

A novel hotel recommender system incorporating review sentiment and contextual information

ABSTRACT

Otel öneri sistemleri, turistlerin gereksiz otel bilgilerini filtrelemelerine yardımcı olarak, tercihleriyle uyumlu konaklama seçeneklerini bulmalarını sağlayan kritik araçlardır. Bu sistemler temel olarak içerik tabanlı filtreleme (Content-Based Filtering, CBF) ile kullanıcı tabanlı (User-Based Collaborative Filtering, UCF) ve öge tabanlı (Item-Based Collaborative Filtering, ICF) olmak üzere işbirlikçi filtreleme (Collaborative Filtering, CF) yöntemlerini kullanır. Buna ek olarak, bağlam farkındalıklı öneri sistemleri (Context-Aware Recommender Systems, CARS), karar verme sürecine ilgili bağlamsal verileri entegre ederek önerilerin kalitesini artırır.

Bununla birlikte, bu sistemlerin karşılaştığı önemli bir zorluk, kullanıcı-öge matrisinde (seyahat eden-otel matrisi olarak da bilinir) veri seyrekliğidir. Veri seyrekliği, tüm kullanıcıların her otele puan vermemesi ve bazılarının hiç puan vermemesi nedeniyle ortaya çıkar; bu durum, çok sayıda eksik puan ve kullanıcı tercihlerini doğru şekilde tahmin etmek için yetersiz veri anlamına gelir.

Bu çalışma, CBF, UCF, ICF ve CARS yöntemlerini bütünleştiren yeni bir hibrit otel öneri sistemi önermektedir. Sistemimizde yenilikçi bir unsur olarak, bağlamsal bilgi türü olarak seyahat türleri kullanılmaktadır. Ayrıca, otel yorumlarından duygu verisi elde etmek amacıyla duygu analizi (Sentiment Analysis, SA) teknikleri uygulanmaktadır. Elde edilen bu duygu verileri, matristeki eksik puanların tahmin edilmesinde kullanılarak veri seyrekliği sorunu ele almaktadır.

TripAdvisor.com'dan alınan bir veri kümesi üzerinde gerçekleştirilen deneysel analizler, önerilen sistemimizin, geleneksel UCF yaklaşımlarının yanı sıra CBF, UCF, ICF ve bağlamsal bilgiyi birleştiren ancak SA içermeyen diğer yöntemlerden anlamlı ölçüde daha iyi performans gösterdiğini ortaya koymuştur. Özellikle, sistemimiz temel modellere kıyasla anlamlı derecede daha düşük Ortalama Mutlak Hata (Mean Absolute Error, MAE) ve Kök Ortalama Kare Hatası (Root Mean Square Error, RMSE) değerleri sergilemiştir.

Bu gelişme, yalnızca veri seyrekliği sorunu azaltmakla kalmayıp, aynı zamanda seyahat edenler için otel önerilerinin doğruluğunu ve güvenilirliğini de artırmaktadır.

1. INTRODUCTION

Son birkaç on yılda turizm sektörü önemli bir büyüme göstermiş ve küresel ölçekte en hızlı büyüyen sektörlerden biri haline gelmiştir [1–5]. Özellikle 2022 yılında, uluslararası turist seyahatleri bir önceki yıla göre iki katına çıkarak 900 milyonu aşmıştır [6]. İnternet teknolojileri ve mobil cihazlardaki gelişmeler, çevrimiçi seyahat bilgilerine erişimi büyük ölçüde kolaylaştırmıştır [7]. Ancak, çevrimiçi bilgi miktarındaki hızlı artış, bilgi fazlalığı (information overload) sorunu ortaya çıkarmıştır [8, 9]. Bu durum, turizm alanına da yansımış ve turistlerin kendileri için gerçekten ilgili olan otel bilgilerini devasa veri yığınları

arasından seçmelerini zorlaştırmıştır [2, 10]. Bu sorunu çözmek amacıyla, gereksiz bilgileri filtreleyerek verimli çözümler sunan otel öneri sistemleri geliştirilmiştir [11–13].

Öneri sistemlerinin (Recommender Systems, RS) temel amacı, büyük veri kümeleri içinden en alakalı bilgi ve hizmetleri seçerek bilgi fazlalığı sorununu hafifletmek ve kişiselleştirilmiş hizmetler sunmaktır [14–16]. RS alanındaki temel yaklaşımlar, içerik tabanlı filtreleme (Content-Based Filtering, CBF) ve işbirlikçi filtreleme (Collaborative Filtering, CF) yöntemleridir. CF, kullanıcı tabanlı (User-Based CF, UCF) ve öge tabanlı (Item-Based CF, ICF) olmak üzere iki alt tür içerir; bu yöntemler sırasıyla kullanıcı veya öge benzerliklerine dayalı olarak öneriler üretir [17–24].

Buna ek olarak, bağlam farkındalıklı öneri sistemleri (Context-Aware Recommender Systems, CARS), zaman, konum veya birlikte seyahat edilen kişiler gibi bağlamsal bilgileri kullanarak önerilerin etkinliğini önemli ölçüde artıran sistemler olarak ortaya çıkmıştır. Bu durum, seyahat türünün (örneğin aile, çift, tek başına, arkadaş veya iş amaçlı seyahat) önemli bir bağlam unsuru olduğu otel öneri sistemleri için özellikle önemlidir [25–27]. Bu tür bağlamsal veriler, öneri doğruluğunu artıran ek bilgi katmanları sağlar. Bu durum, genellikle yalnızca kullanıcı–öge etkileşimlerine dayalı olarak öneri üreten uygulamalarda özellikle faydalıdır [27–33].

Otel öneri sistemleri alanındaki araştırmalar, CARS yöntemlerinin geleneksel öneri sistemlerine kıyasla müşteri bilgilerini toplamda daha başarılı olduğunu göstermiştir [34, 35]. Ancak, pratikte her kullanıcının her ögeye puan vermemesi, hatta bazı kullanıcıların hiç puan vermemesi, son derece seyrek bir kullanıcı–öge puanlama matrisi oluşmasına neden olmaktadır [36, 37]. Bu tür veri seyrekliği, öneri sistemleri için ciddi bir engel teşkil eder; zira kullanıcı puanlarının eksikliği, doğru öneriler üretmeyi zorlaştırır. Bu sorun, bağlamsal bilgiye göre veri kümelerinin alt gruplara ayrıldığı CARS yaklaşımlarında daha da belirgin hale gelir; zira bu bölünme, matrislerin yoğunluğunu daha da azaltarak seyrekliği artırır [28].

Duygu analizi (Sentiment Analysis, SA), kullanıcı yorumlarından gizli içgörülerini ortaya çıkarmak için çeşitli çalışmalarda kullanılmış ve veri seyrekliği sorununu hafifletmede önemli bir rol oynamıştır [38–41]. Bu çalışmadaki hedefimiz, otel önerileri için özel olarak tasarlanmış, bağlam ve duygu farkındalıklı yöntemleri geleneksel CBF ve CF teknikleriyle entegre eden yenilikçi bir hibrit öneri sistemi geliştirmektir. Bu kapsamda, eksik puanları tahmin etmeye yönelik yeni bir strateji önererek CARS içindeki veri seyrekliği sorununu etkili bir şekilde ele almaktayız. Bu çalışma, otel öneri sistemleri alanındaki mevcut araştırmalara katkı sağlamayı, henüz keşfedilmemiş alanlara ışık tutmayı ve bu alanın en kalıcı sorunlarından birine umut vadeden bir çözüm sunmayı amaçlamaktadır.

Makale şu şekilde yapılandırılmıştır: Bölüm 2, otel önerilerinde bağlam ve duygu farkındalıklı teknikleri incelemekte; Bölüm 3, önerilen hibrit sistemi açıklamakta; Bölüm 4, deneysel düzeni ve sonuçları detaylandırmakta; Bölüm 5 ise çalışmayı sonuçlandırmaktadır.

2 Literature review

2.1 Context-aware recommender systems

CARS (Bağlam Farkındalıklı Öneri Sistemleri) içinde, kullanıcıya ait bağlamsal bilgiler, CBF ve/veya CF teknikleri kullanılarak öneri sürecine entegre edilir. Kullanıcının mevcut görevi,

günün saati, bulunduğu konum ve kullandığı cihaz gibi bağlamsal unsurlar, sunulan önerilerin uygunluğunu önemli ölçüde etkiler [42]. Adomavicius ve Tuzhilin [43], bağlamı; yerler, olaylar, kişiler ve nesneler gibi ilgi alanı nesneleriyle ilişkili durumları karakterize eden bilgiler olarak tanımlar ve bağlam kavramının, yalnızca sözlükteki “bir durumu etkileyen koşullar” anlamının ötesine geçtiğini, ilgili disiplin veya problem alanına özgü bir nitelik taşıdığını belirtir.

CARS üzerine kapsamlı bir inceleme sunan Adomavicius ve Tuzhilin [43], CARS teknolojilerine genel bir bakış sağlamaktadır. Abbas ve arkadaşları [28] ise hesaplamalı zekâ teknikleri kullanarak CARS yaklaşımlarını karşılaştırmaktadır. Bennett ve arkadaşları [44], bağlamsal bilgilerden yararlanan görev odaklı arama süreçlerini derinlemesine incelemiştir. Ayrıca Casillo ve arkadaşları [25], CARS hakkında önemli içgörüler sunarak bu sistemlerin kültürel miras sektöründeki uygulama zorluklarını vurgulamaktadır. Son olarak Nadeem ve arkadaşları [26], CARS’ın farklı alanlardaki avantajlarını inceleyerek, bu yaklaşımın öneri sistemlerini geliştirmedeki çok yönlülüğünü ve etkisini öne çıkarmaktadır.

2.2 Sentiment analysis in RSs

SA (Sentiment Analysis – Duygu Analizi), aynı zamanda *görüş madenciliği* (opinion mining) olarak da bilinir ve bireylerin çeşitli varlıklar hakkındaki görüşlerini, tutumlarını ve duygularını belirlemek ve çıkarmak için doğal dil işleme (Natural Language Processing), metin analizi ve hesaplamalı dilbilim (computational linguistics) tekniklerinden yararlanır [45–48]. Özellikle SA’nın amacı, bir cümle veya paragrafın olumlu, olumsuz ya da nötr bir duygu içerip içermediğini belirlemek ve böylece belirli ürünler veya hizmetler hakkındaki müşteri tercihleri hakkında içgörüler elde etmektir.

SA temelde iki ana yaklaşımla gerçekleştirilir: **sınıflandırma temelli yaklaşım** ve **sözlük temelli (lexicon-based) yaklaşım**.

Sınıflandırma temelli yaklaşım, makine öğrenmesi yöntemlerini kullanır; burada doğal dil işleme teknikleriyle elde edilen özenle seçilmiş metinsel özellikler kullanılarak sınıflandırıcılar (classifiers) geliştirilir. Bu sınıflandırıcılar, metnin duygu yönelimini belirlemek üzere tasarlanmıştır ve metni “olumlu”, “olumsuz” veya “nötr” olarak kategorize eder.

Buna karşılık, sözlük temelli yaklaşım bir **duygu sözlüğüne (sentiment lexicon)** dayanır — bu sözlük, duygu yüklü kelime, ifade ve deyimlerden oluşan özel olarak oluşturulmuş bir koleksiyondur. Sözlük, manuel olarak derlenebileceği gibi, başlangıçta belirlenmiş bir kelime çekirdeğinden otomatik olarak da üretilebilir. Bu görüş yüklü kelime derlemesi sayesinde, metinlerin duygu yönelimi tanımlanabilir ve değerlendirilebilir.

Öneri sistemleri (RS) alanında SA, kişiselleştirme ve öneri doğruluğunu artırmak amacıyla çeşitli alanlara özgü problemler üzerinde uygulanmıştır. Bu uygulamalara örnek olarak giyim [29], eğitim/ders [49], otel [50] ve diğer ürün önerileri [51] verilebilir. Bu durum, SA’nın kullanıcı tercihlerini anlama ve onlara uygun öneriler sunma konusundaki çok yönlülüğünü ve etkinliğini açıkça ortaya koymaktadır.

2.3 Related works on hotel RSs

Tablo 1, çeşitli otel öneri sistemlerinin karşılaştırmalı bir analizini sunmakta olup; veri kaynakları, öneri teknikleri (CBF ve CF dâhil), bağlamsal bilgiler ve SA (Duygu Analizi) tekniklerinin uygulanması gibi unsurları öne çıkarmaktadır. Analiz, öneri sürecine bağlamsal bilginin entegre edilmesine yönelik belirgin bir odaklanmanın olduğunu ve bunun umut verici sonuçlar doğurduğunu ortaya koymaktadır. Ayrıca, UCF (Kullanıcı Tabanlı İşbirlikçi Filtreleme), hedef kullanıcıya benzer kullanıcıların belirlenmesinde kritik bir teknik olarak tanımlanmakta ve bu sayede kişiselleştirilmiş önerilerin kolaylaştırıldığı vurgulanmaktadır. Çoğu durumda UCF, otel puanlamalarını daha etkili şekilde ele almak amacıyla CBF (İçerik Tabanlı Filtreleme) ile hibrit bir yaklaşımla birleştirilmektedir.

Hu ve Chen [52], otel yorumlarının öneri sonuçlarını iyileştirebilecek zengin bir bilgi kaynağı olduğunu belirtmektedir. Kumar ve arkadaşlarının [18] çalışması gibi yakın dönem araştırmalar, otel yorumlarının duygu durumunu değerlendirmek için SA kullanımını incelemiştir. Bu çalışmada, yedi sınıflandırıcı ve üç farklı kelime gömme (word embedding) tekniğini karşılaştıran bir topluluk öğrenme (ensemble learning) modeli kullanılmış ve puan tahmininde doğruluğu artırmak amaçlanmıştır.

Bu sistemlerde öne çıkan önemli bir zorluk, kullanıcı–öge puanlama matrisinin seyrekliğidir; bu matris genellikle her kullanıcıdan kapsamlı otel puanlarını içermemektedir. Bu durum, benzer otellerin tanımlanması ve hedef bir otel için puan tahmini yapılması süreçlerinde ICF'nin (Öge Tabanlı İşbirlikçi Filtreleme) yeterince kullanılmamasına yol açmaktadır. Bu sorunu ele almak amacıyla Hu ve arkadaşları [53], kullanıcı özelliklerine dayalı kullanıcı–kullanıcı benzerliğini hesaplayarak eksik puanları tahmin eden bir yöntem geliştirmiş ve hem UCF hem de ICF tekniklerinden tam anlamıyla yararlanmayı hedeflemiştir.

Sonuç olarak, bağlamsal bilgilerin, otel yorumlarından elde edilen duygu analizinin ve eksik puanların stratejik biçimde tahmin edilmesinin birleşimi, etkili bir otel öneri sistemi geliştirmeye yönelik kapsamlı bir yaklaşım sunmaktadır. UCF, ICF ve CBF'nin bütünleştirildiği bu çok boyutlu strateji, öneri performansını artırmak için yenilikçi ve potansiyel olarak daha sağlam bir çözüm ortaya koymakta; otel öneri sistemleri alanında öncü bir adımı temsil etmektedir.

Table 1 Comparisons of hotel recommender systems

Work	Data source	SA	Contextual information	CBF	CF	
					UCF	ICF
Zheng et al. [59]	TripAdvisor.com		v	v	v	
Levi et al. [60]	Venere.com TripAdvisor.com	v	v		v	
Sharma et al. [61]	Booking.co			v	v	
Zhang et al. [62]	Ctrip.com		v	v	v	
Hu et al. [53]	TripAdvisor.com		v	v	v	v
Shrote and Deorankar [63]	Not mentioned				v	
Jalan and Gawande [64]	TripAdvisor.com	v	v	v		
Ray et al. [65]	TripAdvisor.com	v		v		
Kumar et al. [50]	TripAdvisor.com	v			v	v
Our study	TripAdvisor.com	v	v	v	v	v

3 The proposed hybrid hotel recommender system

Şekil 1, önerilen hibrit otel öneri sisteminin akış şemasını göstermektedir. İlk aşamada veriler, TripAdvisor.com'dan alınmakta olup; oteller, gezgin profilleri, yorumlar, puanlar ve seyahat türleri gibi bilgileri kapsamaktadır. Gezinler tarafından verilen otel puanlarına dayanarak bir **gezin-otel matrisi** oluşturulmakta ve matris değerleri, ilgili oteller için verilen puanları yansıtmaktadır. Bu matris daha sonra bağlamsal bilgi olarak kullanılan seyahat türlerine göre — **İş, Aile, Çift, Arkadaş ve Tekil (Solo)** — beş ayrı matrise bölünmektedir. Bunun sonucunda, bu matrisler yüksek derecede seyrek hale gelmekte (Bkz. Bölüm 3.1) ve bu durum çok sayıda eksik puan içermelerine, dolayısıyla **eksik değer tahmini (missing value imputation)** gerekliliğine yol açmaktadır [54].

Eksik verilerin tahmini süreci için, gezin tercihlerini belirlemeye yönelik olarak **yorum ön işleme (review preprocessing)** işlemleri gerçekleştirilir. Bu aşama; yazım denetimi, sözcük türü etiketleme (part-of-speech tagging) ve otel özelliklerinin tanımlanması adımlarını içerir (Bkz. Bölüm 2). Bu işlemlerin ardından, gezin yorumlarından daha derin içgörüler elde etmek amacıyla **duygu analizi (Sentiment Analysis, SA)** uygulanır ve böylece bir **otel-özellik matrisi** oluşturulur. Bu matris, **otel-otel benzerliği hesaplamaları, komşuluk seçimi** ve **puan tahmini** işlemleriyle birlikte kullanılarak, bağlama dayalı olarak seyrek matrislerin iyileştirilmesi için ICF (Öğre Tabanlı İşbirlikçi Filtreleme) yaklaşımında değerlendirilir. Sonuç olarak, beş adet zenginleştirilmiş gezin-otel matrisi elde edilir (Bkz. Bölüm 3.3).

Son aşama olan **nihai tahmin sürecinde, UCF (Kullanıcı Tabanlı İşbirlikçi Filtreleme)** yaklaşımı uygulanır. Bu aşamada **gezin-gezin benzerliği** ve **komşuluk seçimi** işlemleri gerçekleştirilerek, son otel önerileri elde edilir (Bkz. Bölüm 3.4).

3.1 Data collection and context partition

Bu çalışmada kullanılan veri kümesi, seyahat bilgileri konusunda önde gelen web sitelerinden biri olan **TripAdvisor.com**'dan elde edilmiştir. TripAdvisor, geniş otel yorumları ve gezin önerileri koleksiyonuyla tanınmakta olup, kullanıcılara oteller hakkında yorum ve puanlama yapma olanağı sağlamaktadır. Ayrıca kullanıcılar, yaptıkları yorumlara bağlamsal bir boyut kazandırmak amacıyla seyahat türlerini de sınıflandırabilmektedir. Bu veri koleksiyonu; otel özellikleri, gezin demografik bilgileri, yorum içerikleri, puanlamalar ve kategorize edilmiş seyahat niyetleri gibi bileşenleri kapsamaktadır.

Veri edinim süreci **Eylül 2009 ile Şubat 2017** tarihleri arasındaki dönemi kapsamaktadır ve **coğrafi olarak ABD'deki otellere** odaklanmaktadır. Web sitesinden elde edilen veri kümesi, otel özellikleri, gezin demografik bilgileri, yorumlar, puanlamalar ve seyahat niyetleriyle ilgili ayrıntıları içermektedir. Bu çalışma, özellikle **TripAdvisor'da 20'den fazla yorum paylaşmış ve "kıdemli katkıcı" (senior contributor)** olarak tanımlanan kullanıcılar tarafından yazılmış yorumlara odaklanmaktadır.

Analitik netlik sağlamak amacıyla, veri kümesi seyahat türüne göre beş farklı segmente ayrılmıştır. Bu ayrıştırma, seyahat türünü temel bir bağlamsal değişken olarak kullanarak, her bir seyahat kategorisi için beş ayrı **gezin-otel matrisi** oluşturulmasına olanak

tanımlanmıştır. İncelenen seyahat türü kategorileri şunlardır: **İş, Aile, Çift, Arkadaş ve Tekil (Solo)**. Bu doğrultuda, veri kümesi beş karşılık gelen alt kümeye bölünmüştür.

Bu bölünmüş veri kümelerine ait temel bilgiler **Tablo 2**'de sunulmuştur. Dikkate değer bir nokta olarak, bir gezginin aynı oteli aynı seyahat kategorisi altında birden fazla kez puanlaması durumunda, bu puanlar **ortalama alınarak birleştirilmiş**, böylece söz konusu otelin ilgili seyahat bağlamındaki konumu doğru bir biçimde yansıtılmıştır.

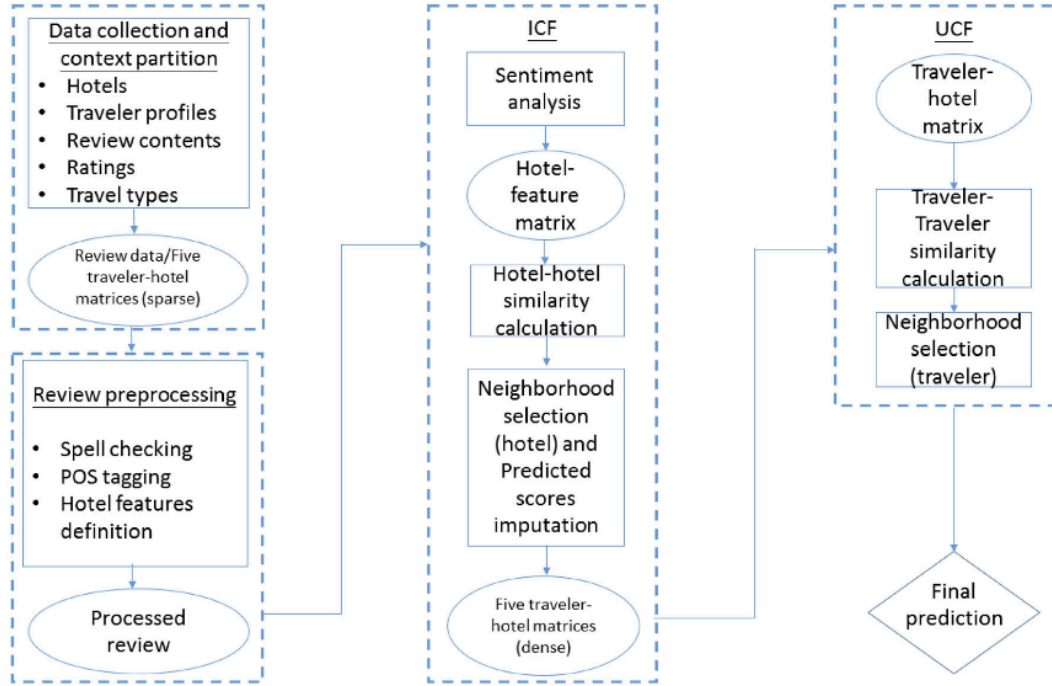


Fig. 1 The flow chart of the proposed system

3.2 Review preprocessing and hotel feature extraction

Gezginlerin otel tercihlerini anlamak amacıyla bir dizi **yorum ön işleme (review preprocessing)** adımı uygulanmıştır. Bu adımlar arasında **yazım denetimi, sözcük türü etiketleme (part-of-speech, POS tagging)** ve **temel otel özelliklerinin belirlenmesi** yer almaktadır. Yazım denetiminin amacı, yorumlardaki yazım ve dilbilgisi hatalarını düzeltmektir; buna, vurgulama amacıyla yapılan kasıtlı tekrarlar da dahildir (örneğin: “Otel çooooook güzeldi.”). Bu işlem için **Google Spell Check** servisi kullanılmıştır.

POS etiketleme işlemi, **Stanford CoreNLP** aracıyla gerçekleştirilmiş olup, kelimeler isim, sıfat, fiil ve zarf gibi kategorilere atanmıştır. Aynı zamanda durak kelimeler (stop words) yönetilmiş ve metin uygun şekilde bölümlere ayrılmıştır.

Bir sonraki adımda, **otel özelliklerinin belirlenmesi ve bu özelliklerle ilişkili terimlerin eşleştirilmesi** gerçekleştirilmiştir. Önceki araştırmalardan elde edilen bulgulara dayanarak, müşteri önceliklerini temsil eden yedi temel otel özelliğine odaklanılmıştır: **yemek (food)**, **tesisler (facilities)**, **temizlik (cleanliness)**, **hizmetler (services)**, **uyku kalitesi (sleep quality)**, **fiyatlar (prices)** ve **konum (location)**. Bu özellikler, Hu ve ark. [53], Sohrabi ve

ark. [55] ve Albaladejo-Pina ile Díaz-Delfa [56] tarafından yapılan çalışmalarda da tanımlanmıştır.

Terimlerin ilgili otel özellikleriyle doğru bir şekilde eşleştirilmesini sağlamak amacıyla, **Normalize Edilmiş Google Uzaklığı (Normalized Google Distance, NGD)** yöntemi kullanılmıştır. Bu teknik, terimler ile otel özellikleri arasındaki anlamsal yakınlığı ölçerek, bunların doğru şekilde kategorize edilmesini sağlar. Örneğin, “bathroom”, “shower” ve “toilet” gibi kelimeler benzer bağlamlarda kullanıldıkları için birbiriyle ilişkili olarak tanımlanır.

NGD tabanlı sınıflandırmanın ardından, iki uzmandan oluşan bir ekip tarafından **manuel inceleme** yapılmıştır. Bu adım, terimlerin otel özellikleriyle olan eşleştirmesini daha da hassaslaştırarak sınıflandırmanın doğruluğunu artırmıştır. **Tablo 3**, otel özellikleriyle bunlara karşılık gelen örnek terimleri göstermektedir.

Table 2 Basic information of the original dataset and five subsets

Dataset	Total ratings	Hotels	Travelers	Density (%)
Original	14,682	8304	278	0.59
Friends	747	673	151	0.70
Solo	1426	1296	156	0.69
Family	1769	1535	193	0.57
Couple	4231	3356	212	0.56
Business	6509	4070	218	0.68

Table 3 Examples of the hotel features and their representative terms

Hotel features	Representative terms
Food	Bar, coffee, meal, food, breakfast, restaurant, menu, drinks, buffet, water, lunch, dinner, etc.
Facilities	Hotel, room, TV, floor, lobby, phone, bathroom, ceiling, oven, gym, plug, etc.
Cleanliness	Smoking, toilet, laundry, housekeeping, cleanliness, etc.
Services	Service, staff, reception, helpfulness, greet, person, etc.
Sleep quality	Bed, night, noise, stay, quiet, sleeping, blanket, pillow, etc.
Prices	Money, charge, price, fee, payment, etc.
Location	Area, location, place, west, city, north, country, etc.

3.3 Hotel–feature matrix development and hotel–hotel similarity calculation

Bir dizi otel $H = \{h_1, h_2, \dots, h_x\}$ ve bir dizi otel özelliği $F = \{f_1, f_2, \dots, f_y\}$ düşünelim. Bu otel özellikleri; **yemek, tesisler, temizlik, hizmetler, uyku kalitesi, fiyatlar ve konum** kategorilerini kapsamaktadır. Her bir otel özelliği için duygu puanlarını hesaplayarak bir **otel-özellik matrisi** oluşturulur. Bu işlem, otel yorumlarındaki her bir cümle için ilgili terimlerin çıkarılması ve bu cümlelerin uygun otel özellikleriyle ilişkilendirilmesiyle gerçekleştirilir. Sonuç olarak, her bir otel özelliği, yorumlardan toplanan bir cümle koleksiyonunu temsil eder ve bu koleksiyonlar üzerinden **duygu puanları (sentiment scores)** hesaplanır.

Yorumlardan duygu ve öznel bilgi çıkarımı, **OpinionFinder** aracı kullanılarak yapılmıştır. Bu araç, **MPQA Corpus** adlı, 8221 kelimeden oluşan bir sözlükten yararlanır [57]. Bu derlemdaki her kelimeye bir **sözcük türü etiketi (part-of-speech category)**, bir **öznel tür (örneğin güçlü veya zayıf)** ve bir **duygu yönelimi (örneğin olumlu, olumsuz veya nötr)** atanmıştır. Olumlu bir duygu belirten kelimeler (örneğin “awesome”, “wonderful”) olumlu bir durumu ifade ederken, olumsuz bir duygu belirten kelimeler (örneğin “bad”, “poor”, “awful”) olumsuz bir durumu yansıtır. Ne olumlu ne de olumsuz duygu içeren kelimeler (örneğin “really”, “very”) **nötr** olarak sınıflandırılır. Bu çalışmada, yalnızca **olumlu duygu yönelimine sahip kelimeler** daha ayrıntılı şekilde işlenmiş ve analiz edilmiştir.

Bir otelin h_i 'sinin f_j özelliğiyle ilişkili **p. cümlesinin** ($p = 1, \dots, m$) duygu puanı SS_{h_i, f_j, s_p} , f_j , s_p olarak tanımlanır. Buna göre, otel h_i 'nin f_j özelliğine ilişkin genel duygu puanı şu formülle hesaplanır:

$$SS_{h_i, f_j} = \frac{\sum_{p=1}^m SS_{h_i, f_j, s_p}}{m}$$

Burada **m**, otel h_i 'nin f_j özelliğiyle ilişkili toplam cümle sayısını ifade eder. Bu yöntem, müşteri yorumlarından elde edilen duygu puanları aracılığıyla otel özelliklerinin ayrıntılı bir biçimde analiz edilmesine olanak tanır.

Örneğin, gezgin **Tom**'un otel **A** hakkındaki şu yorumunu ele alalım:

“Lobi çok güzel ve yemekler lezzetli.”

Bu cümle, iki farklı otel özelliğine — **tesisler (facilities)** ve **yemek (food)** — yönelik **olumlu bir duygu** ifade etmektedir. Bu durumda, otel **A**'nın hem tesisler hem de yemek özelliklerine ilişkin duygu puanları olumlu yönde etkilenebilir ve buna göre kaydedilir. Bir yorum belirli bir özelliği tekrarlayan biçimde olumlu olarak anıyorsa, bu özelliğin duygu puanı da kümülatif biçimde artar; bu da artan olumlu geri bildirimi yansıtır. Bu yaklaşım, yorumlarda yer alan her bir memnuniyet veya övgü ifadesinin, ilgili otel özelliklerinin genel duygu puanına doğrudan katkı sağlamasını garanti eder.

Otel–özellik matrisi oluşturulduktan sonra, iki otel arasındaki benzerliği ölçmek için **kosinüs benzerliği (cosine similarity)** yöntemi kullanılır. Özellikle, **n boyutlu vektörlerle** (özellikler) temsil edilen iki otel \mathbf{h}_x ve \mathbf{h}_y için kosinüs benzerliği, **Öklid iç çarpımı (Euclidean dot product)** formülüyle aşağıdaki şekilde hesaplanır:

$$\text{sim}(\mathbf{h}_x, \mathbf{h}_y) = \cos(\theta) = \frac{\vec{h}_x \cdot \vec{h}_y}{\|\vec{h}_x\| \|\vec{h}_y\|} = \frac{\sum_{j=1}^n h_{xj} \cdot h_{yj}}{\sqrt{\sum_{j=1}^n h_{xj}^2} \sqrt{\sum_{j=1}^n h_{yj}^2}}$$

Burada \mathbf{h}_x ve \mathbf{h}_y , sırasıyla otel \mathbf{h}_x ve otel \mathbf{h}_y 'nin \mathbf{f} özelliğine ilişkin duygu puanlarını temsil etmektedir.

3.4 Hotel rating imputation from the ratings of top-k nearest neighbor hotels

Gezgin–otel matrisindeki **eksik otel puanlarını** tahmin etmek (impute etmek) amacıyla, **otel–özellik matrisinden elde edilen otel–otel benzerliklerinden** yararlanılmıştır. Her bir otelin **en yakın k komşusu (top-k nearest neighbors)**, otel–otel benzerlik matrisi temel alınarak belirlenmiştir ve bu komşular, otel puanlarının tahmininde temel referans noktası olarak kullanılmıştır.

Bir gezgin \mathbf{tra} 'nın hedef otel \mathbf{h} için vermesi beklenen puan $\mathbf{R}_{a,i}$ olarak gösterilir ve **Resnick ve ark. [58]** tarafından önerilen aşağıdaki formül kullanılarak hesaplanır:

$$R_{a,t} = R_{h_t} + \frac{\sum_{i=1}^k (R_{a,h_i} - R_{h_i}) \cdot \text{sim}(\mathbf{h}_t, \mathbf{h}_i)}{\sum_{i=1}^k |\text{sim}(\mathbf{h}_t, \mathbf{h}_i)|}$$

Burada:

- $\mathbf{R}_{a,i}$, hedef otel \mathbf{h} 'nin ortalama puanını,
- $\mathbf{R}_{a,i}$, **i. komşu otel \mathbf{h}_i** 'nin ortalama puanını,
- $\mathbf{R}_{a,i}$, gezgin \mathbf{tra} tarafından **otel \mathbf{h}_i** 'ye verilen puanı ifade eder.

Bu formülün uygulanması sonucunda, **beş farklı seyahat türüne göre oluşturulmuş seyrek gezgin–otel matrislerindeki eksik puanlar**, yukarıdaki **Eşitlik (3)**'ten elde edilen değerlerle doldurulmuştur.

Ancak şu durumun altı çizilmelidir: eğer gezgin \mathbf{tra} 'nın hedef oteli \mathbf{h} 'ye en benzer **k komşu otellerin hiçbirisi** gezgin tarafından puanlanmamışsa (yani $\mathbf{R}_{a,i}$ eksikse), bu durumda $\mathbf{R}_{a,i}$ değeri için herhangi bir tahmin yapılması **mümkün değildir**.

3.5 User-based collaborative filtering for hotel recommendation

Otel puanları tahmin edildikten (impute edildikten) sonra, her gezgin için puan tahmini yapmak amacıyla Kullanıcı Tabanlı İşbirlikçi Filtreleme (User Collaborative Filtering, UCF) yöntemi uygulanmıştır. Bu işlem, gezginler arasındaki kosinüs benzerliğinin (cosine similarity), birlikte puanladıkları otellerin değerlendirmelerine göre hesaplanmasını içerir.

Burada, $\rightarrow tra$ ve $\rightarrow trb$, sırasıyla gezgin tra ve gezgin trb 'nin birlikte puanladıkları otellere ilişkin ortak puanlama vektörlerini temsil etmektedir. q , bu iki gezgin tarafından ortak puanlanan otel sayısını göstermektedir. $\rightarrow tra$ ve $\rightarrow trb$ arasındaki kosinüs benzerliği Eşitlik (4) kullanılarak hesaplanmaktadır:

$$\text{sim}(\text{tr}_a, \text{tr}_b) = \cos(\theta) = \frac{\vec{\text{tr}}_a \cdot \vec{\text{tr}}_b}{\|\vec{\text{tr}}_a\| \|\vec{\text{tr}}_b\|} = \frac{\sum_{i=1}^q \text{tr}_a \cdot \text{tr}_b}{\sqrt{\sum_{i=1}^q \text{tr}_a^2} \sqrt{\sum_{i=1}^q \text{tr}_b^2}} \quad (4)$$

Bu eşitlik yardımıyla, **gezgin-gezgin benzerlik matrisi** oluşturulabilir ve her hedef gezgin için **en yakın k komşu gezgin** belirlenebilir. Ardından, bu komşu gezginlerin puanları, hedef gezgin tra 'nın hedef otel h için vereceği tahmini puanı ($R'_{ra, h}$) hesaplamakta kullanılır. Tahmin, aşağıdaki formül ile gerçekleştirilir:

$$R_{\text{tr}_a, h_t}' = \bar{R}_{\text{tr}_a} + \frac{\sum_{i=1}^k (R_{\text{tr}_i, h_t} - \bar{R}_{\text{tr}_i}) \cdot \text{sim}(\text{tr}_a, \text{tr}_i)}{\sum_{i=1}^k \text{sim}(\text{tr}_a, \text{tr}_i)} \quad (5)$$

Burada:

- R_{ra} : hedef gezgin tra 'nın ortalama otel puanını,
- R_{ri} : **i. komşu gezgin** tr_i 'nin ortalama otel puanını,
- k : komşu gezgin sayısını,
- $R_{ri, h}$: gezgin tr_i tarafından hedef otel h 'ye verilen puanı ifade eder.

Bu yaklaşım, **benzer gezginlerin ortak görüşlerinden yararlanarak**, hedef gezgin için **kişiselleştirilmiş otel puanı tahminleri** yapılmasını sağlamaktadır.

4 Experimental evaluation

4.1 Experimental setup

Önerilen öneri sisteminin etkinliğini değerlendirmek amacıyla, karşılaştırma için temel referans olarak geleneksel bir Kullanıcı Tabanlı İşbirlikçi Filtreleme (User Collaborative

Filtering, UCF) yöntemi kullanılmıştır. Bu yöntem, Tablo 1’de de belirtildiği üzere, birçok öneri sisteminde temel bir teknik olarak yaygın biçimde kabul edilmektedir. Ayrıca, önerilen sistemin performansı, CAPH_UCF yaklaşımı [53] ile karşılaştırılmıştır. Bu karşılaştırma, sistemimizin göreceli etkinliğine ilişkin ayrıntılı bir bakış sunarak, gelişmiş UCF tekniklerini içeren bir yöntemle kıyasla sistemimizin güçlü yönlerini ve potansiyel iyileştirme alanlarını ortaya koymaktadır.

Kapsamlı bir değerlendirme için, veri kümeleri “leave-one-out cross-validation” (birini dışarıda bırakmalı çapraz doğrulama) yöntemiyle ayrı eğitim ve test bölümlerine ayrılmıştır. Bu strateji, sistemin farklı veri senaryoları altındaki performansının sağlam biçimde ölçülmesini sağlamaktadır.

Önerilen sistemin temel amacı, bir gezginin belirli bir otele vereceği puanı doğru bir şekilde tahmin etmektir. Sistem performansını değerlendirmek için iki temel ölçüt kullanılmaktadır: Ortalama Mutlak Hata (Mean Absolute Error, MAE) ve Kök Ortalama Kare Hatası (Root Mean Squared Error, RMSE). Her iki metrik de model tarafından tahmin edilen puanlarla gezginler tarafından verilen gerçek puanlar arasındaki farkı nicel olarak ifade eder.

MAE, tahmin edilen puanlarla gerçek puanlar arasındaki mutlak farkların ortalaması olarak hesaplanır ve modelin ortalama tahmin hatasını yansıtır. Daha düşük MAE değeri, daha yüksek tahmin doğruluğunu gösterir.

Buna karşılık, RMSE, tahmin edilen ve gerçek puanlar arasındaki farkların karelerinin ortalamasının karekökü olarak hesaplanır ve büyük hataların etkisini daha fazla vurgular. MAE’de olduğu gibi, daha düşük bir RMSE değeri de daha yüksek tahmin doğruluğuna işaret eder.

MAE ve RMSE için kullanılan formüller sırasıyla Eşitlik (6) ve Eşitlik (7)’de verilmiştir. Burada:

- T , toplam tahmin sayısını,
- y_i , otele verilen gerçek puanı,
- \hat{y}_i (y şapkalı), model tarafından tahmin edilen puanı ifade etmektedir.

Table 4 The comparative results

Dataset	MAE	RMSE	Coverage (%)	Num. of recommendations
Original	0.6833	0.8649	89.96	12,246
Friends	NA	NA	0	0
Solo	NA	NA	0	0
Family	NA	NA	0	0
Couple	0.6234	0.7642	13.56	542
Business	0.6919	0.8728	72.34	4384

$$MAE = \frac{1}{T} \sum_{i=1}^T |y_i - \hat{y}_i| \quad (6)$$

$$RMSE = \sqrt{\frac{1}{T} \sum_{i=1}^T (y_i - \hat{y}_i)^2} \quad (7)$$

4.2 Results

4.2.1 Study I

Çalışma I (Study I), tüm orijinal veri kümesinin **UCF (Kullanıcı Tabanlı İşbirlikçi Filtreleme)** yaklaşımıyla kullanılması ile orijinal veri kümesinin bağlamsal bilgilere dayanarak beş alt kümeye bölünmesi durumları arasındaki **performans farklarını değerlendirmeye** odaklanmaktadır. **Bölüm 1’de** vurgulandığı üzere, bağlamsal bilgilerin eklenmesi **seyreklik (sparsity)** sorununu daha da kötüleştirmekte; bu da beşe ayrılmış alt kümelere dayalı tahminlerin, bölünmemiş orijinal veri kümesine dayalı tahminlere kıyasla daha zor hale gelmesine neden olmaktadır.

Tablo 4, en yakın **k komşu sayısının 30** olarak sabitlendiği durumda elde edilen sonuçları göstermektedir. Gözlemler, **arkadaş, tekil (solo) ve aile** alt kümelerinin herhangi bir öneri üretilmediğini ortaya koymaktadır. Buna karşılık, **çift ve işalt** kümeleri öneri üretebilmekte, ancak bu alt kümelerin **kapsama oranları (coverage rates)**, orijinal veri kümesinin tamamının kullanıldığı duruma kıyasla daha düşük kalmaktadır.

Dolayısıyla, orijinal veri kümesinin bağlamsal bilgilere göre daha küçük parçalara ayrılmasının, **UCF tabanlı öneri sistemleri için tahmin sürecini zorlaştırdığı** sonucuna varılmaktadır. Bu zorluk, alt kümelerde **komşuları etkili biçimde tanımlayacak yeterli verinin bulunmamasından**, diğer bir ifadeyle, bütünleşik veri kümesinin sunduğu bilgi yoğunluğunun alt kümelerde kaybolmasından kaynaklanmaktadır.

4.2.2 Study II

Çalışma II (Study II), **UCF, CAPH_UCF** ve önerdiğimiz yeni yaklaşımı kullanan öneri sistemlerinin performanslarını karşılaştırmayı amaçlamaktadır. Adil bir karşılaştırma sağlamak ve farklı seyahat türleri arasında dengesiz tahminleri azaltmak için, hem **CAPH_UCF** hem de önerdiğimiz yöntem için **beş gezgin–otel matrisinin yoğunlukları (densities)** birbirine daha yakın olacak şekilde ayarlanmıştır. Ayrıca, önerilen sistemimizi daha zorlu koşullar altında test etmek amacıyla, matrislerimiz **CAPH_UCF**’inkinden daha seyrek olacak biçimde tasarlanmıştır. Bu doğrultuda, **komşu tahmini boyutları (neighbor imputation sizes)** CAPH_UCF için **50**, önerilen yöntemimiz için ise **100** olarak belirlenmiştir. Her iki yaklaşım için oluşturulan beş matrisin sonuçta elde edilen yoğunluk değerleri **Tablo 5**’te sunulmuştur.

Tablo 6 ve Şekil 2, **UCF, CAPH_UCF** ve önerilen yöntemimizin **MAE (Ortalama Mutlak Hata) ve RMSE (Kök Ortalama Kare Hatası)** sonuçlarını göstermekte; **Tablo 7** ise **kapsama oranları (coverage rates) ve öneri sayıları** hakkında ayrıntılı bilgi sunmaktadır. Analizimizde, **en yakın komşu sayısı (top-k)** 10, 20 ve 30 olarak belirlenen farklı durumlar karşılaştırılmıştır.

Tablo 6 ve Şekil 2’ye göre, önerilen yöntemimiz, **k** değerindeki değişimden bağımsız olarak tüm alt kümelerde **CAPH_UCF**’yi açık ara geride bırakmış; hem CAPH_UCF hem de UCF yöntemlerine kıyasla anlamlı ölçüde **daha düşük MAE ve RMSE** değerleri elde etmiştir.

Buna karşın, **UCF** yöntemi veri seyrekliği sorununu çözmede zorlanmakta; bu durum, **Tablo 7’de** belirtildiği üzere **%100 kapsama oranına ulaşamamasıyla** açıkça görülmektedir.

Ayrıca, **top-k komşu sayısı arttıkça**, UCF yönteminde **kapsama oranının kademeli olarak azaldığı** gözlemlenmiştir. Bu da, söz konusu yöntemin **seyrek veriyle etkin biçimde başa çıkma konusundaki sınırlılıklarını** vurgulamaktadır.

Table 5 Density comparison of the five traveler–hotel matrices: CAPH_UCF versus our approach

Dataset	CAPH_UCF (%)	Our approach (%)
Friends	39.40	53.64
Solo	34.10	55.82
Family	32.55	45.99
Couple	33.66	39.60
Business	39.30	41.16

4.2.3 Study III

Önerilen yöntemimiz, iki temel yaklaşıma kıyasla üstün performans sergilemiştir. **Çalışma III (Study III)**, tahmin edilen komşu boyutlarının (**imputed neighbor sizes**) değişiminin, önerilen yöntemimizin tahmin doğruluğu üzerindeki etkisini incelemektedir. Bu amaçla, α ile gösterilen tahmini komşu boyutu sırasıyla **50, 100, 150 ve 200** olarak ayarlanmış; ayrıca **en yakın komşu sayısı (top-k)** değerleri **10, 20 ve 30** olarak belirlenmiştir.

Tablo 8, α değerlerinin artmasıyla birlikte **matris yoğunluğunun (matrix density)** da arttığını ve daha yoğun gezgin–otel matrislerinin **daha yüksek performans** sağladığını göstermektedir.

Tablo 9, farklı tahmini komşu otel boyutları (α) ve komşu gezgin sayıları (k) için elde edilen **MAE (Ortalama Mutlak Hata)** ve **RMSE (Kök Ortalama Kare Hatası)** metriklerini sunmaktadır. Dikkat çekici bir şekilde, α değeri arttıkça hem MAE hem de RMSE değerlerinin — veri kümesinin hangi alt kümesine ait olduğuna bakılmaksızın — **azaldığı** gözlemlenmiştir. En iyi sonuçlar, $\alpha = 200$ olarak ayarlandığında elde edilmiştir.

Bu bulgu, **orijinal veri kümesinin bağlamsal bilgilere göre bölünmesinin seyreklik sorununu şiddetlendirdiğini** ve bu sorunu azaltmak için **ek veriye olan gereksinimi** açıkça ortaya koymaktadır. Çalışma sonuçlarımız, **gezgin yorumlarından elde edilen gizil duygu verilerinin (latent sentiment data)** öneri sistemlerinde kullanılmasının, **seyreklik problemini etkin bir şekilde hafiflettiğini** doğrulamaktadır.

Ayrıca, bu çalışma göstermektedir ki; gezgin–otel matrislerinin **daha kapsamlı bilgiyle zenginleştirilmesi** ve buna bağlı olarak **yoğunluğunun artırılması**, öneri sisteminin performansını anlamlı ölçüde iyileştirmektedir.

Table 6 Comparative results of MAE and RMSE metrics: evaluating UCF, CAPH_UCF, and our approach

Dataset	k	MAE			RMSE		
		UCF	CAPH_UCF	Our approach	UCF	CAPH_UCF	Our approach
Friends	10	NA	0.5816	0.19	NA	0.7738	0.3005
Solo		0.657	0.6389	0.185	0.7974	0.8207	0.315
Family		0.5935	0.6114	0.192	0.7386	0.7755	0.3142
Couple		0.6591	0.6336	0.264	0.833	0.8139	0.4015
Business		0.7107	0.6862	0.3567	0.8976	0.8704	0.5229
Friends	20	NA	0.5816	0.1819	NA	0.7738	0.2844
Solo		NA	0.6389	0.1862	NA	0.8207	0.3078
Family		NA	0.6114	0.191	NA	0.7755	0.3091
Couple		0.6239	0.6337	0.2571	0.7761	0.814	0.3936
Business		0.7062	0.6862	0.3504	0.8909	0.8704	0.5179
Friends	30	NA	0.5816	0.1999	NA	0.7738	0.2974
Solo		NA	0.6389	0.2102	NA	0.8207	0.3247
Family		NA	0.6114	0.2084	NA	0.7755	0.322
Couple		0.6234	0.6337	0.2701	0.7642	0.814	0.4024
Business		0.6919	0.6862	0.3547	0.8728	0.8704	0.5198

5 Conclusion

Etkili otel öneri sistemlerinin geliştirilmesi, turizm sektöründe giderek daha hayati bir önem kazanmıştır. Bu sistemler, gezginler için çevrimiçi otel bilgilerinin oluşturduğu bilgi karmaşasını azaltmak amacıyla **CBF (İçerik Tabanlı Filtreleme)**, **UCF (Kullanıcı Tabanlı İşbirlikçi Filtreleme)** ve **ICF (Öğe Tabanlı İşbirlikçi Filtreleme)** gibi yerleşik yöntemleri kullanmaktadır. Ayrıca, **CARS (Bağlam Farkındalıklı Öneri Sistemleri)** yaklaşımının dahil edilmesi, seyahat türleri gibi bağlamsal otel bilgilerini karar verme sürecine entegre ederek öneri kalitesini daha da artırabilmektedir.

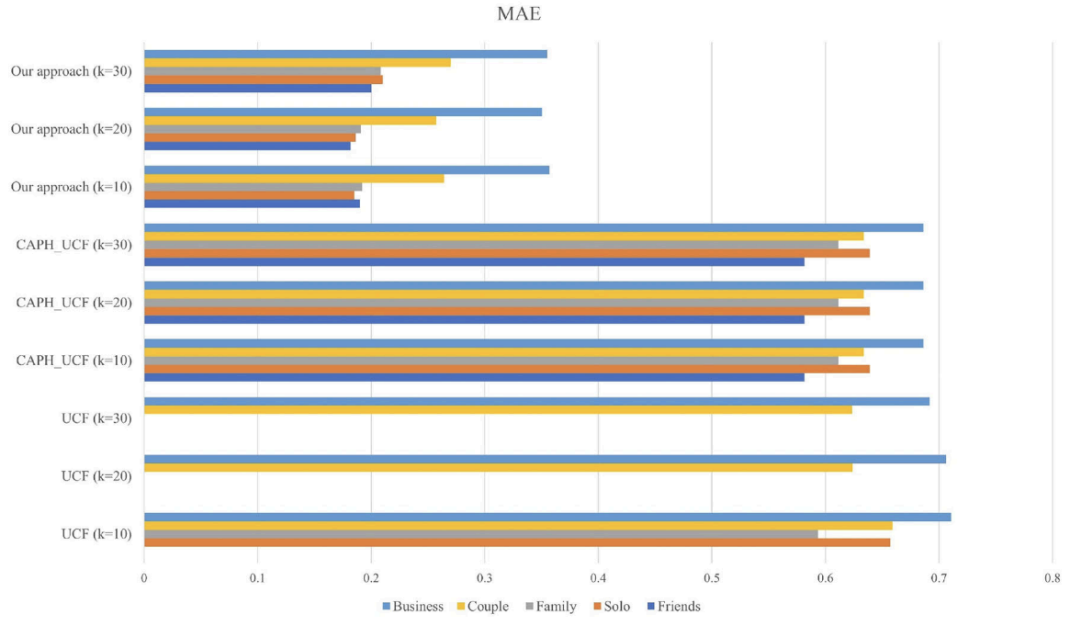
Bu ilerlemelere rağmen, literatür incelendiğinde, **gezgin–otel matrisindeki veri seyreklik (data sparsity)** sorununu ele alan çalışmaların oldukça sınırlı olduğu görülmektedir. Bu matris, hedef kullanıcıya benzer gezginlerin belirlenmesi ve onların otel puanlarının tahmin edilmesi açısından kritik bir rol oynamaktadır. Ancak, gezginlerin genellikle her oteli değerlendirmemesi nedeniyle, bu matris çoğu zaman çok sayıda **eksik puan (missing rating)** içermekte ve kapsamlı bilgi barındırmamaktadır. Bu seyrek matrislere dayanmak, sıklıkla **optimal olmayan önerilere** yol açmaktadır.

Bu çalışma, **CBF, UCF ve ICF** yöntemlerinin güçlü yönlerini bir araya getiren ve **seyahat türlerini (İş, Aile, Çift, Arkadaş, Tekil)** bağlamsal bilgi olarak dahil eden **yenilikçi bir hibrit otel öneri sistemi** önermektedir. Ayrıca, gezgin yorumlarından duygu verisi elde etmek amacıyla **Duygu Analizi (Sentiment Analysis, SA)** yönteminden yararlanılmıştır. Elde edilen bu duygu verisi, gezgin tercihlerini daha iyi anlamayı sağlamakta ve seyrek matrislerdeki eksik puanların doldurulmasında kullanılmaktadır. Bu sayede, öneri doğruluğunun artırılması hedeflenmiştir.

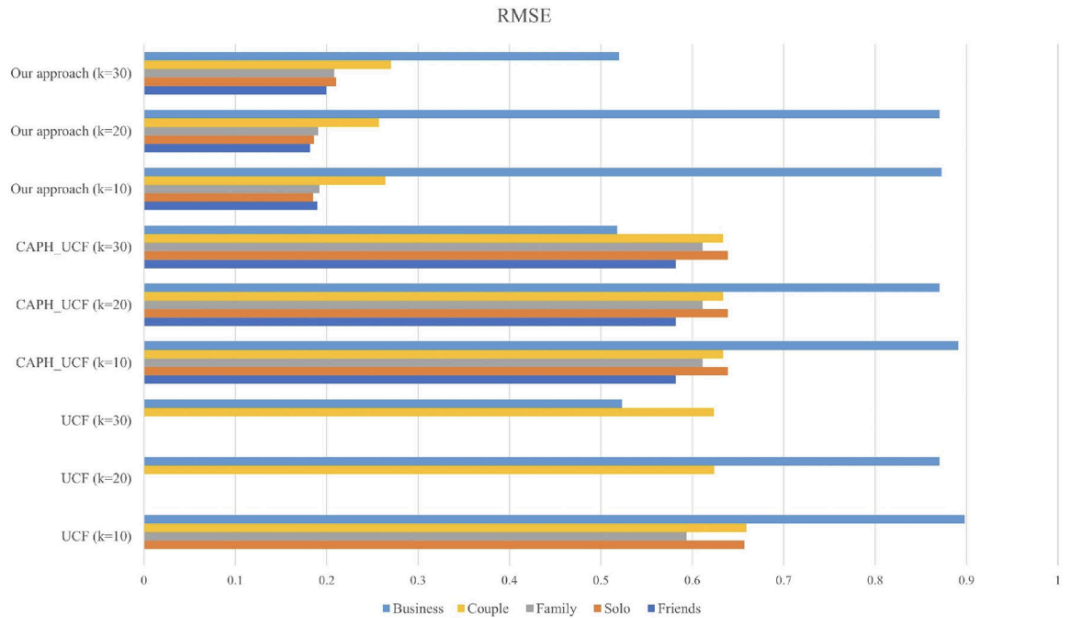
Deneyisel analizlerimiz, **TripAdvisor.com**'dan elde edilen bir veri kümesi üzerinde gerçekleştirilmiştir. Bulgular, bağlamsal bilgilerin eklenmesinin matrislerin seyreklik düzeyini artırdığını ve yüksek performansa ulaşmayı zorlaştırdığını göstermektedir. Özellikle, orijinal seyrek matrisin seyahat türlerine göre beş alt matrise bölünmesi bu sorunu daha da belirgin hale getirmiştir. Ancak, önerdiğimiz **eksik puan tamamlama (missing rating)**

imputation) yöntemi, geleneksel **UCF yaklaşımlarının** ve **CBF, UCF, ICF ile bağlamsal bilgiyi** birleştiren ancak **duygu analizini içermeyen yöntemlerin** performansını anlamlı biçimde aşmıştır.

Özellikle, sistemimiz **MAE (Ortalama Mutlak Hata)** ve **RMSE (Kök Ortalama Kare Hatası)** değerlerinde belirgin bir düşüş sağlayarak, **üstün tahmin doğruluğunu** açık bir şekilde ortaya koymuştur.



(a) MAE



(b) RMSE

Fig. 2 Bar chart illustrating the comparative results of MAE and RMSE metrics

Table 7 Comparative results of coverage rate and number of recommendations: evaluating UCF, CAPH_UCF, and our approach

Dataset	k	Coverage			Num. of recommendations		
		UCF (%)	CAPH_UCF (%)	Our approach (%)	UCF	CAPH_UCF	Our approach
Friends	10	0	100	100	0	716	716
		14.76	100	100	206	1396	1396
		24.05	100	100	404	1680	1680
		64.34	100	100	2571	3996	3996
		92.97	100	100	5634	6060	6060
Friends	20	0	100	100	0	716	716
		0	100	100	0	1396	1396
		0	100	100	0	1680	1680
		30.76	100	100	1229	3996	3996
		85.18	100	100	5162	6060	6060
Friends	30	0	100	100	0	716	716
		0	100	100	0	1396	1396
		0	100	100	0	1680	1680
		13.56	100	100	542	3996	3996
		72.34	100	100	4384	6060	6060

Table 8 The densities of the five matrices by different imputed neighbor sizes

Dataset	$\alpha = 50$ (%)	$\alpha = 100$ (%)	$\alpha = 150$ (%)	$\alpha = 200$ (%)
Friends	24.48	39.40	50.37	59.12
Solo	22.16	34.10	42.71	49.32
Family	20.25	32.55	41.58	48.72
Couple	20.95	33.66	42.75	49.75
Business	25.44	39.30	48.32	54.73

Table 9 MAE and RMSE results based on different imputed neighbor hotel sizes (α) and number of neighbor travelers (k)

Dataset	k	MAE				RMSE			
		$\alpha = 50$	$\alpha = 100$	$\alpha = 150$	$\alpha = 200$	$\alpha = 50$	$\alpha = 100$	$\alpha = 150$	$\alpha = 200$
Friends	10	0.2915	0.1900	0.1531	0.1354	0.4140	0.3005	0.2597	0.2392
	20	0.3136	0.1819	0.1263	0.1013	0.4357	0.2844	0.2356	0.2165
	30	0.3411	0.1999	0.1311	0.0966	0.4653	0.2974	0.2359	0.2122
Solo	10	0.2773	0.1850	0.1572	0.1473	0.4145	0.3150	0.2889	0.2809
	20	0.3010	0.1862	0.1385	0.1182	0.4200	0.3078	0.2747	0.2618
	30	0.3374	0.2102	0.1487	0.1177	0.4554	0.3247	0.2764	0.2571
Family	10	0.2878	0.1920	0.1608	0.1500	0.4141	0.3142	0.2859	0.2736
	20	0.3050	0.1910	0.1436	0.1234	0.4289	0.3091	0.2695	0.2551
	30	0.3280	0.2084	0.1531	0.1237	0.4540	0.3220	0.2738	0.2541
Couple	10	0.3650	0.2640	0.2261	0.2099	0.5020	0.4015	0.3719	0.3606
	20	0.3704	0.2571	0.2099	0.1866	0.5018	0.3936	0.3614	0.3488
	30	0.3853	0.2701	0.2143	0.1864	0.5159	0.4024	0.3628	0.3485
Business	10	0.4382	0.3567	0.3298	0.3187	0.5897	0.5229	0.5077	0.5025
	20	0.4395	0.3504	0.3149	0.2992	0.5866	0.5179	0.5007	0.4957
	30	0.4510	0.3547	0.3143	0.2963	0.5962	0.5198	0.5003	0.4946