Öneri Sistemleri İçin İşbirlikçi Derin Öğrenme ile Bayesci Negatif Olmayan Matris Ayrımlaştırmasının Karşılaştırılması
Comparison of Collaborative Deep Learning and Nonnegative Matrix Factorization for Recommender Systems

Mine Öğretir, Ali Taylan Cemgil
Bilgisayar Mühendisliği Bölümü, Boğaziçi Üniversitesi, İstanbul, Türkiye
mine.ogretir@boun.edu.tr, taylan.cemgil@boun.edu.tr

***Özetçe* —İşbirlikçi filtreleme ve içerik bazlı metodlar öneri sistemlerinde kullanılan iki temel yaklaşım iken birleşik modeller ikisinin de avantajlarını kullanmaktadır. Bu makalede Bayesci İstiflenmiş Gürültü Giderici Otokodlayıcılarını içerik bilgisinin öğrenilmesi amacıyla kullanıldığı birleşik bir model ile işbirlikçi filtreleme yöntemi olarak kullanılan Bayesci Negatif Olmayan Matris Ayrıştırma modeli karşılaştırılmıştır. Sıkı bağlı birleşik model olan İşbirlikçi Derin Öğrenme'nin işbirlikçi filtrelemeye göre daha başarılı sonuç verdiği gösterilmiştir.**

Anahtar Kelimeler—Öneri Sistemleri, İstiflenmiş Gürültü Giderici Otokodlayıcılar, Birleşik Metodlar

***Abstract*—Collaborative filtering and content-based methods are two main approaches for recommender systems, and hybrid models use advantages of both. In this paper, we made a comparison of a hybrid model, which uses Bayesian Staked Denoising Autoencoders for content learning, and a collaborative filtering method, Bayesian Nonnegative Matrix Factorisation. It is shown that the tightly coupled hybrid model,Collaborative Deep Learning, gave more successful results comparing to collaborative filtering methods.**

Keywords—Recommender Systems, Stacked Denoising Autoencoders, Hybrid Models

I. GİRİŞ

Internet servislerindeki içerik artışından dolayı kişilerin bu bilgilere kendi eğilimleri, beğenileri vs. doğrultusunda etkin olarak erişebilmeleri de gittikçe zorlaşmaktadır. Kullanıcıların ilgili oldukları bilgilere etkin erişebilmelerinin sağlanması için öneri sistemleri geliştirilmektedir. Öneri sistemleri akademik makaleler, filmler, haberler, müzik gibi ürün kalemlerinin doğru kişilere ulaştırılabilmesi amacıyla kullanılan araç ve tekniklerdir. Bu konuda kullanılan metodlar [17] işbirlikçi filtreleme, içerik bazlı metodlar ve bu iki metodun bir arada kullanıldığı birleşik metodlar olarak temelde üçe ayrılmaktadır. İşbirlikçi filtreleme yöntemleri sadece derecelendirme bilgisini kullanırken [15] [6], içerik tabanlı metodlar kişi profilleri, ürün açıklamaları gibi yardımcı içerik bilgilerini kullanmaktadır [5] [16]. Birleşik modeller ise iki yöntemi de içerirler. Birleşik modelller de sıkı bağlı ve gevşek bağlı olmak üzere, iki yaklaşım arasında karşılıklı bilgi alışverişi olması ya da tek taraflı bilgi akışı olmasına göre ikiye ayrılabilmektedir [2].

İşbirlikçi filtreleme yöntemleri doğrudan derecelendirmeleri kullandığı için yeni bir kalemin eklenmesi durumunda öneri yapamamaktadır (Soğuk başlangıç problemi). Ayrıca bilinen derecelendirmelerin seyrek olması durumunda da başarılı önerilerde bulunamamaktadır. Diğer taraftan içerik bazlı metodlarda derecelendirme bilgileri kullanılamamaktadır. Bu çalışmada birleşik bir model ile işbirlikçi filtreleme yöntemi karşılaştırılmıştır. Kullanılan birleşik model sıkı bağlı birleşik modellere örnek teşkil etmektedir [2] ve özellikle son günlerde gündemde olan derin öğrenme tekniklerinin bu sınıf modellere uygulanmasının bir örneğidir. Karşılaştırma için Bayesci negatif olmayan matris ayrıştırma yöntemi kullanılmıştır. [1]

Sıkı bağlı birleşik modeller arasında olan işbirlikçi konu bağlanımı çalışmasında [4] gizli Dirichlet dağıtımı konu modeli ile olasılıksal matris ayrıştırması entegre edilmiştir. Bu modelde içerik bilgisinin işlenmesinde kelime çantası kullanılmıştır. Bazı modeller [9]–[11] alışılmış matris ayrıştırması yerine kısıtlı Boltzmann Makinelerini ya da yinelenen sinir ağlarını kullanmışlardır. Ancak bunlar içerik bilgisini kullanmayan, derin öğrenme tekniklerinin işbirlikçi filtrelemeye uygulanması örneğidir. Müzik tavsiyesi için sunulan bazı çalışmalar [12] [13] içerik bilgisinin kullanımında kelime çantasında dolambaçlı sinir ağlarını ve derin inanç ağlarını kullanmışlardır. Ancak bu modeller belirlenimcidirler ve gürültü modellemesine sahip değildir. Bu nedenle de gürbüzlülükleri düşüktür. İşbirlikçi derin öğrenme'de [2] (İDÖ) içerik bilgisinin kelime çantası temsilinden olasılıksal yığın gürültü giderici otokodlaması aracılığıyla özellikler çıkarılması hedeflenmiştir. Bu özellikler de olasılıksal işbirlikçi filtrelemede kullanılmıştır. Ayrıca öğrenme sırasında bu iki bölümün bilgileri diğer tarafı da besleyecek bir yapı kurulmuştur. Bu makalede, işbirlikçi filtreleme yöntemi olarak negatif olmayan

978-1-5090-6494-6/17/$31.00

matris ayrıştırmasında Bayesci çıkarım yöntemi kullanılarak (BNOMA) İDÖ ile karşılaştırma yapılmıştır.

II. Model- İşbirlikçi Derin Öğrenme

İDÖ, hiyerarşik Bayesci bir modeldir. İçerik bilgisi çok katmanlı otokodlayıcı sinir ağları modelinden çıkarılan özelliklerle modele dahil edilmektedir. Derecelendirme matrisi,$\mathbf{R}$, kalemleri (makale, film, haber vs.) ifade eden matris $\mathbf{V}$ ile kişileri ifade eden matrisin $\mathbf{U}$ çarpımı aracılığıylayla oluşturulmaktadır. Her bir kalemi ifade eden vektör, içerik bilgisi ağının darboğaz katmanının çıktısı ile veri bağıl konum vektörünün toplamından üretilmektedir. İşbirlikçi filtreleme kısmında olasılıksal matris ayrıştırma yöntemi kullanılmıştır. Temelde problem, seyrek veriye sahip olan $\mathbf{R}$ matrisindeki eksik verileri bulmaktır.

Model, eğitim ve test verisi olarak kesinlik içermeyen, imalı geri bildirimleri kullanmaktadır. Sadece kullanıcıların hangi kalemleri tercih ettiği bilgisine erişim vardır. Hangi kalemleri tercih etmedikleri, kalemle karışılaşıp karşılaşmadıkları bilgisi mevcut değildir. Bir kişinin bir kalemi tercih etmemesi o kalemi hiç görmediğini de, görüp tercih etmediğini de gösterebilir. Bu ikisi arasındaki ayrım yapılamamaktadır.

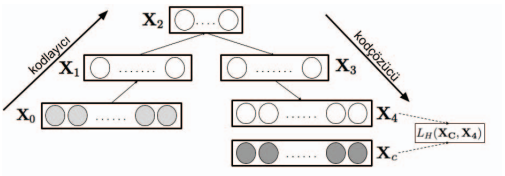
Bundan sonraki ilk alt bölümde istifleyici gürültü giderci otokodlayıcılar, sonra İDÖ ve son alt bölümde BNOMA kısaca açıklanmıştır.

A. İstiflenmiş Gürültü Giderici Otokodlayıcılar ve Genelleştirilmiş Bayesci İGGO

İstiflenmiş Gürültü Giderici Otokodlayıcılar [3] girdi olarak verinin gürültülü halini alan ve çıktı olarak verinin gürültüsüz halini kullanan, verinin özellik gösterimini öğrenmeyi amaçlayan ileri beslemeli sinir ağlarıdır. İki seviye İGGO Şekil 1'de gösterilmiştir. Genelde grafikte en üstte görülen gizli katman özelliklerin belirlendiği darboğaz olarak kısıtlandırılır. İGGO aşağıdaki optimizasyon problemini çözer:

$$\min_{\{W_l\},\{b_l\}} \|\mathbf{X}_c - \mathbf{X}_L\|_F^2 + \lambda \sum_l (\|\mathbf{W}_l\|_F^2 + \|\mathbf{b}_l\|_2^2) \quad (1)$$

İfadede λ düzenlileştirme parametresi ve $\|.\|_F$ Frobenius normunu göstermektedir. $\mathbf{X}_c$, J tane kalemin (film, makale ya da kitap vs olabilir) metin bilgisinin S boyutundaki sözlüğe göre oluşturulan tf-idf (terim frekansı – ters metin frekansı) vektörlerin birleşimi olan JxS boyutlu temiz matristir. $\mathbf{X}_L$ İGGO'nun son gizli katmanını, $\mathbf{W}_l$ ve $\mathbf{b}_l$ yapay sinir ağının l. katmanındaki ağırlık parametrelerini ifade etmektedir.

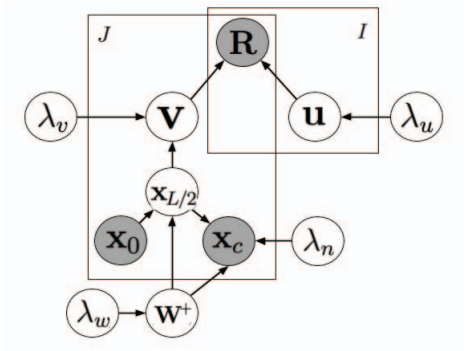


Şekil 1: İki seviye İGGO, L=4

İGGO'nun genelleştirilmiş Bayesci versiyonunun üretici süreci şu şekildedir:

1) İGGO ağının her l katmanı için aşağıdaki dağılımdan değerleri çek,
   a) Ağırlık matrisleri $\mathbf{W}_l$ ve $\mathbf{b}_l$'nin her n. sütunları için,
   $$\mathbf{W}_{l,*n} \sim \mathcal{N}(\mathbf{0}, \lambda_w^{-1}\mathbf{I}_{K_l}); \mathbf{b}_l \sim \mathcal{N}(\mathbf{0}, \lambda_w^{-1}\mathbf{I}_{K_l})$$
   b) $\mathbf{X}_l$'in her bir j satırı için,
   $$\mathbf{X}_{l,j*} \sim \mathcal{N}(\sigma(\mathbf{X}_{l-1,j*}\mathbf{W}_l + \mathbf{b}_l), \lambda_s^{-1}\mathbf{I}_{K_l})$$
2) Her j kalemi için,
   a) Temiz girdi çek, $\mathbf{X}_{c,j*} \sim \mathcal{N}(\mathbf{X}_{L,j*}, \lambda_n^{-1}\mathbf{I}_J)$

$\sigma(.)$, sigmoid fonksiyonunu göstermektedir. Bu sinir ağı yapısında ilk $L/2$ katman kodlayıcı olarak görev görürken son $L/2$ katman da kodçözücü olarak görev almaktadır. Bu sürecin sonsal olasılık dağılımının maksimizasyonu tekrar oluşturma hatasının minimizasyonuna denk gelmektedir.



Şekil 2: İDÖ modelinin grafik gösterimi.

B. İşbirlikçi Derin Öğrenme

Tanımlanan Bayesci İGGO'nun kullanıldığı İDÖ üretici süreç şu şekilde tanımlanmaktadır:

1) İGGO ağının her l katmanı için aşağıdaki dağılımdan değerleri çek,
   a) Ağırlık matrisleri $\mathbf{W}_l$ ve $\mathbf{b}_l$'nin her n. sütunları için,
   $$\mathbf{W}_{l,*n} \sim \mathcal{N}(\mathbf{0}, \lambda_w^{-1}\mathbf{I}_{K_l}); \mathbf{b}_l \sim \mathcal{N}(\mathbf{0}, \lambda_w^{-1}\mathbf{I}_{K_l})$$
   b) $\mathbf{X}_l$'in her bir j satırı için,
   $$\mathbf{X}_{l,j*} \sim \mathcal{N}(\sigma(\mathbf{X}_{l-1,j*}\mathbf{W}_l + \mathbf{b}_l), \lambda_s^{-1}\mathbf{I}_{K_l})$$
2) Her j kalemi için,
   a) Temiz girdi çek,
   $$\mathbf{X}_{c,j*} \sim \mathcal{N}(\mathbf{X}_{L,j*}, \lambda_n^{-1}\mathbf{I}_J)$$
   b) Bir gizli bağıl konum vektörü çek,
   $\epsilon_j \sim \mathcal{N}(\mathbf{0}, \lambda_v^{-1}\mathbf{I}_J)$
   ve gizli kalem vektörünü şu şekilde hesapla,
   $$\mathbf{v}_j = \epsilon_j + \mathbf{X}^T_{\frac{L}{2},j*}$$
   c) Her kullanıcı i için gizli kullanıcı vektörü çek,
   $$\mathbf{u}_i \sim \mathcal{N}(\mathbf{0}, \lambda_u^{-1}\mathbf{I}_K)$$

d) Her (i, j) kullanıcı kalem eşleniği için $\mathbf{R}_{i,j}$ derecelendirmesi çek,

$$\mathbf{R}_{i,j} \sim \mathcal{N}(\mathbf{u}_i^T \mathbf{v}_j, \mathbf{C}_{ij}^{-1})$$

Burada λ_u, λ_v, λ_w, λ_s ve λ_n hiperparametredir. I kullanıcının J adet kalemle ilgili puanlaması IxJ boyutlarındaki $\mathbf{R}$ matrisiyle gösterilmektedir. Kullanılan veri setinde $\mathbf{R}_{ij}$ ikili değerlerden oluşmaktadır. Eğer i kullanıcısının listesinde j makalesi varsa $\mathbf{R}_{ij} = 1$, değilse 0'dır. $\mathbf{X}_o$, $\mathbf{X}_c$matrisinin İGGO'ya girdi olarak verilen gürültülendirilmiş versiyonudur. $\mathbf{X}_l$, İGGO'nun l. katmanın JxK_l boyutlu çıktı matrisidir. L, toplam katman sayısını belirtmektedir. Özellik çıkarılması ortadaki katmandan, $L/2$. katmandan yapıldığı için L katmanlı bir İGGO'ya $L/2$ seviye İGGO denilmektedir. $\mathbf{W}_l$ ve $\mathbf{b}_l$, ağırlıkları ifade etmektedir. $\mathbf{W}_+$ ise tüm katmanların tüm ağırlık değerlerini ifade etmektedir. $\mathbf{C}_{ij}$ [4]'da belirtildiği gibi güvenirlilik parametresidir (eğer $\mathbf{R}_{ij} = 1$ $\mathbf{C}_{ij} = a$, değilse $\mathbf{C}_{ij} = b$'dir). Burada $L/2$. katman derecelendirme ve içerik bilgisi arasında bağlantıyı sağlamaktadır. Hem özellik yapılanması derecelendirme bilgisinden yararlanabilmekte, hem de derecelendirmede içerik bilgisi kullanılabilmektedir. Hesaplamanın kolaylaştırılması için λ_s sonsuz olarak alınmıştır (Gauss dağılımı Dirac Delta dağılımı haline gelmektedir). Modelin grafik gösterimi simgeler basitleştirilerek Şekil 2'de verilmiştir.

C. Maksimum Sonsal Kestirimleri

Modelin hesaplamasında beklenti-enbüyütme tarzında maksimum sonsal kestirimi yapılmıştır. [4]'te olduğu gibi sonsal dağılımı maksimize etmek λ_u, λ_v, λ_w, λ_s ve λ_n hiperparametreleri verildiğinde $\mathbf{U}, \mathbf{V}$,$\{\mathbf{X}_l\}$, $\mathbf{X}_c$, $\{\mathbf{W}_l\}$,$\{\mathbf{b}_l\}$,$\mathbf{R}$ parametrelerinin logaritmik olabilirliğini maksimize etmeye eşit olacaktır. [2]'de tam ifadesi verilen ve λ_s sonsuza gittiğindeki olabilirlik şu şekildedir:

$$\begin{aligned}\mathcal{L} = &-\frac{\lambda_u}{2}\sum_i \|\mathbf{u}_i\|_2^2 - \frac{\lambda_w}{2}\sum_l (\|\mathbf{W}_l\|_F^2 + \|\mathbf{b}_l\|_2^2) \\ &- \frac{\lambda_v}{2}\sum_j \|\mathbf{v}_j - f_e(\mathbf{X}_{0,j*}, \mathbf{W}^+)^T\|_2^2 \\ &- \frac{\lambda_n}{2}\sum_j \|f_r(\mathbf{X}_{0,j*}, \mathbf{W}^+) - \mathbf{X}_{c,j*}\|_2^2 \\ &- \sum_{i,j} \frac{\mathbf{C}_{ij}}{2}(\mathbf{R}_{ij} - \mathbf{u}_i^T \mathbf{v}_j)^2\end{aligned}$$

Burada belirtilen $f_e(.)$ fonksiyonu gürütülendirilmiş içerik bilgisini alıp kodlanmış halini veren $L/2$. katmanın çıktısını ifade etmektedir. $f_r(.)$ ise kodlanıp tekrar oluşturulan içerik bilgisi çıktısının fonksiyonudur, yani son katmanın çıktısıdır.

[4] ve [18]'deki gibi mevcut $\mathbf{W}^+$ için $\mathbf{u}_i$ ve $\mathbf{v}_j$ değerlerini $\mathcal{L}$'den koordinat tırmanışı yöntemiyle güncelleyebiliriz:

$$\begin{aligned}\mathbf{u}_i &\leftarrow (\mathbf{V}\mathbf{C}_i\mathbf{V}^T + \lambda_u \mathbf{I}_K)^{-1}\mathbf{V}\mathbf{C}_i\mathbf{R}_i \\ \mathbf{v}_j &\leftarrow (\mathbf{U}\mathbf{C}_j\mathbf{U}^T + \lambda_v \mathbf{I}_K)^{-1}(\mathbf{U}\mathbf{C}_j\mathbf{R}_j + \lambda_v f_e(\mathbf{X}_{0,j*}, \mathbf{W}^+)^T)\end{aligned}$$

$\mathbf{U}$ ve $\mathbf{V}$ tüm kullanıcı ve kalemleri ifade eden matrisleri, $\mathbf{C}_i = diag(\mathbf{C}_{i1}, ..., \mathbf{C}_{iJ})$ güvenilirliği ifade eden çapraz matrisi, $\mathbf{R}_i$ ve $\mathbf{R}_j$ i. kullanıcının tüm kalemler için veya j. kalemin tüm kullanıcılar için derecelendirme bilgilerini gösteren vektörleri ifade etmektedir.

Hiperparametre değerleri ile ilgili iki uç nokta bulunmaktadır. Bunlardan birisi λ_n/λ_v oranının pozitif sonsuza yaklaştığı zamandır. Bu durumda, bölümlerin birbirini beslemesi kesilmekte ve İGGO'da öğrenilen özellikler doğrudan modelde kullanılmaktadır. Diğer durumda, λ_n/λ_v oranının sıfıra yaklaştığında, İGGO'nun kod çözücü kısmı ortadan kalkmaktadır. Her iki durumda da performans ciddi oranda azalmaktadır [2].

$\mathbf{U}$ ve $\mathbf{V}$ verildiğinde her katman için $\mathbf{W}_l$ ve $\mathbf{b}_l$ değerlerini geri yayılımla öğrenebiliriz. $\mathbf{U}, \mathbf{V}, \mathbf{W}_l$ ve $\mathbf{b}_l$ değerlerini sırayla dönüşümlü olarak güncellediğimizde olabilirliğin yerel optimum noktasına ulaşabiliriz. Olabilirliğin $\mathbf{W}_l$'e göre gradyanı aşağıda verilmiştir.

$$\begin{aligned}\nabla_{\mathbf{W}_l}\mathcal{L} = &-\lambda_w \mathbf{W}_l \\ &- \lambda_v \sum_j \nabla_{\mathbf{W}_l} f_e(\mathbf{X}_{0,j*}, \mathbf{W}^+)^T (f_e(\mathbf{X}_{0,j*}, \mathbf{W}^+)^T - \mathbf{v}_j) \\ &- \lambda_n \sum_j \nabla_{\mathbf{W}_l} f_r(\mathbf{X}_{0,j*}, \mathbf{W}^+)(f_r(\mathbf{X}_{0,j*}, \mathbf{W}^+) - \mathbf{X}_{c,j*})\end{aligned}$$

D. Tahmin

D, gözlemlenen test verisi olmak üzere, [4]'da yapılana benzer şekilde derecelendirmeyi şu şekilde tahmin edebiliriz:

$$E[\mathbf{R}_{ij}|D] \approx E[\mathbf{u}_i|D]^T (E[(f_e(\mathbf{X}_{0,j*}, \mathbf{W}^+)^T|D]) + E[\epsilon_j|D])$$

Tahmin edilen derecelendirmeyi şu şekilde hesaplayabiliriz:

$$\mathbf{R}_{ij}^* \approx (\mathbf{u}_j^*)^T (f_e(\mathbf{X}_{0,j*}, \mathbf{W}^{+*})^T + \epsilon_j^*) = (\mathbf{u}_i^*)^T \mathbf{v}_j^*$$

Burada j. kalem için henüz herhangi bir derecelendirme yapılmamışsa ϵ_j^* 0 olacaktır.

III. Model-Bayesci Negatif Olmayan Matris Ayrıştırması

Negatif olmayan matris ayrıştırmasında verilen IxJ boyutlu $\mathbf{R} = \{R_{i,j}\}$ matrisi, kalemlerin içerik bilgisi olan $\mathbf{X}$ ile birleştirilerek oluşturan $\mathbf{T}$ matrisinin ayrıştırması araştırılmıştır. $\mathbf{T} = \mathbf{UV}$ olacak şekilde U ve V pozitif matrisleri araştırılır. Bu makelede kullanılan model ve çıkarım yönteminde [1] baz alınmıştır.

Üretici model şu şekildedir:

$$\begin{aligned}u_{i,k} \sim \mathcal{G}(u_{i,k}; a_{i,k}^u, \frac{b_{i,k}^u}{a_{i,k}^u}), v_{k,j} \sim \mathcal{G}(v_{k,j}; a_{k,j}^v, \frac{b_{k,j}^v}{a_{k,j}^v}), \\ s_{i,k,j} \sim \mathcal{PO}(s_{i,k,j}; u_{i,k}v_{k,j}), t_{i,j} = \sum_k s_{i,k,j}.\end{aligned}$$

Burada $\mathcal{G}(x; a, b)$ şekli a, oranı b olan x rastgele değişkeninin Gamma dağılımını, $\mathcal{PO}(s, \lambda)$ ise λ negatif olmayan yeğinlik parametresiyle $s \in \mathbb{N}_0$ rastgele değişkeninin Poisson dağılımını ifade etmektedir.

Problemin çözümünde bu modelin değişimsel Bayes ile çıkarım yöntemi kullanılmıştır. İDÖ'de olduğu gibi güvenilirlik değerleri ve verilerin değişkenlere yapacağı etkinin ayarlanması maskeleme matrisi üzerinden sağlanmıştır.

IV. Deneyler

Bu bölümde, belirtilen modellerin CiteULike-a verisiyle yapılan deney sonuçları gösterilmektedir. Bu veri setinde 5551 kullanıcı, 16980 makale bulunmaktadır. Setin derecelendirme yoğunluğu %0.22'dir.

Kalemlerinin içerik bilgisinde [4]'daki gibi S kelimeden oluşan sözlük ile terim frekansı-ters metin frekansı üzerinden kelime çantası gösterimi kullanılmıştır. Makalelerin başlıkları ve özetleri baz alınmış, etkisiz sözcükler çıkarıldıktan sonra sözlük oluşturulmuştur. S değeri citeulike-a verisi için 8000'dir. Eğitim seti için veri tabanında her bir kullanıcı için derecelendirme yaptığı rastgele P kalem seçilmiş ve kalanı da test kümesi olarak kullanılmıştır.

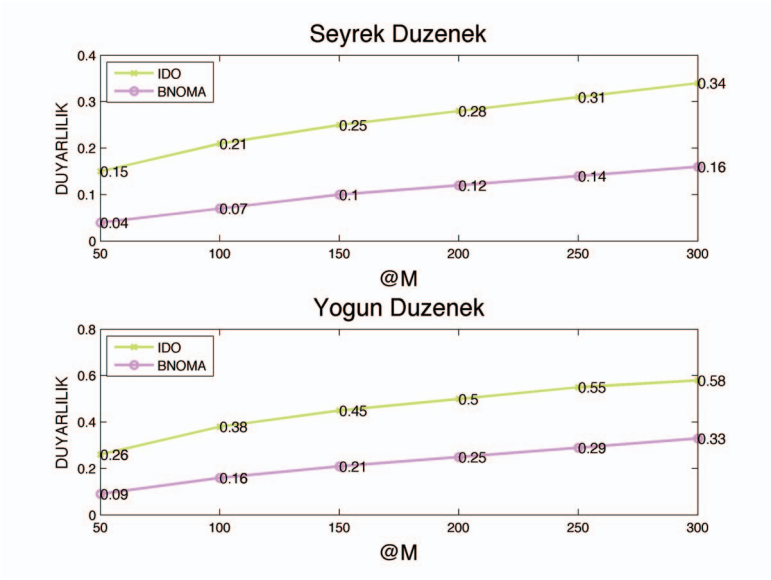
Hem İDÖ hem de BNOMA'da sonucu veren hiperparametrelerin bulunması için grid araması yöntemi kullanılmıştır. İDÖ'de ilk etapta içerik bilgisi için önöğrenme yapılmış ve İDÖ'nün öğrenmesinde bu öğrenilen ağ ile sürece başlanmıştır. BNOMA'nın öğrenmesinde değişimsel Bayes kullanılmıştır. Tüm parametreler bağımsız olarak ele alınmış ve döngülerde hiperparametrelerin güncellenmesi sağlanmıştır. Klasik negatif olmayan matris ayrıştırmasında ise toplama güncellemesi kullanılmıştır.

Derecelendirme [8]'da olduğu gibi sadece pozitif bilgi üzerinden sağlanmaktadır. Bir kişinin bir kalemle ilgili derecelendirmesinin boş olması kişinin o kalemle ilgilenmediğini de, o kalemi hiç görmediğini de gösteriyor olabilir. Bu nedenle deney sonuçlarının değerlendirilmesinde kesinlik oranından ziyade duyarlılık oranı kullanılmıştır.

Duyarlılık perfomansı şu şekilde hesaplanmaktadır:

$$\text{duyarlılık@M} = \frac{\text{İlk M kalemden kişinin seçtiklerinin sayısı}}{\text{Kişinin seçtiği kalem sayısı}}$$

"



Şekil 3: İki farklı konfigürasyonda İDÖ ve BNOMA duyarlılık@M deney sonuçları

V. Sonuçlar ve Vargılar

Eğitim seti, seyrek ve yoğun olmak üzere farklı iki düzenekte hazırlanmıştır. Seyrek düzenekte P sayısı 1 olarak hazırlanmış, yoğun düzenekte ise P sayısı 10 olarak alınmıştır.

Şekil-3'te belirtilen modellerin farklı n değerleri için duyarlılık sonuçları verilmiştir. İDÖ hem seyrek hem de yoğun düzenekte diğer modellerden daha iyi sonuç almıştır.

VI. Gelecek Çalışmalar

Deney sonuçları da göstermiştir ki, birleşik model işbirlikçi filtrelemeye göre daha iyi sonuçlar vermiştir. İçerik bilgisinin sinir ağları aracılığıyla ifade edilmesiyle başarı artmaktadır.

İleride yapılabilecek çalışmalarda içeriğin [14]'da olduğu gibi kelimelerin sırasını da dahil eden farklı yapay sinir ağları yöntemleriyle temsil edilmesi bu tür modellerin başarısını arttırabilir. Nitekim bir örnek olarak [7]'da gürbüz yinelenen ağlar kullanılarak performans artışı sağlanmıştır. Ayrıca bir sonraki çalışmada İDÖ'de kullanılan olasılıksal matris ayrıştırması yerine negatif olmayan matris ayrıştırması kullanılmasıyla ilgili bir çalışma yapılabilir.

Kaynakça

[1] A.T. Cemgil. "Bayesian inference for nonnegative matrix factorisation models." Computational Intelligence and Neuroscience 2009 (2009).

[2] H. Wang, N. Wang, and D. Yeung. Collaborative deep learning for recommender systems. In KDD, 2015.

[3] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. JMLR, 11:3371–3408, 2010.

[4] C. Wang and D. M. Blei. Collaborative topic modeling for recommending scientific articles. In KDD, pages 448–456, 2011.

[5] F. Ricci, L. Rokach, and B. Shapira. Introduction to Recommender Systems Handbook. Springer, 2011.

[6] R. Salakhutdinov and A. Mnih. Probabilistic matrix factorization. In NIPS, 2007.

[7] H. Wang, X. Shi, and D.Y. Yeung. Collaborative recurrent autoencoder: recommend while learning to fill in the blanks. In NIPS, 2016.

[8] H. Wang, B. Chen, and W.-J. Li. Collaborative topic regression with social regularization for tag recommendation. In IJCAI, 2013.

[9] R. Salakhutdinov, A. Mnih, and G. E. Hinton. Restricted Boltzmann machines for collaborative filtering. In ICML, 2007.

[10] K. Georgiev and P. Nakov. A non-iid framework for collaborative filtering with restricted Boltzmann machines. In ICML, 2013.

[11] B. Hidasi, A. Karatzoglou, L. Baltrunas, and D. Tikk. Session-based recommendations with recurrent neural networks. arXiv preprint arXiv:1511.06939, 2015.

[12] A. V. D. Oord, S. Dieleman, and B. Schrauwen. Deep content-based music recommendation. In NIPS, 2013.

[13] X. Wang and Y. Wang. Improving content-based and hybrid music recommendation using deep learning. In ACM MM, 2014.

[14] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In NIPS, pages 3111–3119, 2013.

[15] S. Rendle, C. Freudenthaler, Z. Gantner, and L. Schmidt-Thieme. BPR: Bayesian personalized ranking from implicit feedback. In UAI, pages 452–461, 2009.

[16] K. Lang. Newsweeder: Learning to filter netnews. In ICML, pages 331–339, 1995.

[17] J. Bobadilla, F. Ortega, A. Hernando, and A. Gutierrez. Recommender systems survey. Knowledge Based Systems, 46:109–132, 2013.

[18] Y. Hu, Y. Koren, and C. Volinsky. Collaborative filtering for implicit feedback datasets. In ICDM, pages 263–272, 2008.