

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

Metin Uslu

235B7014

BLM-5116 Veri Madenciliği ve Bilgi Keşfi

Prof. Dr. Songül Varlı

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

Özet

Bir öneri sistemi, kullanıcıyı bilgi arama zahmetinden kurtarır ve en uygun içeriği önermek için profilini analiz eder. Önerilerin gerçekleştirilmesi için içeriğe dayalı, işbirliğine dayalı ve bilgiye dayalı teknikler dahil olmak üzere çeşitli teknikler önerilmiştir. Kitap, müzik ve video gibi içerikleri önermek için öneri sistemleri kullanılır; ayrıca e-ticarete sıklıkla kullanılırlar. Özellikle, Güney Kore film endüstrisi, film öneri sistemlerinde yaygın olarak kullanılan, türlere dayalı işbirlikçi bir filtreleme yöntemini kullanan filmleri önermektedir. Ancak kullanıcılar ilk etapta film tavsiye hizmetleriyle karşılaştıklarında ya da filmlere ilişkin olarak oyuncu ya da yönetmen tercihleri gibi belirli tercihlere sahip olduklarında bu yöntem etkisiz olabiliyor. Bu bizi, oyuncuların filmografi bilgilerini ve 509 Güney Kore filminin türlerini dikkate alan içerik bazlı filtrelemeyi kullanan, oyuncu bazlı bir öneri sistemi önermeye motive etti. Önerilen sistemin etkinliği ve performansı, geleneksel tür odaklı öneri sistemiyle karşılaştırılarak değerlendirilir.

1. Giriş

İnternet ve veri teknolojilerindeki hızlı ilerlemeler, kullanıcılara çeşitli içerikler önermek için çok sayıda öneri sisteminin geliştirilmesine yol açmıştır. Bu kullanışlıdır çünkü kullanıcıların ilgilerini çeken içerikleri bulmak için çok sayıda web sitesinde arama yapmasına gerek yoktur. İçeriklerin birçok benzersiz özelliği ve özelliği olması nedeniyle, bilim adamları ve kuruluşlar rekabetçi pazarlarda başarıya ulaşmak için öneri sistemlerinin ve çözümlerinin kullanımını araştırmışlardır [1]–[3].

Bu eğilim göz önüne alındığında, içeriğe dayalı, işbirlikçi, bilgi odaklı ve hibrit teknikler dahil olmak üzere birçok dikkate değer öneri yaklaşımı tanıtılmıştır [4], [5].

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 [8].

Öneri sistemleri, kitapçılar ve video mağazaları da dahil olmak üzere çoğu çevrimiçi hizmet için vazgeçilmez hale geldi [9]. Bu tür hizmetler arasında Netflix [10] ve YouTube [11] 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 [12]. Her iki hizmet de piyasada son derece başarılı olmasına rağmen, bazı önemli sınırlamalar bu tekniklerin bu platformlar tarafından kullanılmasını engellemiştir. Yeterli sayıda kullanıcı 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 [13], [14].

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 [15], [16]. Örneğin Netflix ayrıca öneri sistemini, bireysel kullanıcıların her türe olan eğilimini yansıtarak çalıştırır [17]. 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.

2. Çalışma Hakkında

A. 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” [18]–[20]. 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 [21].

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 [22].

Hem öğe özelliklerini hem de kullanıcı profillerini dikkate alan içerik tabanlı öneri sistemlerinin temel amacı, kullanıcılara daha uygun öğeler önermektir. Bu, sistemin daha buluşsal odaklı yöntemleri ve sınıflandırma çözümlerini kullanması gerektiği anlamına gelir [23], [24].

B. İçerik Tabanlı Tavsiye Sistemleri ile İlgili Problemler

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 [25]. 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 [1]. Bu, belirli öğeleri önerirken ve tavsiye ederken dikkate değer kutuplaşma sorunlarına yol açabilir [26].

Bu sınırlamalar nedeniyle içerik tabanlı öneri sistemlerinin filmleri etkili bir şekilde önermesi her zaman mümkün olmayabilir. Bu eksikliği gidermek için içerik tabanlı film öneri sistemlerine yönelik film endüstrisi odaklı bakış açılarını kullandık. Her ne kadar bir dizi özellik belirli bir filmle ilişkilendirilse de, 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 [27]–[29]. Bu nedenle çalışmamız, Güney Kore film endüstrisindeki sinema oyuncularını dikkate alan içerik tabanlı bir öneri sistemi geliştirmeyi amaçlamaktadır.

3. Yöntem

A. Veri Toplama ve Ön İşlemler

Verileri Kore Film Konseyi'nden aldık.¹ Bu listede filmlerin listesi, ayrıntıları, film şirketleri, aktörler ve yönetmenler yer alıyordu. Her filmin benzersiz kodunu, başlığını, yapım yılını, açılış yılını, türünü, yapım bildirimini, yapımcı ülkeleri ve türünü çıkardık. Ayrıca yönetmenlerin isimlerini ve filmografilerini de aldık; oyuncuların isimleri ve filmografisi, doğum tarihleri, cinsiyetleri ve rolleri; filmografisi ve şirket sınıflandırmasıyla birlikte yapım şirketinin ve CEO'sunun adı. Veriler, her filme benzersiz bir numara atayarak belirlediğimiz 3429 kopya film başlığını içeriyordu. Veri kümesindeki filmler arasında uzun metrajlı filmler, kısa öyküler ve çok amaçlı kitaplar yer alıyordu. Yalnızca vizyona giren uzun metrajlı filmleri kullanmaya karar verdik.

Buna ek olarak, Naver film sitesinden² alınan, başlık, yıl, izleyici puanları ve izleyici sayısı ile her filmin İnternet kullanıcılarının puanları ve İnternet kullanıcılarının sayısının yanı sıra tarafından sağlanan derecelendirmeleri içeren verileri de kullandık. gazeteciler ve eleştirmenler. Veriler ayrıca her film hakkında şu bilgileri de içeriyordu: özellikler, yönetmen, oyuncular, sınıf, kümülatif izleyici katılımı, katılım oranı, cinsiyete göre ziyaretçi memnuniyeti, yaşa göre ziyaretçi memnuniyeti, yönetmenlik, oyunculuk, hikaye, görsel efektlerin kalitesi ve arka plan müzik. 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.

Kore Gişe Bilgi Sisteminden³ ek veriler topladık; veriler arasında film sıralaması, 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, 15 539 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 (Şekil 1).

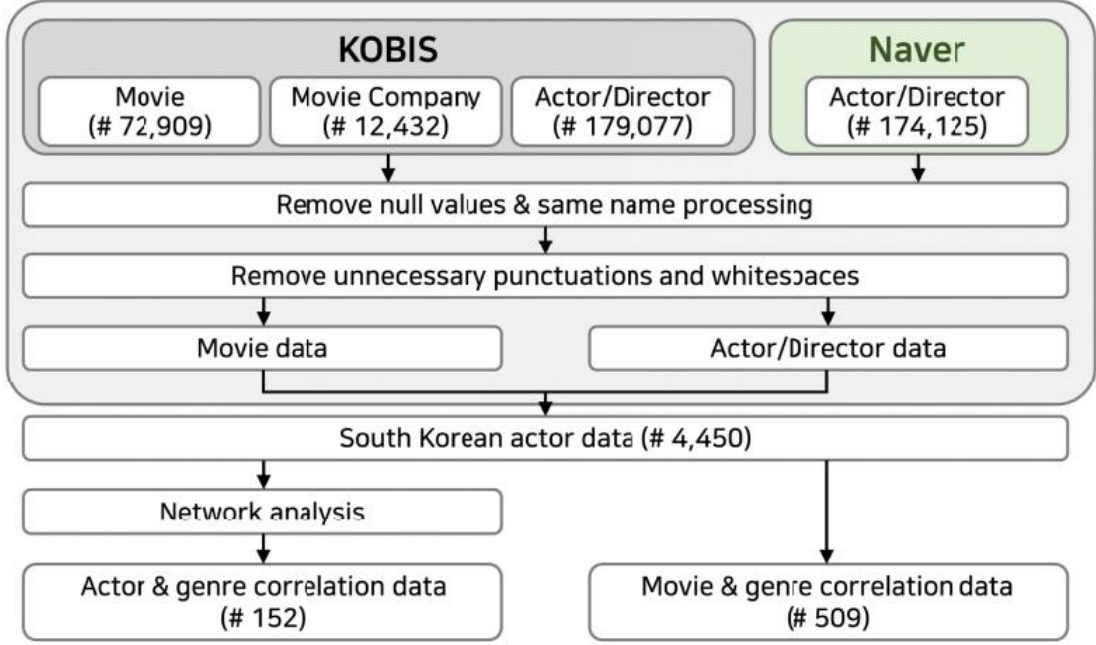


Fig. 1. Data preprocessing procedure for correlation analysis.

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 türlerine göre “animasyon türü”, “performans türü” ve “müzik türü”nü hariç tutarak “batı tiyatrosu türü”nü “çeşitli” kategorilerle birleştirdik. Tablo I’de bu çalışmada kullanılan film türleri sunulmaktadır.

Ayrıca içerik bazlı öneri sistemini uygulamak için Naver serisinden 4 509 veri grafiği çıkarıldı. Topladığımız veri grafikleri için veri ön işleme prosedürü Şekil 2’de gösterilmektedir. Verileri topladıktan sonra her grafikten gereksiz noktalama işaretlerini ve beyaz boşlukları kaldırdık. Daha

sonra kullanıcıların gönderilerini tokenize etmek için bir Python paketi olan python (konlpy)'da Kore doğal dil işlemeyi kullandık.

Aktör sütununu, yönetmen sütununu, tür sütununu ve olay örgüsü sütununu içerik sütunu olarak dize tipi bir sütun oluşturacak şekilde birleştirerek içerik tabanlı öneri sistemi için veri kümesini oluşturduk. Ayrıca kullanıcılara film önerirken gişe başarısının unsurlarını dikkate alarak nihai veri setine izleyicilerin reytingleri, satışları, seyircileri ve satış oranlarına ilişkin veriler ekledik.

B. Film ve Film Türü Arasında Rank Korelasyonunun 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.)

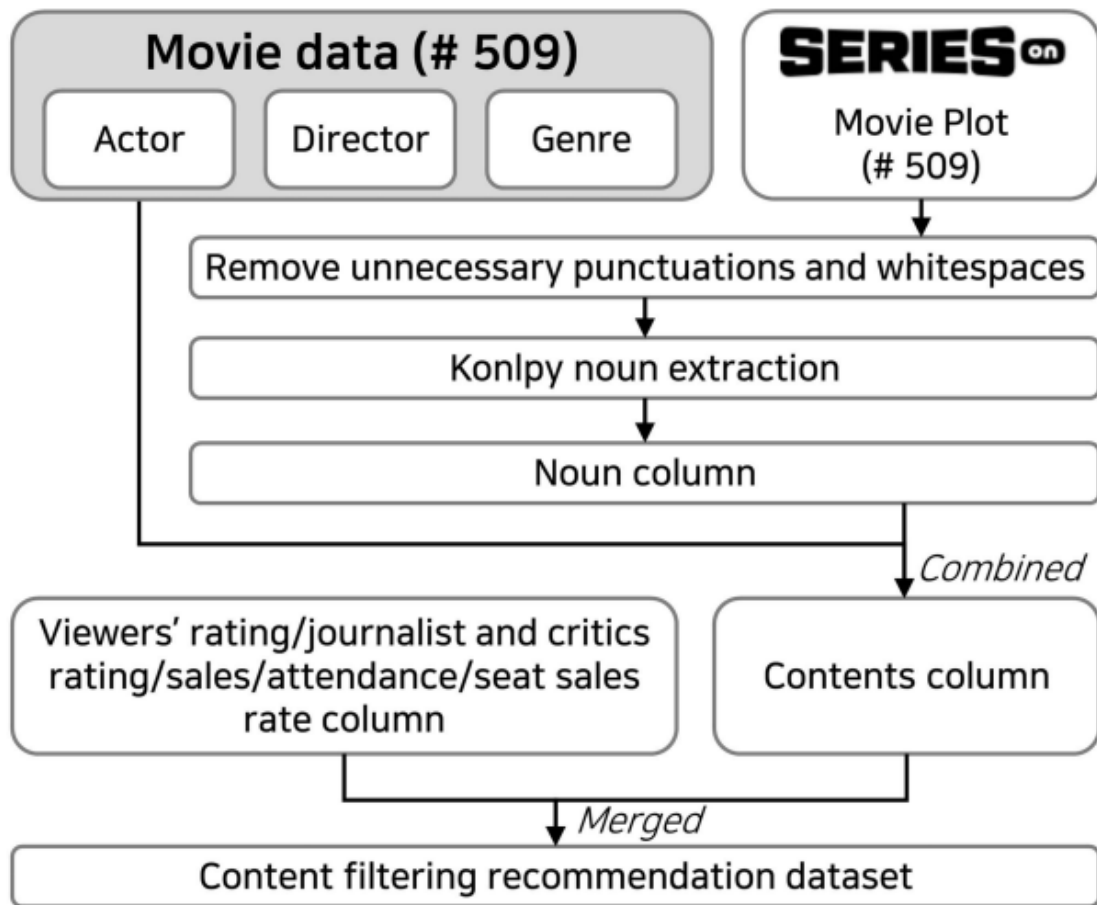


Fig. 2. Data preprocessing procedure for the content-based recommendation system.

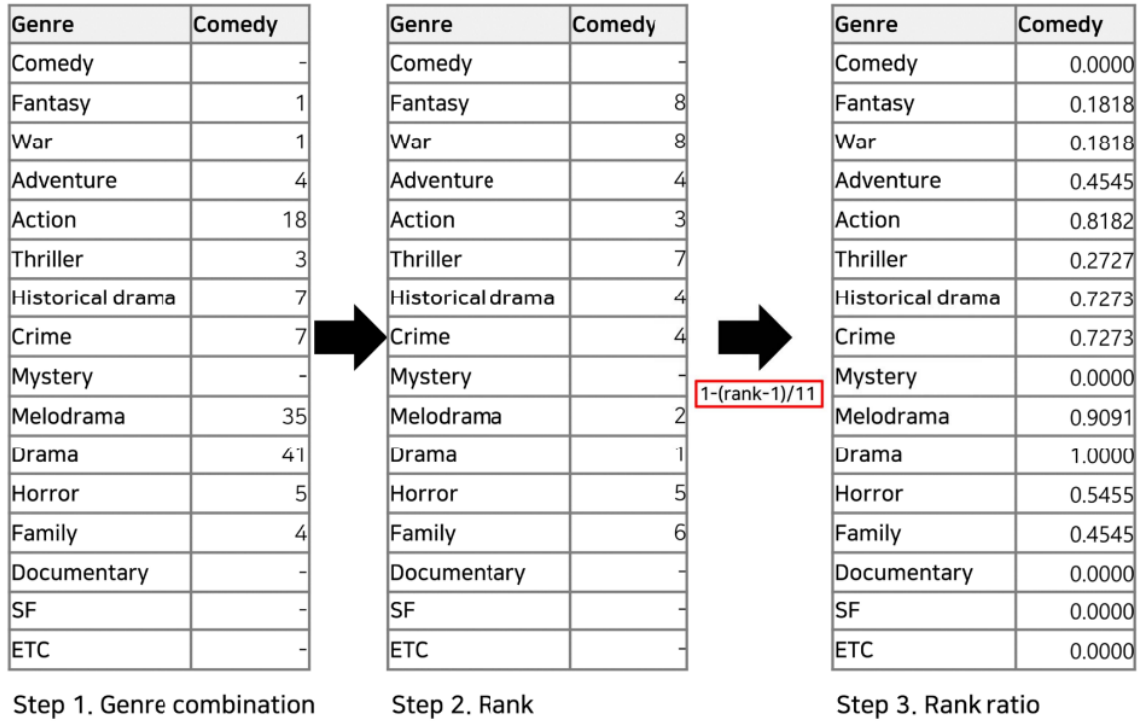


Fig. 3. Course that calculates rank correlations.

Şekil 3, sıra korelasyonu için hesaplama prosedürlerine genel bir bakış göstermektedir. Tablo II, kırmızıdaki değerlerin mavidekilerden farklı olduğunu ve film ile tür arasındaki sıra korelasyonunun son olarak bu iki bölümün ortalaması alınarak hesaplandığını göstermektedir. Örneğin “komedi” ve “fantezi”nin kırmızı renkteki sıralama düzeltmesi 0,33, mavi renkteki sıralama düzeltmesi ise 0,09'dur. Bu değerlerin ortalaması alınarak “komedi” ve “fantezi” kategorilerinin sıralama düzeltmesi 0,21 olarak belirlendi.

TABLE II
INCONSISTENCY BETWEEN THE UPPER SIDE (RED) AND LOWER SIDE (BLUE)

	Comedy	Fantasy	War	Adventure	Action	Thriller	Historical drama	Crime	Mystery	Melodrama	Drama	Horror	Family	Documentary	SF	ETC
Comedy	0.00	0.33	0.50	0.75	0.77	0.25	0.40	0.43	0.00	0.91	0.92	0.75	0.60	0.00	0.00	0.00
Fantasy	0.09	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.29	0.64	0.23	0.38	0.00	0.00	0.00	0.00
War	0.09	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.36	0.38	0.00	0.00	0.00	0.00	0.00
Adventure	0.36	0.00	0.00	0.00	0.54	0.00	0.20	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00
Action	0.73	0.33	0.50	0.50	0.00	0.75	0.80	0.86	0.43	0.36	0.85	0.25	0.40	0.00	0.80	0.00
Thriller	0.18	0.00	0.00	0.00	0.69	0.00	0.00	0.57	0.86	0.64	0.62	0.88	0.00	0.00	0.60	0.00
Historical drama	0.64	0.00	0.00	0.25	0.62	0.00	0.00	0.00	0.00	0.64	0.54	0.00	0.00	0.00	0.00	0.00
Crime	0.64	0.00	0.00	0.00	0.85	0.63	0.00	0.00	0.29	0.36	0.77	0.00	0.00	0.00	0.00	1.00
Mystery	0.00	0.33	0.00	0.00	0.54	0.63	0.00	0.29	0.00	0.36	0.54	0.75	0.00	0.00	0.00	0.00
Melodrama	0.82	0.67	0.50	0.00	0.31	0.13	0.20	0.29	0.29	0.00	0.77	0.50	0.40	0.00	0.00	0.00
Drama	0.91	0.83	0.75	0.00	0.92	0.88	0.60	0.71	0.71	0.82	0.00	0.25	0.80	0.00	0.60	0.00
Horror	0.45	0.67	0.00	0.00	0.31	0.38	0.00	0.00	0.57	0.73	0.15	0.00	0.00	0.00	0.60	0.00
Family	0.36	0.00	0.00	0.00	0.31	0.00	0.00	0.00	0.00	0.36	0.31	0.00	0.00	0.00	0.60	0.00
Documentary	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
SF	0.00	0.00	0.00	0.00	0.54	0.00	0.00	0.00	0.00	0.00	0.15	0.25	0.40	0.00	0.00	0.00
ETC	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.29	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00

TABLE III
GENRES OF MOVIES STARRING THE ACTOR “DON LEE”

Genre	Count	Genre	Count
Comedy	4	Mystery	-
Fantasy	2	Melodrama	2
War	-	Drama	13
Adventure	1	Horror	-
Action	10	Family	-
Thriller	6	Documentary	-
Historical drama	1	SF	-
Crime	9	ETC	-

C. Aktör ve Film Türü Arasında Korelasyonunun 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 on aksiyon ve dokuz 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.

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

(1)'de X, seçilen türdür ve σ_X , bir oyuncunun o belirli türe ait bir filmde kaç kez görüldüğünün standart sapmasıdır. Diğer tür Y olsun. Pearson korelasyon katsayısının değeri -1 ile 1 arasındadır. Eğer X ve Y arasındaki Pearson korelasyonu 1'e yakınsa, X ve Y gibi iki tür, belirli bir oyuncu ile kendi türü arasındaki korelasyon açısından benzerdir.

4. Sonuçlar

Şekil 4-6 sonuçları özetlemektedir. Bir film ile türü arasındaki korelasyon, bir oyuncu ile film türü arasındaki korelasyondan önemli ölçüde farklıdır. Filmler ve türleri arasında (film yapımcılarının bakış açıları) önemli bir korelasyon bulunabilse de, filmler ve aktörler arasındaki ilişki (izleyicilerin bakış açıları) ilişkilendirilemez. Örneğin, filmler ve türler dikkate alındığında “komedi” türü ile “macera” türü arasında yüksek korelasyon (0,56) varken, aktörler ve türler arasında düşük bir korelasyon görülmektedir. Tablo IV-VI, filmler ve türler arasındaki korelasyon matrisini özetlemektedir.

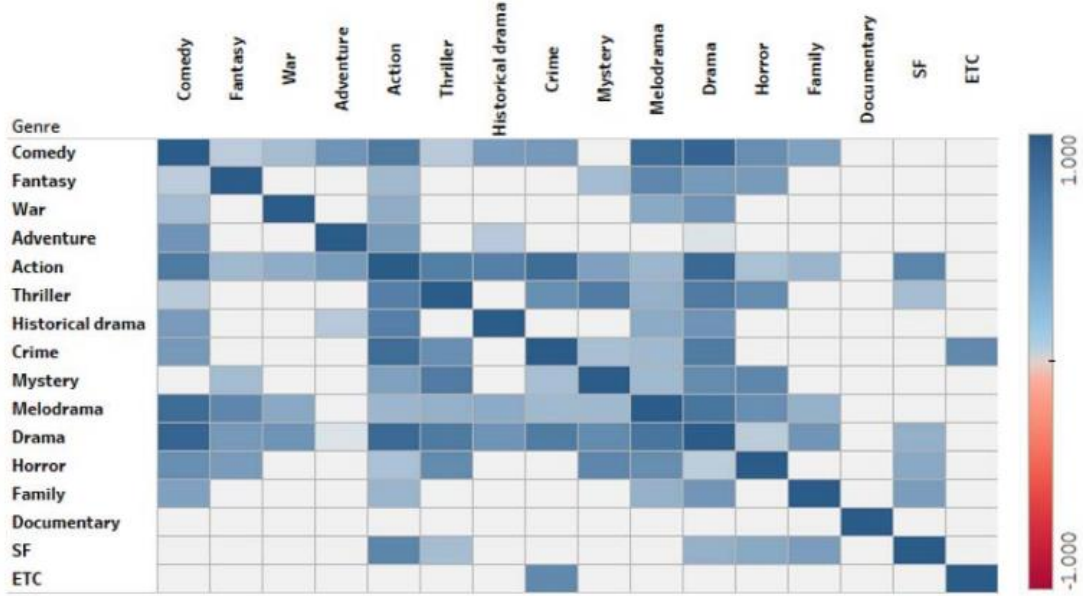


Fig. 4. Rank correlation plot between a movie and its genres.

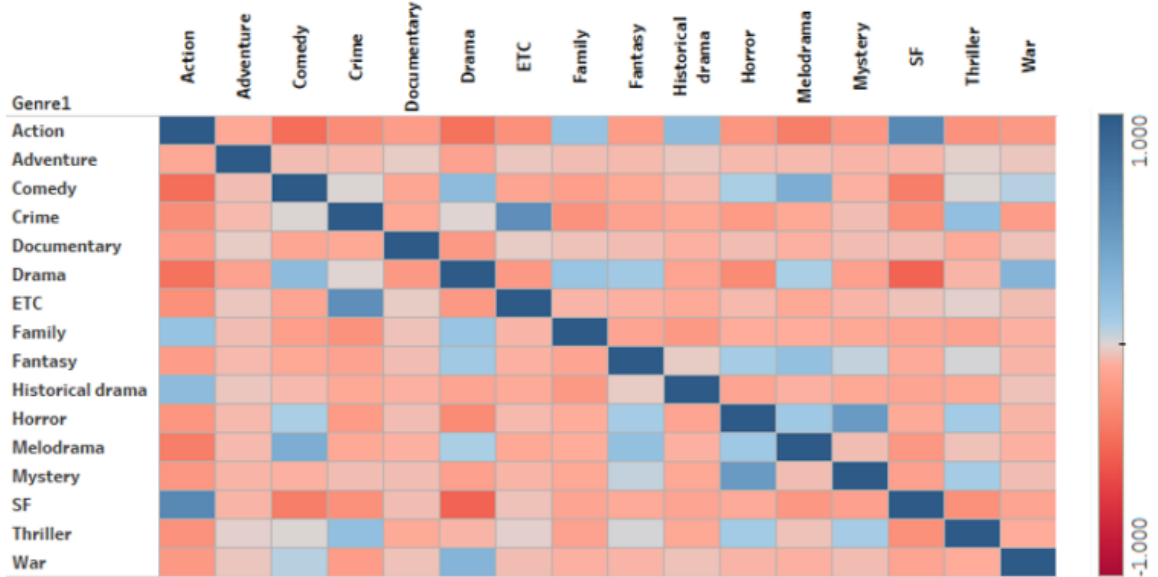


Fig. 5. Correlation plot between an actor and their genres.

Önerilen öneri sistemini değerlendirmek için on katılımcıyla bir kullanıcı anketi gerçekleştirdik. Beş katılımcı film türlerini dikkate alarak filmlerini seçtiklerini söylerken, diğer beş katılımcı film seçimlerini filmlerde rol alan oyunculara göre belirlediklerini belirtmiştir.

TABLE IV
GENRES WITH THE RANK CORRELATION GREATER THAN 0.5 BETWEEN A MOVIE AND ITS GENRES, AND
GENRES WITH THE CORRELATION OF 0.1 TO -0.1 BETWEEN AN ACTOR AND THEIR GENRES

	Rank correlation between movie and genre	Correlation between actor and genre
Comedy	Adventure, Action, Historical drama, Crime, Melodrama, Drama, Horror	War, Adventure, Thriller, Historical drama, Crime, Mystery, Horror
Fantasy	Melodrama, Drama, Horror	War, Adventure, Thriller, Historical drama, Mystery, Document, SF, ETC
War	Drama	Comedy, Fantasy, Adventure, Thriller, Historical drama, Mystery, Document, SF, ETC
Adventure	Comedy, Action	Comedy, Fantasy, War, Thriller, Historical drama, Crime, Mystery, Melodrama, Horror, Family, Document, SF, ETC
Action	Comedy, Adventure, Thriller, Historical drama, Crime, Drama, SF	-
Thriller	Action, Crime, Mystery, Drama, Horror	Comedy, Fantasy, War, Adventure, Melodrama, Drama, Document, ETC
Historical drama	Comedy, Action, Drama	Comedy, Fantasy, War, Adventure, Melodrama, Document, ETC
Crime	Comedy, Action, Thriller, Drama, ETC	Comedy, Adventure, Mystery, Drama
Mystery	Thriller, Drama, Horror	Comedy, Fantasy, War, Adventure, Crime, Melodrama, Document, ETC
Melodrama	Comedy, Fantasy, Drama, Horror	War, Adventure, Thriller, Historical drama, Mystery, Drama, Family, Document
Drama	Comedy, Fantasy, War, Action, Thriller, Historical drama, Crime, Mystery, Melodrama, Family	Thriller, Crime, Melodrama
Horror	Comedy, Fantasy, Thriller, Mystery, Melodrama	Comedy, War, Adventure, Family, Document, SF, ETC
Family	Drama, SF	War, Adventure, Melodrama, Horror, Document, ETC
Documentary	-	Fantasy, War, Adventure, Thriller, Historical drama, Mystery, Melodrama, Horror, Family, SF, ETC
SF	Action, Family	Fantasy, Adventure, Horror, Document, ETC
ETC	Crime	Fantasy, War, Adventure, Thriller, Historical drama, Mystery, Horror, Family, Document, SF

TABLE V
RANK CORRELATION BETWEEN MOVIES AND GENRES

	Comedy	Fantasy	War	Adventure	Action	Thriller	Historical drama	Crime	Mystery	Melodrama	Drama	Horror	Family	Documentary	SF	ETC
Comedy	1.00	0.21	0.30	0.56	0.75	0.22	0.52	0.53	0.00	0.86	0.92	0.60	0.48	0.00	0.00	0.00
Fantasy		1.00	0.00	0.00	0.32	0.00	0.00	0.00	0.31	0.65	0.53	0.52	0.00	0.00	0.00	0.00
War			1.00	0.00	0.40	0.00	0.00	0.00	0.00	0.43	0.57	0.00	0.00	0.00	0.00	0.00
Adventure				1.00	0.52	0.00	0.23	0.00	0.00	0.00	0.08	0.00	0.00	0.00	0.00	0.00
Action					1.00	0.72	0.71	0.85	0.48	0.34	0.88	0.28	0.35	0.00	0.67	0.00
Thriller						1.00	0.00	0.60	0.74	0.38	0.75	0.63	0.00	0.00	0.30	0.00
Historical drama							1.00	0.00	0.00	0.42	0.57	0.00	0.00	0.00	0.00	0.00
Crime								1.00	0.00	0.32	0.74	0.00	0.00	0.00	0.00	0.64
Mystery									1.00	0.32	0.63	0.66	0.00	0.00	0.00	0.00
Melodrama										1.00	0.79	0.61	0.38	0.00	0.00	0.00
Drama											1.00	0.20	0.55	0.00	0.38	0.00
Horror												1.00	0.00	0.00	0.43	0.00
Family													1.00	0.00	0.50	0.00
Documentary														1.00	0.00	0.00
SF															1.00	0.00
ETC																1.00

TABLE VI
GAP BETWEEN THE CORRELATION OF ACTOR-GENRE AND RANK CORRELATION OF MOVIE-GENRE

	Comedy	Fantasy	War	Adventure	Action	Thriller	Historical drama	Crime	Mystery	Melodrama	Drama	Horror	Family	Documentary	SF	ETC
Comedy	0.00	0.10	0.23	0.51	0.33	0.21	0.46	0.52	0.08	0.51	0.68	0.51	0.32	0.12	0.34	0.13
Fantasy		0.00	0.07	0.06	0.15	0.02	0.02	0.14	0.26	0.45	0.40	0.41	0.13	0.05	0.10	0.08
War			0.00	0.03	0.21	0.09	0.04	0.17	0.05	0.35	0.28	0.07	0.08	0.04	0.13	0.05
Adventure				0.00	0.41	0.01	0.20	0.06	0.07	-0.06	0.06	0.06	0.05	0.02	0.07	0.03
Action					0.00	0.49	0.47	0.59	0.28	0.00	0.48	0.07	0.17	0.17	0.04	0.24
Thriller						0.00	0.11	0.39	0.63	0.34	0.68	0.51	0.14	0.10	0.06	0.01
Historical drama							0.00	0.11	0.11	0.34	0.44	0.13	0.19	0.08	0.13	0.10
Crime								0.00	0.24	0.21	0.74	0.18	0.23	0.11	0.24	0.06
Mystery									0.00	0.27	0.48	0.17	0.11	0.05	0.15	0.07
Melodrama										0.00	0.70	0.47	0.29	0.08	0.20	0.11
Drama											0.00	0.07	0.38	0.19	0.09	0.19
Horror												0.00	0.09	0.05	0.33	0.06
Family													0.00	0.04	0.37	0.07
Documentary														0.00	0.05	0.02
SF															0.00	0.04
ETC																0.00

Ankette, bir katılımcı en sevdiği filmi seçtikten sonra sistem, türleriyle en yüksek korelasyona sahip on film ve aktörler ve türler arasında en yüksek korelasyona sahip on film olarak düzenlenen 20 filmi önerdi. Daha sonra her katılımcı, önerilen filmleri on puanlık Likert ölçeğinde (1: "kesinlikle tercih

etmiyorum" ile 10: "kesinlikle tercih ediyorum") tercihlerini belirtti. Örneğin Tablo VII, Parazit'i en sevdiği film olarak seçen bir katılımcıya sunulan film listesini içermektedir.

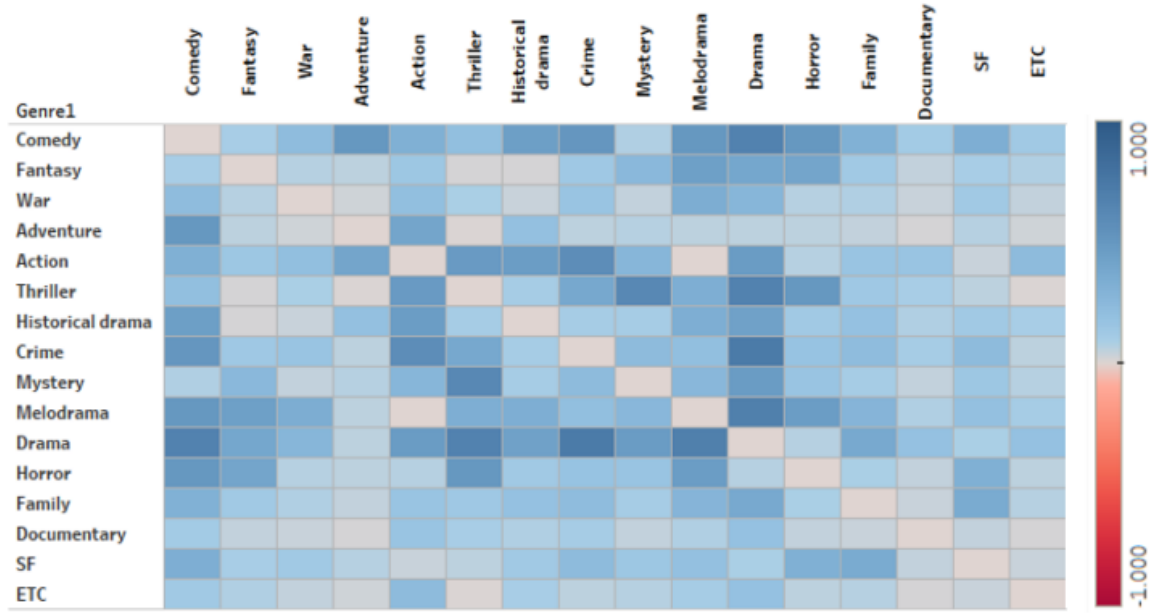


Fig. 6. Difference between the correlation between an actor and their genre and the rank correlation between a movie and its genre.

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>

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)

Tablo VIII anket sonuçlarını göstermektedir. Filmleri oyunculara göre seçen kullanıcıların, tür odaklı yaklaşımlara göre oyuncu türü yaklaşımları tarafından önerilen filmleri tercih etme olasılıkları daha yüksekti ($p < 0,05$).

5. Değerlendirme ve Tartışma

Bu çalışma, tür ve aktörleri dikkate alarak belirli filmleri önermek için içerik tabanlı filtrelemeyi kullanmayı amaçlamıştır. Bu yaklaşımın etkinliğini iki içerik bazlı öneri sistemi oluşturarak değerlendirdik: filmler ve türler arasındaki korelasyona dayalı bir sistem ve aktörler 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. Çalışmamızın bulgularının Güney Kore film endüstrisi için çeşitli pratik ve akademik çıkarımları var. Akademik açıdan bakıldığında, çalışmanın sonuçları, içeriğe dayalı öneri sistemleri geliştirilirken kullanılabilecek gelecekteki potansiyel teknikleri vurgulamaktadır. Ayrıca, bir dizi hibrit öneri sistemi,

içeriğe dayalı öneri sistemlerinin çeşitli prosedürlerini içerdiğinden, çalışmamızın sonuçları hibrit öneri sistemlerinin geliştirilmesine de katkıda bulunabilir.

Pratik açıdan bakıldığında çalışmamız, çeşitli içeriklere sahip hizmetler için içerik bazlı öneri sistemlerinin oluşturulabileceği dikkate değer bir temel sağlar. Ayrıca, Güney Kore film endüstrisinde kullanıcı tarafından kişiselleştirilmiş içeriklere ilişkin yalnızca birkaç çalışma yürütüldüğü için, bu çalışmanın sonuçları, potansiyel müşterilere daha kısa ve doğru tavsiye hizmetleri önermek için kullanılabilir. Ayrıca, daha müşteri odaklı bakış açıları Güney Kore film endüstrisinin daha da geliştirilmesine hizmet edebilir.

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.

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>