanıcının birden fazla varyasyonda yer alması durumu) yoktur. Bu şekilde tüm veri kümesinin değerlendirilmesi sağlanmıştır. Yinelenen her 5-Katlamalı çapraz doğrulama işleminde, seçilen eğitim verisi farklı olacağından ve buna bağlı olarak KKT ile oluşturulan kriter ağırlıkları değişeceğinden, işlem 100 kez tekrarlanarak sistemin genel davranışının analiz edilmesi sağlanmıştır. Buna göre, Yahoo!Movies veri kümelerinin tamamında uygulanan 5-Kesinlik ve 7-Kesinlik sonuçları Tablo 6.14'te yer almaktadır.

Tablo 6.14. *Yahoo!Movies veri kümeleri üzerinde KKT kullanılarak oluşturulmuş ÇÖOF sisteminin 5-Kesinlik ve 7-Kesinlik verileri. 1 – 5 ve 1 – 13 oy yelpazeleri*

Veri Kümesi	Oy Aralığı	5-Kesinlik	7-Kesinlik
YM5-5.Overall	1-5 Aralığı	**70,83**	**79,90**
	1-13 Aralığı	**77,94**	**85,77**
YM10-10.Overall	1-5 Aralığı	66,31	74,76
	1-13 Aralığı	73,87	72,54
YM20-20.Overall	1-5 Aralığı	62,05	64,77

	1-13 Aralığı	67,91	72,77

Bir diğer çalışmada Bilge ve Kaleli (2014)'nin çalışmalarında önermiş oldukları benzerlik-tabanlı ÇÖOF sistemi ve tarafımızca önerilen RoughMCCF yaklaşımı karşılaştırılmıştır. Bilge ve Kaleli (2014) çalışmalarında benzerlik tabanlı ÇÖOF sistemlerini farklı mesafe ve benzerlik metrikleri kullanarak irdelemiş, ürünler arasındaki ilişkileri tahmin etmek için PCC ve DKB (Düzeltilmiş Kosinüs Benzerliği – Adjusted Cosine Similarity), mesafe ölçümü tabanlı benzerlik yaklaşımı için ise Öklit, Chebyshev ve Manhattan mesafe metriklerini kullanarak sistemin tahmin doğruluğunu incelemiştir. Yaptıkları çalışmada kullandıkları tüm metriklerin sistemin doğruluğunu artırma konusunda önemli ölçüde faydalı olduklarını belirtmişlerdir. Tablo 6.15'te Bilge ve Kaleli (2014)'nin deney sonuçları ve kendi yaklaşımımız olan RoughMCCF yaklaşımının sonuçları alt alta verilmiştir. İki yöntem arasındaki tek fark ortalama kriter için tahmin üretimi konusundaki yaklaşımdır. Her iki yöntemde de kriterler için aynı benzerlik ve mesafe metrikleri kullanılmış, komşuluk seçimi olan *k* değeri 40 olarak belirlenmiş, veri kümesi eğitim ve test verilerine bölünürken 5-Katlamalı Çapraz Doğrulama yöntemi kullanılmıştır. Deneysel bulgular RoughMCCF yaklaşımının daha başarılı sonuç verdiğini ortaya koymaktadır.

Tablo 6.15. *Bilge ve Kaleli (2014)'nin çalışmalarında önerdiği ÇÖOF sistemi ile RoughMCCF yaklaşımının YM10 ve YM20 veri setlerinde OMH değerleri üzerinden karşılaştırılması*

Tavsiye Yaklaşımı				**YM10**	**YM20**
Çoklu-Ölçütlü Ürün Tabanlı OF	**Benzerlik**	**PCC**	**Ort.**	0.7670	0.7789
			Min.	0.7670	0.7445
		DKB	**Ort.**	0.7953	0.7911
			Min.	0.7475	0.7273
	Mesafe	**Öklit**		**0.7293**	**0.6322**
		Manhattan		0.7328	0.6798
		Chebyshev		0.7335	0..6493
RoughMCCF	**Benzerlik**	**PCC**		**0.5513**	**0.5398**
		DKB		0.5825	0.5610
	Mesafe	**Euclidean**		0.6439	0.5596
		Manhattan		0.7034	0.6172
		Chebyshev		0.7046	0.5995

Yapılan deneylerden bir diğeri ise KKT kullanılarak oluşturulan toplama fonksiyonu tabanlı ÇÖOF sistemimizin (RoughMCCF) literatürde kabul görmüş bazı çalışmalarla karşılaştırılması üzerinedir. Karşılaştırmanın doğru yapılabilmesi için diğer çalışmalarda önerilen yöntemlerle birebir aynı yöntem uygulanmıştır. Buna göre, 1-13 oy yelpazesine sahip YM10-10 veri kümesi 5-Katlamalı Çapraz doğrulama ile eğitim ve test verilerine ayrılmış ve veri kümesinin bütününe tahmin üretilerek sistemin doğruluğu 5-Kesinlik, 7-Kesinlik ve OMH metrikleri üzerinden değerlendirilmiştir. Önerilen yöntemin karşılaştırıldığı çalışmalar; Adomavicius ve Kwon (2007) tarafından önerilen geleneksel OF sistemi StandartCF ve benzerlik-tabanlı ÇÖOF sistemi MC-Similarity, Jannach vd. (2012) tarafından önerilen 200 komşu sayılı geleneksel *k*-NN yaklaşımlı UB-Pearson, ürün tabanlı regresyon yaklaşımı TotalReg, Kullanıcı ve Ürün tabanlı Destek Vektör Regresyonu yöntemlerinin birleştirildiği WeightedSVM, Lemire ve Maclachlan (2008) tarafından önerilen geleneksel OF sistemi SlopeOne ve Koren (2010) tarafından önerilen Matrix Faktorizasyonu kullanan FunkSVD yöntemidir. Yapılan deney ile elde edilen 5-Kesinlik, 7-Kesinlik ve OMH bulguları Tablo 6.16'da yer almaktadır.

Tablo 6.16. *YM10-10 veri kümesi üzerinde önerilmiş yöntemler ile RoughMCCF yönteminin N-Kesinlik ve OMH metrikleri üzerinden karşılaştırılması. Oy yelpazesi = 1-13*

Algoritma	5-Kesinlik	7-Kesinlik	OMH
StandartCF	68,70	69,00	2,06
MC-Similarity	67,69	71,12	2,06
TotalReg	70,90	70,40	-
UB-Pearson	71,68	70,38	2,07
SlopeOne	72,95	71,49	1,89
FunkSVD	73,15	71,65	1,89
RoughMCCF	73,87	72,54	1,30
WeightedSVM	**75,62**	**73,26**	**1,05**

Bu deneyden elde edilen sonuçlara göre, ÇÖOF sistemine KKT'nin dahil edilmesinin sistemin gelişmesinde önemli derecede katkı sağladığı anlaşılmıştır. Önerilen sistem, literatürde kabul görmüş ve ölçü alınabilecek çalışmalardan daha iyi sonuçlar almıştır. Önerdiğimiz sistemin, Jannach vd. (2012) tarafından önerilen WeightedSVM yönteminden daha az başarılı olmasının arkasında WeightedSVM yönetiminin sistem karmaşıklığı bulunmaktadır. Yapılan zaman karmaşıklığı analizinde tarafımızca önerilen sistemin ağırlık hesaplama işleminin zaman karmaşıklığının $O(N^2)$ olduğu

anlaşılmaktadır (*N*: Kullanılan eğitim verisi sayısını ifade eder). WeightedSVM yöntemi ise kullanıcı tabanlı ve ürün tabanlı iki ayrı Destek Vektör Regresyonu'nun kombinasyonu ile oluşturulmuştur. Araştırmacıların, SV-Regress-I ve SV-Regress-U olarak adlandırdıkları bu iki regresyon yönteminin her birinin zaman karmaşıklığı ise $O(N^2)$'dir. Bu sebeple WeightedSVM yaklaşımının önerdiğimiz yaklaşımdan iki kat daha fazla hesaplama zamanına ihtiyaç duyduğu söylenebilir. Bu açıdan bakıldığında önerdiğimiz yöntemin daha hızlı hesaplama yaptığı ve yine karşılaştırılan yöntemlere nazaran daha makul ölçülerde tahmin doğruluğu ürettiği söylenebilir.

Bir diğer deneyde ise 1-5 oy yelpazesine sahip YM5-5, YM10-10 ve YM20-20 veri kümeleri, benzer bir çalışma ile StandartCF, MC-Similarity, SlopeOne ve FunkSVD yöntemleri ile KOKH ve Kapsam metrikleri üzerinden karşılaştırılmış ve önerilen yöntemin (RoughMCCF) bu metrikler üzerinden de başarılı sonuçlar aldığı kanıtlanmıştır. Yahoo Movies! veri kümeleri üzerinde KOKH ve Kapsam metrikleri üzerinden yapılan karşılaştırma sonuçlarının yer aldığı veriler Tablo 6.17'de yer almaktadır.

Tablo 6.17. *Yahoo!Movies veri kümeleri üzerinde önerilmiş yöntemler ile RoughMCCF yönteminin KOKH ve Kapsam metrikleri üzerinden karşılaştırılması. Oy yelpazesi = 1-5*

Algoritma	**YM5-5**		**YM10-10**		**YM20-20**	
	KOKH	Kapsam	KOKH	Kapsam	KOKH	Kapsam
StandartCF	0,83	1,0	0,81	1,0	0,78	1,0
MC-Similarity	0,99	0,24	0,93	0,56	0,87	0,99
SlopeOne	0,97	1,0	0,89	1,0	0,81	1,0
FunkSVD	0,91	1,0	0,87	1,0	0,83	1,0
RoughMCCF	**0,80**	1,0	**0,76**	1,0	**0,74**	1,0

7. SONUÇ

Günümüzde İnternet'in yaygınlaşması ile birlikte gündelik olarak yaptığımız birçok işlemi yanı başımızda duran bilgisayar, mobil cihazlar vb. ile kolaylıkla yapabilmekteyiz. Eğitim kurumları e-öğrenme siteleri sayesinde uzaktan eğitim verebilmekte, bankalar internet bankacılığı hizmetleri ile neredeyse tüm bankacılık işlemleri için müşterilerine çevrimiçi hizmetler sunmakta, e-ticaret siteleri ile insanlar gıdadan tekstile, elektronikten mobilya ürünlerine kadar, aradıkları ürünleri mağaza dolaşmaya gerek kalmadan kolaylıkla bulabilmektedirler. Özellikle çevrimiçi alışveriş siteleri olmak üzere birçok alanda sunulan hizmet ve ürünlerin sayısı gün be gün çoğalmakta ancak bu durum bilgi bombardımanı denilen probleme yol açmakta yani müşterilerin aradıkları ürünleri bulmalarında problemler yaşamalarına neden olmaktadır. Bu problemin üstesinden gelebilmek için tavsiye sistemleri geliştirilmiştir.

Tavsiye sistemleri, sunulan binlerce ürün içerisinden müşterilerinin beğenebileceği ürünleri bulmaya çalışan sistemlerdir. Sistemin müşterilerini tanıyabilmesi için ürünler hakkında tercihlerini belirtmeleri gerekmektedir. Bu tercihler genellikle müşterilerin ürünler hakkındaki yorumları veya tercihlerini belirten sayısal değerler şeklindedir. Müşterilerin tercihlerini toplayan öneri sistemleri bu tercihleri kullanarak ürünleri değerlendirebilir ve müşterilerine beğenebilecekleri ürünleri tavsiye edebilir. Tavsiye sistemleri içerisinde en hızlı ve başarılı olan sistemler OF sistemleridir. OF sistemleri, kullanıcılarının ürünlere verdikleri sayısal tercihleri kullanarak kullanıcılara benzeyen kullanıcıları tespit ederek, hakkında fikir sahibi olmadıkları ürünler hakkında kendilerine benzeyen kullanıcıların ilgili ürüne olan tercihlerini kullanarak tavsiye üretir.

OF sistemlerinde kullanıcılar ürünler için tekil tercihlerde bulunurlar fakat bazı durumlarda ürünün geneline tekil tercihte bulunmaktansa ölçütleri için ayrı ayrı tercih belirtmek sistemin kullanıcılarını daha iyi tanımasını sağlayarak daha kaliteli tavsiyeler üretebilmesine olanak sağlar. Örneğin izlenilen bir film hakkında tercihte bulunurken filmin geneline tercih belirtmenin yanında filmin kurgusu, senaryosu, oyunculukları gibi farklı boyutları hakkında tercih belirtmek hem kullanıcıların karakteristiğini çözmede hem de filmi farklı boyutlardan değerlendirmede kolaylık sağlayacağı için ÇÖOF sistemleri geliştirilmiştir.

ÇÖOF sistemleri algoritmik yaklaşımlarına göre iki kategoriye ayrılmıştır: Benzerlik tabanlı yaklaşımlar ve Toplama fonksiyonu tabanlı yaklaşımlar. Benzerlik tabanlı yaklaşımlar OF sistemlerinin 3 boyutlu uzayı olarak düşünülebilir. Bu sistemlerde,

kriterlerin tahmin üretimi için geleneksel OF uygulanır. Genel tercihler için oy tahmini içinse, geleneksel OF sistemlerinden farklı olarak, hedef ürün veya kullanıcının benzerlikleri hesaplanırken iki farklı yaklaşım vardır: En Kötü Senaryo (Worst-Case) ve Ortalama Senaryo (Average-Case). En kötü senaryoda hedef varlığa komşuluk seçerken minimum değere sahip kriter hesaba katılırken ortalama senaryoda ise ortalama değere sahip kriter hesaba katılır. Toplama fonksiyonu tabanlı yaklaşımlarda ise kriterlerin önem dereceleri irdelenerek, her bir kriter için bir ağırlık hesabı yapılır. Genel tercihler için oy tahmini yapılırken bu ağırlıklar hesaba katılarak bir toplama fonksiyonu yardımı ile tahmin üretilir. Ağırlıkların ortaya çıkarılmasında regresyon analizi, sinir ağları, bulanık sistemler, makine öğrenimi gibi istatistiki yöntemlere başvurulur.

ÇÖOF sistemleri geleneksel OF sistemlerine oranla daha kişiselleştirilmiş tavsiyeler üretebildiğinden daha kaliteli tavsiye üretebilir. Ancak ölçütlerin sayısı arttıkça sistemin yükü ve çevrimiçi tepki süresi artacaktır. Bunların yanı sıra hâlihazırda geleneksel OF sistemlerinin ölçeklenebilirlik, gizlilik, doğruluk, performans, seyreklik, soğuk başlangıç gibi problemleri mevcut iken, ÇÖOF sistemlerinde bu problemler çok daha büyük boyutlardadır. Bu tez çalışmasında Kaba Kümeler Teorisi (KKT) yardımıyla tahmin üretimi gerçekleştiren bir toplama fonksiyonu tabanlı tavsiye sistemi ortamı oluşturulmuştur. Oluşturulan sistem ile, ÇÖOF sistemlerinin en önemli problemlerinden doğruluk, seyreklik ve performans problemlerinin önüne geçmek hedeflenmiştir. Deneysel bulgular, ÇÖOF sistemlerinde KKT kullanımının bu tür problemlerin ortadan kaldırmada etkili olduğunu göstermiştir.

Bu tez çalışmasında ortaya atılan "ÇOKLU-ÖLÇÜTLÜ ORTAK FİLTRELEME SİSTEMLERİNDE KABA KÜMELER KULLANIMI" ile ÇÖOF sistemlerinin tahmin doğruluğunu artıracak yeni bir yöntem geliştirilmiş, literatüre KKT kullanılarak oluşturulan yeni bir toplama fonksiyonu tabanlı tavsiye sistemi yaklaşımı kazandırılmış, yine KKT yardımıyla seyreklik ve soğuk başlangıç gibi problemlerin önüne önemli ölçüde geçilebilmiştir.

KAYNAKÇA

Adomavicius, G., & Kwon, Y. (2007). New recommendation techniques for multicriteria rating systems. *Intelligent Systems, IEEE*, 22 (3), 48-55.

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

Adomavicius, G., Manouselis, N., & Kwon, Y. (2011). Multi-criteria recommender systems. *Recommender Systems Handbook*, Springer, 769-803.

Ahn, H. (2008). A new similarity measure for collaborative filtering to alleviate the new user cold-starting problem. *Information Sciences*, 178 (1), 37-51.

Aydemir, S. B., & Gündüz, S. Y. (2020). Interval Valued q-Rung Orthopair Fuzzy Prioritized Dual Muirhead Mean Operator and Their Application in Group Decision Making. *International Conference on Intelligent and Fuzzy Systems* (s. 440-447). Springer, Cham.

Beel, J., Langer, S., Nürnberger, A., & Genzmehr, M. (2013). The impact of demographics (age and gender) and other user-characteristics on evaluating recommender systems. *Proceedings of the 17th International Conference on Theory and Practice of Digital Libraries (TPDL'13)* (s. 400-404). Valetta, Malta: Springer.

Benayoun, R., Roy, B., & Sussman, B. (1966). ELECTRE: Une méthode pour guider le choix en présence de points de vue multiple. *Direction Scientifique*, Note de Travail, No. 49.

Bilge, A., & Kaleli, C. (2014). A multi-criteria item-based collaborative filtering framework. *11th International Joint Conference on Computer Science and Software Engineering (JCSSE'14)*, (s. 18-22).

Bobadilla, J., Ortega, F., Hernando, A., & Bernal, J. (2012). A collaborative filtering approach to mitigate the new user cold start problem. *Knowledge-Based Systems*, 26, 225-238.

Bonikowski, Z., Bryniarski, E., & Wybraniec-Skardowska, U. (1998). Extensions and intentions in the rough set theory. *Information Sciences*, 107 (1-4), 149-167.

Breese, J. S., Heckerman, D., & Kadie, C. (1998). Empirical analysis of predictive algorithms for collaborative filtering. *Proceedings of the 14th conference on Uncertainty in Artificial Intelligence (UAI'98)* (s. 43-52). San Francisco, California: Morgan-Kaufmann.

Chee, S. H., Han, J., & Wang, K. (2001). Rectree: An efficient collaborative filtering method. *Proceedings of the 3rd International Conference on Data Warehousing and Knowledge Discovery*, (s. 141-151).

Chen, H., Li, X., & Huang, Z. (2005). Link prediction approach to collaborative filtering. *Proceedings of the 5th ACM/IEEECS joint conference on Digital libraries* (s. 141-142). New York: ACM Press.

Chen, M. H., Teng, C. H., & Chang, P. C. (2015). Applying artificial immune systems to collaborative filtering for movie recommendation. *Advanced Engineering Informatics*, 29 (4), 830-839.

Chirita, P. A., Nejdl, W., & Zamfir, C. (2005). Preventing shilling attacks in online recommender systems. *Proceedings of the 7th annual ACM international workshop on Web information and data management (WIDM'05)*, (s. 67-74).

Cleger-Tamayo, S., Fernández-Luna, J. M., & Huete, J. F. (2012). Top-N news recommendations in digital newspapers. *Knowledge-Based Systems*, 27, 180-189.

Çekik, R., & Telçeken, S. (2016). A new classification method based on rough sets theory. *Soft Computing*, 22 (6), 1881-1889.

Dat, L. Q., Thong, N. T., Ali, M., Smarandache, F., Abdel-Basset, M., & Long, H. V. (2019). Linguistic approaches to interval complex neutrosophic sets in decision making. *IEEE Access*, 38902-38917.

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

Elahi, M., Ricci, F., & Rubens, N. (2016). A survey of active learning in collaborative filtering recommender systems. *Computer Science Review*, 20, 29-50.

Fernández-Tobías, I., Braunhofer, M., Elahi, M., Ricci, F., & Cantador, I. (2016). Alleviating the new user problem in collaborative filtering by exploiting personality information. *User Modeling and User-Adapted Interaction*, 26 (2-3), 221-255.

Fuchs, M., & Zanker, M. (2012). Multi-criteria ratings for recommender systems: an empirical analysis in the tourism domain. *International Conference on Electronic Commerce and Web Technologies* (s. 100-111). Springer, Berlin, Heidelberg.

Gao, M., Wu, Z., & Jiang, F. (2011). Userrank for item-based collaborative filtering recommendation. *Information Processing Letters*, 111 (9), 440-446.

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

Goldberg, K., Roeder, T., Gupta, D., & Perkins, C. (2001). Eigentaste: A constant time collaborative filtering algorithm. *Information Retrieval*, 4 (2), 133-151.

Gravetter, F. J., Wallnau, L. B., Forzano, L. A., & Witnauer, J. E. (2020). *Essentials of statistics for the behavioral sciences.* New York: Cengage Learning.

Greco, S., Matarazzo, B., & Slowinski, R. (2001). Rough sets theory for multicriteria decision analysis. *European Journal of Operational Research*, 129 (3), 1-47.

Guo, S. (2014). Analysis and evaluation of similarity metrics in collaborative filtering recommender system. *Master's Thesis, Lapland University of Applied Sciences*.

Han, J., & Kamber, M. (2006). *Data Mining: Concepts and techniques.* Morgan Kaufmann.

Herlocker, J. L., Konstan, J. A., Borchers, A., & Riedl, J. (1999). An algorithmic framework for performing collaborative filtering. *Proceedings of the 22nd Annual International Conference on Research and Development in Information Retrieval (ACM SIGIR'99)* (s. 230-237). Berkeley California: ACM.

Hinduja, A., & Pandey, M. (2018). An intuitionistic fuzzy AHP based multi criteria recommender system for life insurance products. *International Journal of Advanced Studies in Computer Science and Engineering*, 7 (1), 1-8.

Hu, Y. C. (2014). Nonadditive similarity-based single-layer perceptron for multi-criteria collaborative filtering. *Neurocomputing*, 129, 306-314.

Hwang, C. L., Lai, Y. J., & Liu, T. Y. (1993). A new approach for multiple objective decision making. *Computers and Operations Research*, 20 (8), 889-899.

Jannach, D., Gedikli, F., Karakaya, Z., & Juwig, O. (2012). Recommending hotels based on multi-dimensional customer ratings. *Information and Communication Technologies in Tourism*, 320-331.

Jannach, D., Karakaya, Z., & Gedikli, F. (2012). Accuracy improvements for multi-criteria recommender systems. *Proceedings of the 13th ACM Conference on Electronic Commerce (EC '12)* (s. 674-689). Valencia, Spain: ACM.

Jia, Z., Yang, Y., Gao, W., & Chen, X. (2015). User-based collaborative filtering for tourist attraction recommendations. *IEEE International Conference on Computational Intelligence & Communication Technology (CICT '15)* (s. 22-25). Ghaziabad, India: CPS.

Jiang, S., Qian, X., Shen, J. , Fu, Y., & Mei, T. (2015). Author topic model based collaborative filtering for personalized POI recommendations. *IEEE Transactions on Multimedia*, 17 (6), 907-918.

Joshi, D., & Kumar, S. (2016). Interval-valued intuitionistic hesitant fuzzy Choquet integral based TOPSIS method for multi-criteria group decision making. *European Journal of Operational Research*, 248 (1), 183-191.

Jung, K. Y., Park, D. H., & Lee, J. H. (2004). Hybrid collaborative filtering and content-based filtering for improved recommender system. *International Conference on Computational Science (ICCS'04)*, (s. 295-302). Krakow, Poland.

Kermany, N. R., & Alizadeh, S. H. (2017). A hybrid multi-criteria recommender system using ontology and neuro-fuzzy techniques. *Electronic Commerce Research and Applications*, 21, 50-64.

Koohi, H., & Kiani, K. (2016). User based collaborative filtering using fuzzy c-means. *Measurement*, 91, 134-139.

Koren, Y. (2010). Factor in the neighbors: Scalable and accurate collaborative filtering. *ACM Transactions on Knowledge Discovery from Data*, 4 (1), 1-24.

Koren, Y., & Sill, J. (2013). Collaborative filtering on ordinal user feedback. *Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI '13)*, (s. 3022-3026). Beijing, China.

Lemire, D., & Maclachlan, A. (2005). Slope one predictors for online rating-based collaborative filtering. *Proceedings of the SIAM International Conference on Data Mining (SDM '05)*, (s. 471-475).

Li, D., Chen, C., Lv, Q., Shang, L., Zhao, Y., Lu, T., & Gu, N. . (2016). An algorithm for efficient privacy-preserving item-based collaborative filtering. *Future Generation Computer Systems*, 55, 311-320.

Li, Q. , & Kim, B. M. (2003). Clustering approach for hybrid recommender systems. *Proceedings of the IEEE/WIC International Conference on Web Intelligence (WI'03)*, (s. 33-38).

Li, Q., Myaeng, S. H., & Kim, B. M. (2007). A probabilistic music recommender considering user opinions and audio features. *Information Processing & Management*, 43 (2), 473-487.

Li, Q., Wang, C., & Geng, G. (2008). Improving personalized services in mobile commerce by a novel multicriteria rating approach. *Proceedings of the 17th International Conference on World Wide Web*, (s. 1235-1236).

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

Liu, L., Mehandijev, N., & Xu, D. L. (2011). Multi-criteria service recommendation based on user criteria preferences. *Proceedings of the 5th ACM conference on Recommender systems (RecSys '11)*, (s. 77-84).

Lu, H., Jiang, S., Song, W., & Ming, X. (2018). A Rough Multi-Criteria Decision-Making Approach for Sustainable Supplier Selection under Vague Environment. *Sustainability*, 10 (8), 2622-2641.

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

Manouselis, N., & Costopolou, C. (2007). Experimental analysis of design choices in multiattribute utility collaborative filtering. *International Journal of Pattern Recognition and Artificial Intelligence*, 21 (2), 311-331.

Mathew, P., Kuriakose, B., & Hegde, V. (2016). Book recommendation system through content based and collaborative filtering. *2016 International Conference on Data Mining and Advanced Computing (SAPIENCE)* (s. 47-52). Ernakulam, India: IEEE.

Medhi, J. (1992). *Statistical methods: an introductory text.* New Delhi: Wiley Eastern Limited.

Mikeli, A., Apostolou, D., & Despotis, D. (2013). A multi-criteria recommendation method for interval scaled ratings. *Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, (s. 9-12). Atlanta, USA.

Mild, A., & Reutterer, T. (2003). An improved collaborative filtering approach for predicting cross-category purchases based on binary market basket data. *Journal of Retailing and Consumer Services*, 10 (3), 123-133.

Nilashi, M., bin Ibrahim, O., & Ithin, N. (2014a). Multi-criteria collaborative filtering with high accuracy using higher order singular value decomposition and Neuro-Fuzzy system. *Knowledge-Based Systems*, 60, 82-101.

Nilashi, M., bin Ibrahim, O., & Ithin, N. (2014b). Hybrid recommendation approaches for multi-criteria collaborative filtering. *Expert Systems with Applications*, 41 (8), 3879-3900.

Nilashi, M., bin Ibrahim, O., Ithin, N., & Sarmin, N. H. (2015). A multi-criteria collaborative filtering recommender system for the tourism domain using

Expectation Maximization (EM) and PCA–ANFIS. *Electronic Commerce Research and Applications*, 14 (6), 542-562.

Nilashi, M., Esfahani, M. D., Roudbaraki, M. Z., Ramayah, T., & bin Ibrahim, O. (2016). A multi-criteria collaborative filtering recommender system using clustering and regression techniques. *Journal of Soft Computing and Decision Support Systems*, 3 (5), 24-30.

Nilashi, M., Jannach, D., bin Ibrahim, O., & Ithin, N. (2015). Clustering-and regression-based multi-criteria collaborative filtering with incremental updates. *Information Sciences*, 293, 235-250.

Oard, D. W., & Kim, J. (1998). Implicit feedback for recommender systems. *Proceedings of the 5th AAAI workshop on Recommender Systems*, (s. 81-83).

O'Mahony, M. P. (2004). *Towards robust and efficient automated collaborative filtering.* University College Dublin, Ireland: Ph.D Dissertation.

Ortega, F., Hernando, A., Bobadilla, J., & Kang, J. H. (2016). Recommending items to group of users using matrix factorization based collaborative filtering. *Information Sciences*, 345, 313-324.

Pamučar, D., Mihajlovič, M., Obradovič, R., & Atanaskovič, P. (2017). Novel approach to group multi-criteria decision making based on interval rough numbers: Hybrid DEMATEL-ANP-MAIRCA model. *Expert Systems with Applications*, 88, 58-80.

Park, S. T., & Pennock, D. M. (2007). Applying collaborative filtering techniques to movie search for better ranking and browsing. *Proceedings of the 13th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'07)*, (s. 550-559).

Pawlak, Z. (1982). Rough sets. *International Journal of Computer & Information Sciences*, 11 (5), 341-356.

Pawlak, Z., & Sowinski, R. (1994). Rough set approach to multi-attribute decision analysis. *European Journal of Operational Research*, 72 (3), 443-459.

Pennock, D. M., Horvitz, E., Lawrence, S., & Giles, C. L. (2000). Collaborative filtering by personality diagnosis: A hybrid memory- and model-based approach.

Proceedings of the 16th Annual Conference on Uncertainty in Artificial Intelligence (UAI-2000) (s. 473-480). San Francisco, California: Morgan Kaufmann.

Polat, H., & Du, W. (2003). Privacy-preserving collaborative filtering using randomized perturbation techniques. *Proceedings of the 3rd IEEE International Conference on Data Mining (ICDM '03)*, (s. 625-628).

Rashid, T., Faizi, S., Xu, Z., & Zafar, S. (2018). ELECTRE-based outranking method for multi-criteria decision making using hesitant intuitionistic fuzzy linguistic term sets. *International Journal of Fuzzy Systems*, 20 (1), 78-92.

Resnick, P., & Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40 (3), 56-58.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., & Riedl, J. (1994). GroupLens: an open architecture for collaborative filtering of netnews. *Proceedings of the 1994 ACM Conference On Computer Supported Cooperative Work*, (s. 175-186).

Ricci, F., Rokach, L., & Shapira, B. (2011). *Introduction to Recommender Systems Handbook.* Springer US.

Roy, B. (1996). *Multicriteria methodology for decision aiding.* Kluwer Academic Publishers.

Saaty, T. L. (1986). Axiomatic foundation of the analytic hierarchy process. *Management Science*, 32 (7), 841-855.

Saaty, T. L. (1996). *Decision making with dependence and feedback: The analytic network process.* Pittsburg, PA: RWS Publications.

Sahoo, N., Krishnan, R., Duncan, G., & Callan, J. P. (2006). Collaborative filtering with multi-component rating for recommender systems. *Proceedings of the 16th Workshop On Information Technologies And Systems (WITS'06).* Milwaukee.

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

Shambour, Q., & Lu, J. (2011). A hybrid multi-criteria semantic-enhanced collaborative filtering approach for personalized recommendations. *Proceedings of the 2011 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT)*, (s. 71-78).

Shan, H., & Banerjee, A. (2010). Generalized probabilistic matrix factorizations for collaborative filtering. *Proceedings of the 2010 IEEE International Conference on Data Mining*, (s. 1025-1030).

Shidpour, H., Da Cunha, C., & Bernard, A. (2016). Group multi-criteria design concept evaluation using combined rough set theory and fuzzy set theory. *Expert Systems with Applications*, 64, 633-644.

Si, L., & Jin, R. (2003). Flexible mixture model for collaborative filtering. *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*, (s. 704-711).

Song, W., & Sakao, T. (2018). An environmentally conscious PSS recommendation method based on users' vague ratings: A rough multi-criteria approach. *Journal of Cleaner Production*, 172, 1592-1606.

Su, X., & Khoshgoftaar, T. M. (2009). A survey of collaborative filtering techniques. *Advances in Artificial Intelligence*, 1-20.

Takacs, G., Pilaszy, I., Nemeth, B. , & Tikk, D. (2009). Scalable collaborative filtering approaches for large recommender systems. *Journal of Machine Learning Research*, 10, 623-656.

Tan, S., Bu, J., Chen, C., Xu, B. , Wang, C., & He, X. (2011). Using rich social media information for music recommendation via hypergraph model. *ACM Transactions on Multimedia Computing Communications and Applications*, 7 (1), 22.

Telçeken, S., & Doğan, M. (2004). Kaba kümeler teorisi yardımı ile büyük veri topluluklarının Analizi. *Eleco2004*, (s. 410-414).

Teng, W. G., & Lee, H. H. (2007). Collaborative recommendation with multi-criteria ratings. *Journal of Computers*, 17 (4), 69-78.

Thakkar, P., Varma, K. , Ukani, V., Mankad, S., & Tanwar, S. (2019). Combining user-based and item-based collaborative filtering using machine learning. *Information and Communication Technology for Intelligent Systems*, 173-180.

Ungar, L. H., & Foster, D. P. (1998). Clustering methods for collaborative filtering. *Proceedings of the Workshop on Recommendation Systems* (s. 114-129). AAAI Press.

Vasilijević, M., Fazlollahtabar, H., Stević, Ž., & Vesković, S. (2018). A rough multicriteria approach for evaluation of the supplier criteria in automotive industry. *Decision Making: Applications in Management and Engineering*, 1 (1), 82-96.

Vozalis, M. G., & Margaritis, K. G. (2007). Using SVD and demographic data for the enhancement of generalized collaborative filtering. *Information Sciences*, 177 (15), 3017-3037.

Wang, J., De Vries, A. P., & Reinders, M. J. (2006). Unifying user-based and item-based collaborative filtering approaches by similarity fusion. *Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (ACM SIGIR'06)*, (s. 501-508).

Wei, J., He, J., Chen, K., Zhou, Y., & Tang, Z. (2017). Collaborative filtering and deep learning based recommendation system for cold start items. *Expert Systems with Applications*, 69, 29-39.

Yang, S., Korayem, M., AlJadda, K., Grainger, T., & Natarajan, S. (2017). Combining content-based and collaborative filtering for job recommendation system: A cost-sensitive Statistical Relational Learning approach. *Knowledge-Based Systems*, 136, 37-45.

Yi, X., Hong, L., Zhong, E., Liu, N. N., & Rajan, S. (2014). Beyond clicks: dwell time for personalization. *Proceedings of the 8th ACM conference on Recommender systems (RecSys '14)*, (s. 113-120).

Yirgu, K. D., & Meier, A. (2016). *Propagation and Aggregation in Trust Enhanced Recommender Systems.* Fribourg, Switzerland.

Zarrinkalam, F., & Kahani, M. (2012). A multi-criteria hybrid citation recommendation system based on linked data. *Proceedings of the 2nd International e-Conference on Computer and Knowledge Engineering (ICCKE'12)*, (s. 283-288).

Zhang, J., Zhu, J., & Zhang, N. (2014). An Improved Privacy-Preserving Collaborative Filtering Recommendation Algorithm. *Proceedings of the Pacific Asia Conference on Information Systems*, (s. 277). Chengdu, China.

Zhang, S., Wang, W., Ford, J., Makedon, F., & Pearlman, J. (2005). Using singular value decomposition approximation for collaborative filtering. *Proceedings of the 7th IEEE International Conference on E-Commerce Technology (CEC'05)*, (s. 257-264).

Zhao, Z. D., & Shang, M. S. (2010). User-based collaborative-filtering recommendation algorithms on hadoop. *Proceedings of the 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'10)*, (s. 478-481).

Zheng, N., & Li, Q. (2011). A recommender system based on tag and time information for social tagging systems. *Expert Systems with Applications*, 38 (4), 4575-4587.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı ve Soyadı : Emin Talip DEMİRKIRAN

Uyruğu : TC

Doğum yeri ve Tarihi

EĞİTİM BİLGİLERİ

Derece	Mezun olduğu okul	Mezuniyet yılı
Lisans	Anadolu Üniversitesi	2011
Lise	Batman Fen Lisesi	2005

GÖREVLER

Görevi	Görev Yeri	Görev Yılı
Araştırma Görevlisi	Eskişehir Teknik Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü Bilgisayar Donanımı Anabilim Dalı	2018 -
Araştırma Görevlisi	Anadolu Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü Bilgisayar Donanımı Anabilim Dalı	2012-2018
Araştırma Görevlisi	Batman Üniversitesi Mühendislik-Mimarlik Fakültesi Bilgisayar Mühendisliği Bölümü Bilgisayar Mühendisliği Kuramsal Temeller Anabilim Dalı	2011-2012

PROJELERDE YAPTIĞI GÖREVLER

1. Developing Techniques for Improving the Robustness of Privacy-Preserving Collaborative Filtering Schemes, TÜBİTAK PROJESİ, Araştırmacı, 01/03/2012 - 01/03/2014 (ULUSAL- Proje No: 2654681)

ULUSLARARASI BİLİMSEL TOPLANTILARDA SUNULAN VE BİLDİRİ KİTAPLARINDA BASILAN BİLDİRİLER

1. Demirkıran E. T. (2016). Recommender Systems for e-Learning Environments. Internation Conference on Research in Education & Science (ICRES2016), 377-381. (Tam Metin Bildiri/Sözlü Sunum) (Yayın No: 3583987)
2. Demirkıran E. T., Türk A. M., & Kaleli C. (2014). A Bicriteria Clustering Approach for Collaborative Filtering. The Third International Conference on Informatics Engineering and Information Science (ICIEIS2014), 11(12), 79-90. (Tam Metin Bildiri/Sözlü Sunum) (Yayın No: 1987159)
3. Demirkıran E. T. (2019). COLLABORATIVE FILTERING RECOMMENDER SYSTEMS: AN APPLICATION. 8th International Conference on Advanced Technologies (ICAT2019), 241-243. (Tam Metin Bildiri/Sözlü Sunum) (Yayın No: 6987654)

ULUSLARARASI HAKEMLİ DERGİLERDE YAYIMLANAN MAKALELER

1. Demirkıran, E. T., Pak, M. Y., & Çekik R. (2021). Multi-Criteria Collaborative Filtering Using Rough Sets Theory. Journal of Intelligent & Fuzzy Systems, 40 (1), 907-917. DOI: 10.3233/JIFS-201073 (Yayın No: 6986752)