

Metin Uslu

Student Id: 235B7014

Movie Recommendation Systems Using Actor-Based Matrix Computations in South Korea

24.12.2023

Sunum Planı

1. Özet & Giriş
2. Çalışma Hakkında
 1. İçerik Tabanlı Tavsiye Sistemleri
 2. İçerik Tabanlı Tavsiye Sistemleri ile İlgili Sorunlar
3. Yöntem
 1. Veri Toplama ve Ön İşlemler
 2. Film ve Film Türü arasında Rank Korelasyonun Hesaplanması
 3. Aktör ve Film Türü arasında Korelasyonun Hesaplanması
4. Sonuçlar
5. Değerlendirme & Tartışma

Makale Hakkında

Article Name: Movie Recommendation Systems Using Actor-Based Matrix Computations in South Korea

Page: 7 & Citation: 6

Article: <https://ieeexplore.ieee.org/document/9566476>

Publisher: IEEE

Date of Publication: 11 October 2021

Authors:

Syjung Hwang: <https://ieeexplore.ieee.org/author/37089549318>

Eunil Park: <https://ieeexplore.ieee.org/author/37891494800>

Özet

Öneri Sistemleri (Recommendation Systems) e-ticaret platformlarında özellikle kitap, müzik, film gibi kategorilerinde son kullanıcıların arama maliyetleri/zahmetleri ortadan kaldırmak ve kullanıcılara daha uygun içerikleri sunarak birlikteliklerini ortaya koymak için kullanılmaktadır. Literatürde İçeriğe Dayalı (Content-Based), İş Birliğine Dayalı (Collaborative Filtering) ve Bilgiye Dayalı (Knowledge-Based) teknikler kullanılmak ile birlikte farklı alternatifler bulunmaktadır. Güney Kore film endüstrisinde yaygın olarak kullanılan Film Türleri (Movie Genre) temel alınarak İş Birlikçi Filtreleme yöntemleri ile film önerileri yapılmaktadır.

Web platformlarda kullanıcıların ilk kez film tavsiyeleri aldıkları anda (cold start) ya da oyunculara ve/veya yönetmenlere ilişkin belirli tercihler sahip olduklarında bu yöntemin etkisiz ve zayıf kaldığı söylenebilir. Bu sebeple oyuncuların film bilgilerini içeren ve 509 adet Güney Kore Film Türlerini dikkate alan İçerik Bazlı(Content Based) filtrelemeyi kullanan, Aktör (oyuncu) bazlı bir öneri sistemi önermeyi motive ettiğini belirtiyorlar. Önerdikleri sistemin etkinliği, performansı ve geleneksel yöntemler ile karşılaştırıldığı belirtiliyor.

Giriş

İşbirlikçi filtreleme teknikleri, benzer özelliklere sahip kullanıcıların benzer tercihlere sahip olma eğiliminde olduğu gerçeğine odaklanmaktadır [6]. İşbirliğine dayalı filtreleme teknikleriyle karşılaştırıldığında, içerik tabanlı filtreleme teknikleri, bir sonraki seçimlerini tahmin etmek için kullanıcılar tarafından gerçekleştirilen önceki değerlendirmeleri dikkate alır [7]. Bu iki temsili öneri tekniği, birçok şirket tarafından kullanıcıya özel hizmetlerde kullanılmıştır.

Öneri sistemleri, kitapçılar ve video mağazaları da dahil olmak üzere çoğu çevrimiçi hizmet için vazgeçilmez hale geldi. Bu tür hizmetler arasında Netflix ve YouTube gibi çevrimiçi film yayın platformları, kendi platformları üzerinden kullanıcıya özel video içerikleri sunmaktadır. Bu platformlardaki kullanıcılardan toplanan büyük miktardaki verilere dayanarak, gelişmiş işbirlikçi filtrelemeye dayalı teknikler kullanılmaktadır. Her iki hizmet de piyasada son derece başarılı olmasına rağmen, bazı önemli sınırlamalar bu tekniklerin bu platformlar tarafından kullanılmasını engellemiştir. Yeterli sayıda kullanıcısı olmayan platformlarda içerik bazlı filtreleme teknikleri de dahil olmak üzere diğer teknikler kullanılmalı; ikinci teknik, belirli kullanıcılar hizmetlerle ilk karşılaştığında da kullanılmalıdır. Bu, bu hizmetlerin başlangıç aşamasında işbirlikçi filtreleme tekniklerinin kullanımının sorunlu olabileceği anlamına gelir.

Yukarıda belirtilen zorluklara dayanarak, film öneri sistemleri üzerinde çalışan birçok bilim insanı, film türleri gibi belirli kategorize edilmiş standartlara sahip işbirlikçi veya hibrit filtreleme öneri sistemlerini kullanmaya çalıştı. Örneğin Netflix ayrıca öneri sistemini, bireysel kullanıcıların her türe olan eğilimini yansıtarak çalıştırır. Bu yaklaşım, film türlerini göz önünde bulundurarak belirli bir filmi seçmeyi amaçlayan kullanıcılar için yararlı olsa da, oyunculara veya yönetmenlere ilişkin tercihler gibi filmlere ilişkin belirli tercihleri olan kullanıcılar söz konusu olduğunda yaklaşım etkisiz olacaktır.

Bu çalışma, oyuncuların filmografi bilgilerini ve 509 Güney Kore filminin türlerini dikkate alan oyuncu bazlı öneri sistemi şeklinde içerik bazlı bir öneri sistemi önermektedir. Daha sonra önerilen sistemin etkinliği ve performansı, geleneksel tür odaklı öneri sistemiyle karşılaştırılarak değerlendirilir.

Çalışma Hakkında

1. İçerik Tabanlı Tavsiye Sistemleri
2. İçeriğe Dayalı Öneri Sistemiyle İlgili Sorunlar

İçerik Tabanlı Tavsiye Sistemleri

Temel olarak bilişsel filtrelemeye dayanan içerik tabanlı öneri sistemi, “alternatif öğeleri kullanıcı profilleriyle ilişkili öğelerle karşılaştırarak öğe önerilerini görüntüleyen bir sistemdir”. Bu sistem esas olarak belirli bir kullanıcı tarafından önceden seçilen belirli öğelerin özelliklerini analiz eder ve benzer kategorilere ait diğer öğeleri önermek için bu öğelerin benzersiz özelliklerini çıkarır. Örneğin, geçmişte çok sayıda romantik film izlemiş olan belirli bir kullanıcı, henüz izlemediği diğer romantik filmler için öneriler alacaktır. Bu tür sistemler halihazırda arama motorları, kitapçılar ve alışveriş merkezleri tarafından kullanılmaktadır.

Genel olarak içerik tabanlı bir öneri sistemi, belirli öğeleri önerirken kullanıcılar arasındaki benzerlikleri dikkate almaz. Bunun yerine sistem tamamen bireysel bir kullanıcı tarafından kullanılan içeriklerin özelliklerine odaklanıyor. Dolayısıyla bireysel kullanıcı açısından bu sistemin, yalnızca bireysel kullanıcı profiline ilişkin içeriğin benzersiz özelliklerini dikkate alması nedeniyle diğer öneri sistemlerine göre daha güvenilir olduğu düşünülmektedir.

İçerik Tabanlı Tavsiye Sistemleri ile İlgili Sorunlar

İçerik tabanlı öneri sistemleri yaygın olarak kullanılmasına rağmen, bazı önemli sınırlamalara sahiptir.

- ❖ Birincisi, sistem içerik özelliklerinin doğası gereği sınırlı olabilir. İyi tanımlanmış öneriler sağlamak için sistem yeterli sayıda özelliğe ihtiyaç duyar. Ancak videolardan veya görsellerden özellik çıkarmak zordur; dolayısıyla bu tür veriler içerik tabanlı öneri sistemi için çok sayıda özellik sağlayamamaktadır.
- ❖ İkincisi, öneriler yapılırken diğer kullanıcıların değerlendirmeleri dikkate alınmadığı için önerilen içeriklerin kalitesini garanti etmek zordur. Bu, sistem öğeleri önerme açısından kayda değer kısıtlamalara sahip olduğunda (örneğin, yalnızca yüksek puan alan öğeleri önermek) sınırlı sayıda öğenin önerilebileceği anlamına gelir. Bu, belirli öğeleri önerirken ve tavsiye ederken dikkate değer kutuplaşma sorunlarına yol açabilir.

Bu sınırlamalar nedeniyle içerik tabanlı öneri sistemlerinin filmleri etkili bir şekilde önermesi her zaman mümkün olmayabilir. Bu eksikliği gidermek için oyuncuların, özellikle de film yıldızı olanların, filmlerin gişe başarısı üzerinde oldukça büyük etkisi vardır. Başka bir deyişle, insanlar filmleri en sevdikleri oyunculara göre seçme eğilimindedir; bu nedenle, kullanıcıların en sevdiği aktörleri dikkate alan bir öneri sisteminin geliştirilmesi, bu türün aktörlere dayalı bir korelasyon göstergesini gerektirecektir. Bu nedenle çalışmamız, Güney Kore film endüstrisindeki sinema oyuncularını dikkate alan içerik tabanlı bir öneri sistemi geliştirmeyi amaçlamaktadır.

Yöntem

1. Veri Seti Oluşturma ve Ön İşlemler
2. Film ve Film Türü arasında Rank Korelasyonun Hesaplanması
3. Aktör ve Film Türü arasında Korelasyonun Hesaplanması

Veri Toplama ve Ön İşlemler

Veri seti 3 ayrı sistem üzerinden toplandığı belirtiliyor.

1. Kore Film Konseyi

Bu veri seti 3429 adet film bilgisini içermekte ve içerisinde; filmlerin listesi, film ayrıntıları, aktörler ve yönetmen bilgileri yer alıyor. Ayrıca aktörlerin ve yönetmenlerin filmografileri de yer almaktadır. Veri kümesindeki filmler arasında uzun metrajlı filmler, kısa öyküler ve çok amaçlı kitaplar yer alıyordu. Yalnızca vizyona giren uzun metrajlı filmleri kullanmaya karar verdik.

2. Naver Film Sitesi

Bu platform üzerinden, Film Başlığı, Yılı, İzleyici Puanı ve İzleyici Sayısı, İnternet Puanı ve Gazeteci & Eleştirmenlerin Derecelendirmeleri yer almaktadır. Derecelendirmelerin güvenilirliğini sağlamak için yalnızca 30'dan fazla kişi tarafından derecelendirilen filmleri seçtik. Her filmin derecelendirmesi izleyicilerin derecelendirmesine dayanıyordu, ancak bu olmadığında bunun yerine İnternet kullanıcılarının derecelendirmeleri kullanıldı. Bu iki derecelendirme türü (0,82) ve benzer dağılım modelleri arasındaki yüksek korelasyon nedeniyle bu değişikliğin uygun olduğu düşünülmektedir.

3. Kore Gişe Bilgi Sistemi

Kore Gişe Bilgi Sistemi üzerinden film adı, gösterim tarihi, koltuk satış oranı, koltuk doluluk oranı, koltuk sayısı, satışlar, kümülatif satışlar, seyirci sayısı ve kümülatif izleyici sayısı yer alıyor. Koleksiyonda dünya çapındaki filmlerin yer aldığı 2346 yönetmen, 15539 oyuncu ve 6647 filme ilişkin veriler yer aldı. Eski filmlere ilişkin veriler yeni filmlere göre eksik olduğundan yalnızca 2010-2019 yıllarına ait film verilerini kullandık. Ayrıca ortalama izleyici kitlesi yüzde 25'in altında olan filmlerin oyuncularını ve yönetmenlerini hariç tutuldu. Verilerin ön işlenmesinden sonra veri seti 4450 Güney Koreli aktör ve 509 Güney Kore filmini içeriyordu

Veri Toplama Ön İşlemler (Devam)

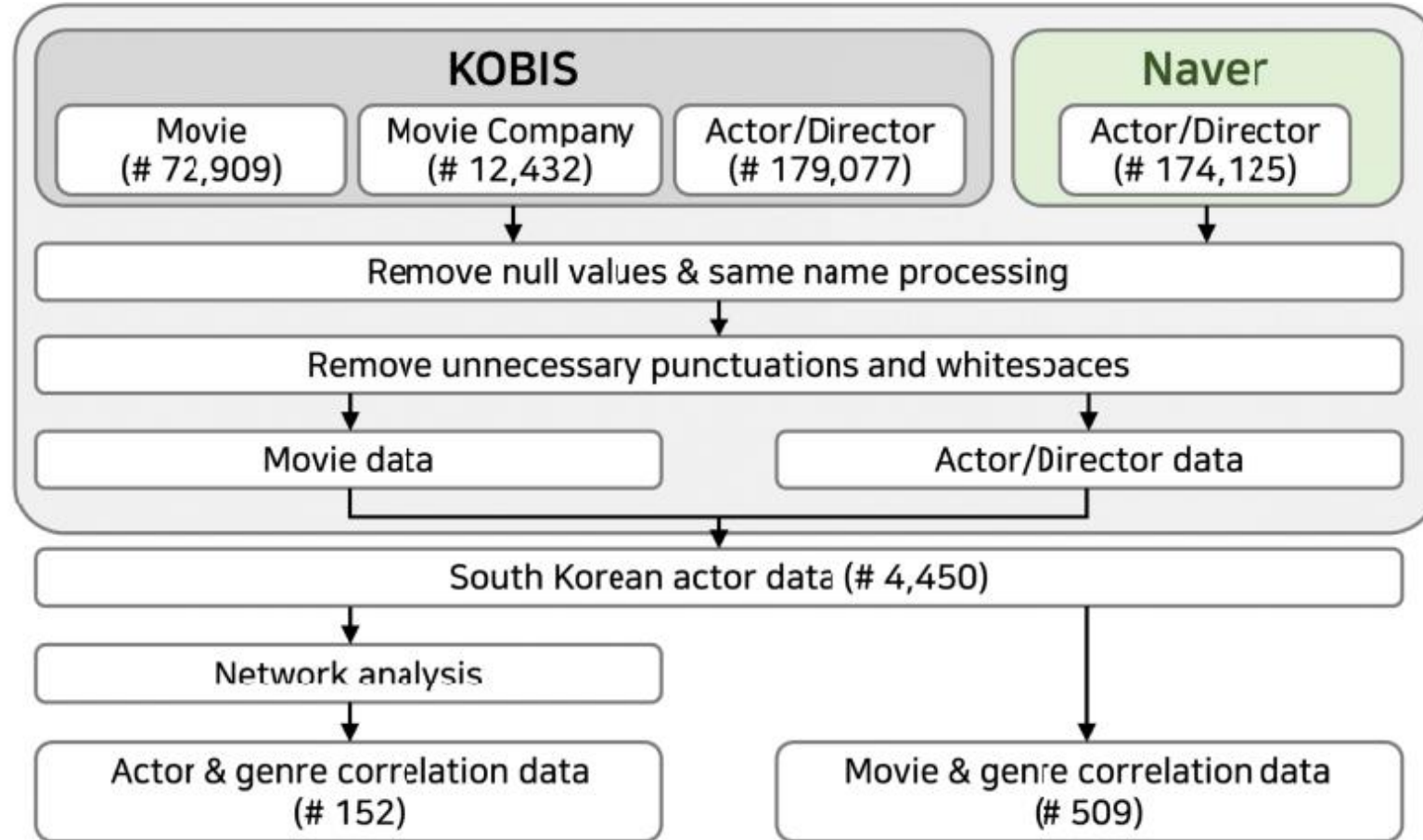


Fig. 1. Data preprocessing procedure for correlation analysis.

Veri Toplama Ön İşlemler (Devam)

TABLE I
GENRE CATEGORY

No.	Genre	No.	Genre
1	Comedy	9	Mystery
2	Fantasy	10	Melodrama
3	War	11	Drama
4	Adventure	12	Horror
5	Action	13	Family
6	Thriller	14	Documentary
7	Historical drama	15	SF
8	Crime	16	miscellaneous

Film(Movie) ve Film Türü(Genre) arasında Rank Korelasyonun Hesaplanması

- Belirli bir film ile türü arasındaki sıra korelasyonunun hesaplanması, film veritabanındaki türlerin kombinasyonuna dayanır. Her filmin birden fazla türü kapsayan bir tür birleşimi vardır. Film türleri film yapımcıları tarafından seçilir. Örneğin 26 Yıl filmi “aksiyon” ve “drama” türlerine ait. Bu, filmin bu iki türe göre kategorize edildiği anlamına gelir.
- Bir film ile türü arasındaki sıra korelasyonunu belirlemek için önce belirli bir filmi seçtik ve film türleri arasına bir tane daha ekledik. Ancak her türe ait filmlerin sayısı açısından önemli bir tutarsızlık mevcuttur. Örneğin “komedi” türünde toplam film sayısı 126, “fantezi” türünde film sayısı ise 12. Böylece sadece bir düzeltme tahmini yapmak yerine her türün her türdeki sıralamasını hesapladık. (Önyargılı dağılımları hariç tutmak için).

Film(Movie) ve Film Türü(Genre) arasında Rank Korelasyonun Hesaplanması

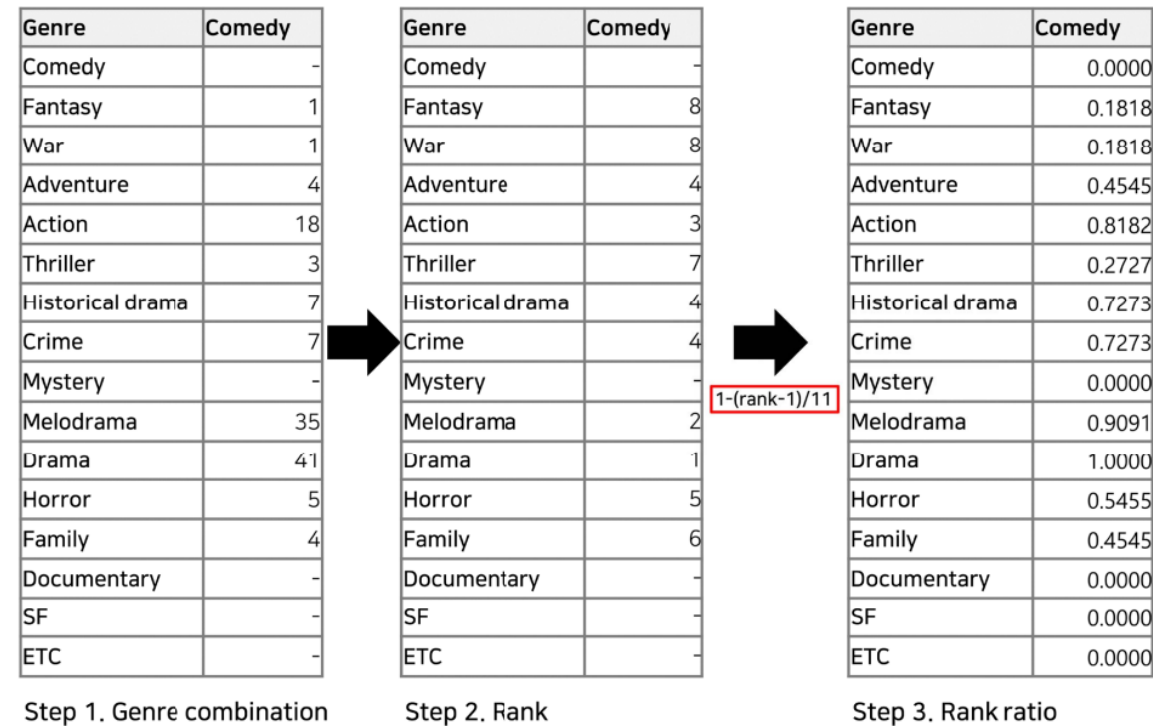


Fig. 3. Course that calculates rank correlations.

Aktör ve Film Türü arasında Korelasyonun Hesaplanması

- Belirli bir oyuncu ile film türü arasındaki korelasyon, oyuncu veri tabanındaki tür kombinasyonu kullanılarak hesaplanır. Her oyuncunun belirli türlerdeki filmlerde diğerlerinden daha sık rol alma eğiliminde olması nedeniyle, oyuncunun sıklıkla rol aldığı belirli türlerle güçlü bir bağlantısı olduğu sonucuna varılabilir. Örneğin, 2010'dan 2019'a kadar aktör “Don Lee” drama türünde 13 filmde rol aldı; ayrıca 10 aksiyon ve 9 polisiye filmde rol aldı. Bu onun bu üç türle de ilişkilendirilebileceği anlamına gelir. (Tablo III).
- Her türe ait filmlerde rol alan oyuncuların sayısı aynı değildir. Bu nedenle, bu önyargıyı azaltmak için her türde yüksek derecede merkeziliğe sahip ilk on aktörü seçmek için bir ağ analizi kullanarak belirli bir aktörü türleriyle ilişkilendirdik. Aktörlerden (satırlar) ve türlerden (sütunlardan) oluşan bir tablo oluşturduk ve **Pearson** korelasyon katsayısını kullanarak belirli bir aktör ile türü arasındaki korelasyonu hesapladık.

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Sonuçlar

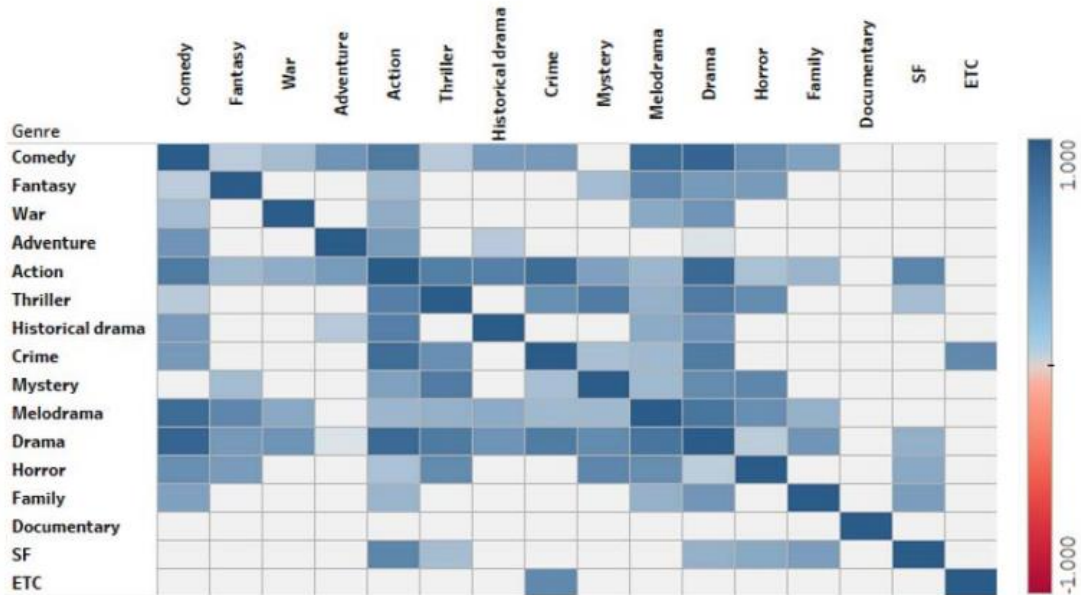


Fig. 4. Rank correlation plot between a movie and its genres.

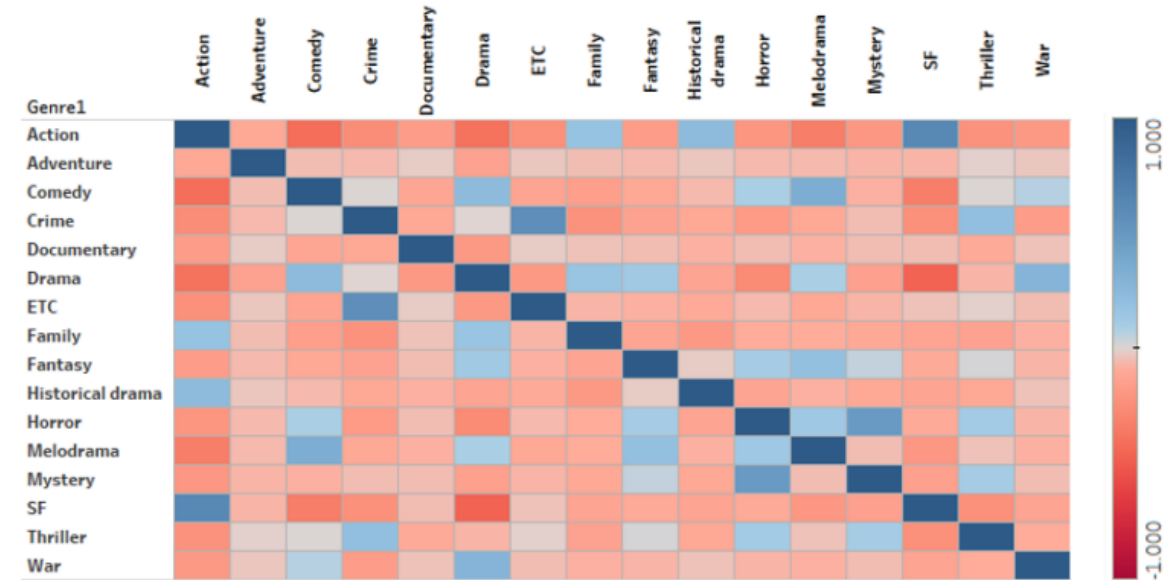


Fig. 5. Correlation plot between an actor and their genres.

Sonuçlar (Devam)

TABLE VII

RESULTS OF MOVIE RECOMMENDATIONS FOR PARASITE; “MOVIE–
GENRE” CORRESPONDS TO THE CASE WHERE THE USER’S MOVIE
SELECTION CRITERION IS THE GENRE, WHILE “ACTOR–GENRE”
CORRESPONDS TO THE CASE WHERE THE USER’S MOVIE
SELECTION CRITERION IS THE ACTOR

Movie - Genre		Actor - Genre	
No.	Title	No.	Title
1	<i>A Taxi Driver</i>	1	<i>A Taxi Driver</i>
2	<i>The Attorney</i>	2	<i>The Attorney</i>
3	<i>Ode to My Father</i>	3	<i>Ode to My Father</i>
4	<i>Blind</i>	4	<i>Blind</i>
5	<i>Secret Reunion</i>	5	<i>Secret Reunion</i>
6	<i>The Neighbors</i>	6	<i>The Neighbors</i>
7	<i>The Age of Shadows</i>	7	<i>i Can Speak</i>
8	<i>Snow piercer</i>	8	<i>1987</i>
9	<i>i Can Speak</i>	9	<i>Default</i>
10	<i>1987</i>	10	<i>All About My Wife</i>

Sonuçlar (Devam)

TABLE VIII
COMPARISON OF TEST RESULTS

User	Preference	Genre-based	Actor-based
1	genre	8	7.5
2	genre	8	8
3	actor	5	6
4	actor	8.5	9.5
5	actor	8	9
6	genre	8	7
7	genre	7	6
8	actor	5	8
9	actor	7.5	9
10	genre	9	9
11	actor	5	6
12	actor	8	9
13	actor	6	6.5
14	actor	7	8
15	genre	8	7
16	genre	9	7
17	actor	6	9
18	actor	7	8.5
19	actor	7.5	8
20	genre	8	7.5
Average	Genre: 8, Actor: 12	7.28 (1.23)	7.78 (1.10)

Değerlendirme & Tartışma

Bu çalışma, Film Tür ve Aktörleri dikkate alarak belirli filmleri önermek için içerik tabanlı filtrelemeyi kullanmayı amaçlamıştır. Bu yaklaşımın etkinliğini iki içerik bazlı öneri sistemi oluşturarak değerlendirdik: Film ve Türler arasındaki korelasyona dayalı bir sistem ve Aktör ve Türler arasındaki korelasyona dayalı bir sistem.

Çeşitli türler arasındaki korelasyonu oyunculara göre hesapladık ve içerik bazlı bir öneri sistemi önerdik. Çalışmamızın sonuçları, oyuncuların film öneri sisteminin önemli bir bileşeni olarak dikkate alınmasının, kullanıcılara daha uygun filmlerin önerilmesine yardımcı olduğunu göstermektedir.

Değerlendirme & Tartışma (Devam)

A. Sınırlamalar ve Gelecek Araştırmalar

- Her ne kadar bu çalışma birçok yararlı bulgu ve sonuç ortaya koymuş olsa da, aşağıdaki dikkate değer sınırlamalara sahiptir. Analizimiz için kullandığımız veri seti yalnızca Güney Kore filmlerini ve oyuncularını içerdiğinden, Güney Kore film endüstrisine katkıda bulunan yabancı filmler ve oyuncular analizde dikkate alınmadı.
- İkinci olarak, film önerirken yalnızca oyuncu temelli tür korelasyonlarını dikkate aldık. Filmlerin bir takım başka önemli özellikleri de olduğundan, gelecekteki araştırmalar analize diğer özellikleri de dahil etmelidir.
- Son olarak, yalnızca 2010'dan 2019'a kadar vizyona giren veya yapım aşamasında olan Güney Kore filmlerini kullandık ve bu nedenle sonuçlar genelleştirilemez. Ayrıca, bu çalışmada diğer film öneri yaklaşımlarının karşılaştırılması tam olarak incelenmediği göz önüne alındığında, gelecekteki araştırmaların diğer film öneri teknikleriyle karşılaştırmayı genişletmesi gerekecektir.

Teşekkürler