

İşbirlikçi Filtreleme ile Film Öneri Sistemi

Halil İbrahim Özhanlı
Elektronik-Haberleşme Mühendisliği
Kocaeli Üniversitesi
Kocaeli, TÜRKİYE
halilibrahimozhanli@gmail.com

Özet— Öneri sistemleri internet teknolojisinin yaygınlaşması ve kullanıcılara ait bilgilerin daha kolay elde edilebiliyor olmasından dolayı son yıllarda araştırmacıların odağında olan bir konu olmuştur. Esas olarak bilgi edinmeyi amaçlayan bu sistemler, pratik uygulamalar konusunda da dikkat çekicidir. Amazon, Netflix gibi internet dünyasındaki çok büyük oyuncular kullanıcılarına gerçek zamanlı olarak önerilerde bulunan öneri sistemleri geliştirmişlerdir. Bu çalışma kapsamında Netflix Priza veri seti kullanılarak kullanıcılara film önerisinde bulunan bir öneri sistemi geliştirilmiştir.

Anahtar Kelimeler— öneri sistemleri, netflix, işbirlikçi filtre

I. GİRİŞ

Öneri sistemleri kullanıcıya fazladan bir çaba içerisine girmeden tercihleri doğrultusunda uygun öğeler tavsiye eden sistemlerdir. 1990'ların ortasından itibaren öneri sistemleri ayrı bir araştırma konusu olmuştur [1]. "Tipik bir tavsiye sisteminde insanlar tavsiyeleri girdi olarak sunar ve sistem yorumlayarak alıcılara ulaştırır." [2]. Öneri sistemlerinin gerçekleştirilmesinde İşbirlikçi Filtreleme (CF) yaygın olarak kullanılan bir yöntemdir ve bir ürünü onu değerlendiren kişilerin fikirleri doğrultusunda başka kişilere önerme süreci olarak tanımlanmaktadır.

İşbirlikçi filtreleme (CF) son yıllarda ortaya çıkmış olmasına karşın temelde insanların asırlardır yapmış olduğu fikir alışverişinden köken almaktadır [3].

Bu çalışmada film izleyicilerine film önerme aşamasında kullanılabilmesi amacı ile İşbirlikçi Filtreleme (CF) algoritması tasarlanmış ve gerçekleştirilmiştir. Netflix Priza'dan alınan oylamalar veri seti olarak kullanılmıştır. Veri setinde eğitim verisi olarak 3.25 milyon oylama ve test verisi olarak ise 100 bin oylama bulunmaktadır.

II. ÖNERİ YÖNTEMLERİ

Bir öneri sisteminin gerçekleştirilmesi üç ana aşamadan oluşur. Öneri sistemine yeterli giriş verisi sağlamak, bu verilerden gerekli çıkarımlar yaparak tahminler oluşturmak ve bu tahminleri kullanıcılara öneri olarak sunmak. Kullandıkları yöntemlere göre öneri sistemleri farklılıklar göstermektedir. Bazı öneri sistemlerinde içerikler arasındaki benzerlikler göz önüne alınırken, diğerleri kullanıcı benzerliklerini veya içerik benzerlikleri kullanmaktadır. Öneri sistemleri fonksiyonel olarak detaylandırılacak olursa genel olarak üç kısımdan oluştuğunu söylemek mümkündür.

A. Kullanıcı-Kullanıcı İşbirlikçi Filtre

Esas amaç birbirine benzeyen kullanıcıları bulmaktır. Örnek olarak bir kullanıcıya öneri yapılacağını varsayalım. Öneri yapılacak olan kullanıcının daha önceden yaptığı tercihler dikkate alınarak kullanıcı ile benzer tercihleri yapan diğer kullanıcılar gruplandırılır ve gruplandırılan bu kullanıcıların tercihleri içerisinde öneri yapılır.

B. Nesne-Nesne İşbirlikçi Filtre

Esas amaç birbirine benzeyen nesneleri bulmaktır. Benzer nesneler bulunduktan sonra kullanıcının geçmişteki tercihleri ile benzer olan diğer elemanlar kullanıcıya öneri olarak sunulur.

C. Hibrit Yaklaşım

Bu yaklaşımda kullanıcı-kullanıcı ve nesne-nesne işbirlikçi filtreleri birlikte kullanılır. Böylelikle tek bir yöntemin sahip olduğu dezavantajları minimuma indirerek hata oranını minimuma indirmektedir.

III. YÖNTEM

Netflix veri setinde kullanıcılar filmleri 1 ile 5 yıldız arasında değerlendirmiştir. Bu çalışmada kullanıcılardan alınan değerlendirmeler kullanılarak, kullanıcı-kullanıcı işbirlikçi filtreleme yöntemi ile diğer kullanıcılara film önerisinde bulunan bir öneri sistemi geliştirilmiştir.

Tablo 1: Netflix kullanıcı oyları gösterimi

	The Matix	Speed	Sideways	Brokeback Mountain
Kullanıcı 1	1	2	5	
Kullanıcı 2		3	5	4
Kullanıcı 3	5	5	2	3

İşbirlikçi filtreleme ilk olarak GroupLens Usenet makale öneri sisteminde [4] kullanılmıştır. Ringo müzik öneri sistemi [5] ve BellCore video öneri sistemi de bu yöntemi kullanmıştır. Kullanıcı-kullanıcı işbirlikçi filtrelemede (CF) kullanıcı benzerliklerini bulmak için Pearson korelasyonu kullanılmıştır. Bu korelasyonda aralarında benzerlik aranan kullanıcılar için iki kullanıcının da oy verdiği filmlerin kesişim kümesi üzerinden Denklem 1 hesaplanmaktadır.

$$s(u, v) = \frac{\sum_{i \in I_u \cap I_v} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I_u \cap I_v} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I_u \cap I_v} (r_{v,i} - \bar{r}_v)^2}}$$

Denklem 1: Pearson Korelasyonu

Kullanıcı benzerliklerinden sonra öneri için ise kullanıcıların ağırlıklı ortalamaları hesaplanmıştır.

$$p_{u,i} = \bar{r}_u + \frac{\sum_{u' \in N} s(u, u')(r_{u',i} - \bar{r}_{u'})}{\sum_{u' \in N} |s(u, u')|}$$

Denklem 2: Ağırlıklı ortalama

Netflix yarışmasını kazanan ekip ise kullanıcı-kullanıcı işbirlikçi filtresini Denklem 3'teki gibi güncellemiş ve lineer regresyon yapısına benzetmiştir.

$$p_{u,i} = \bar{r}_u + \sum \theta_{i,j}^u (r_{u,i} - \bar{r}_u)$$

Denklem 3: Kullanıcı-Kullanıcı CF Linear Regrasyon Benzetimi

Elde edilen sonuçlar "Predictions.txt", tahmin değeri 4.0 değerinden büyük olan sonuçlar ayrıca "Recommendations.txt" dosyasına yazılmıştır.

Son olarak ise MAE(Mean Absolute Error) ve RMSE(Root Mean Square Error) değerleri hesaplanmıştır.

IV. SONUÇLAR

Hesaplamalar sonunda MAE değeri 0.6949234976496851 olarak, RMSE değeri ise 0.8844601266219841 olarak hesaplanmıştır. 53 adet oylama sıfır hata ile hesaplanmıştır ve bu sonuç iyi bir sonuç olarak gösterilebilir. Ayrıca MAE ve RMSE değerleri de dikkate alındığında bu sistemin bir öneri sistemi olarak kullanılabileceği söylenebilir. Veri setindeki kişi sayısı ve kişilerin oyladığı film sayıları artırılarak daha doğru sonuçlar elde edilebileceği öngörülmektedir.

KAYNAKLAR

- [1] Adomavicius, G. and Tuzhilin, A (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. Knowledge and Data Engineering, IEEE Transactions on, 17(6):734–749
- [2] Resnick, P., Varian, H. R., Recommender systems, Communications of the ACM, 40(3), 56-58, 1997
- [3] J. Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen (2006), Collaborative Filtering Recommender Systems
- [4] P. Resnick, N. Iacovou, M. Suchak, P. Bergstrom, and J. Riedl, "GroupLens: an open architecture for collaborative filtering of netnews," in ACM CSCW '94, pp. 175–186, ACM, 1994
- [5] U. Shardanand and P. Maes, "Social information filtering: Algorithms for automating "word of mouth"," in ACM CHI '95, pp. 210–217, ACM Press/Addison-Wesley Publishing Co., 1995