

Bağlantı Tahmini İçin Bağlaşımlı Simetrik Tensör Ayrışım Modelleri

Generalized Coupled Symmetric Tensor Factorization for Link Prediction

Beyza Ermiş, A. Taylan Cemgil
Bilgisayar Mühendisliği Bölümü
Boğaziçi Üniversitesi
İstanbul, Türkiye

Email: beyza.ermis,taylan.cemgil@boun.edu.tr

Evrin Acar
Yaşam Bilimleri Bölümü
Kopenhag Üniversitesi
Kopenhag, Danimarka
Email: evrim@life.ku.dk

Özetçe —Bu çalışmada gözlemlenemeyen bağlantı tahmini problemi için Genelleştirilmiş Bağlaşımlı Tensör Ayrışımı (GBTA) çerçevesinde tanımlanmış modeller sunulmaktadır. GBTA ortak tensörler içeren modellerin eşzamanlı ayrışımı ile ortak saklı faktörler çıkarabilen bir algoritmik çerçevedir. Bu bildiride GBTA çerçevesine ek olarak simetrik yapıdaki matrislerin ayrışımı için kullanılan algoritma sağlanmaktadır. Heterojen verilerin ayrışımında kullanılan önceki çalışmalar ya tek bir iraksaya ya da belirli bir tensör ayrışım modeline odaklanmaktadır; ancak, heterojen veri analizinde temel zorluklardan biri doğru tensör modelini ve iraksayı bulmaktır. Bu nedenle, bu çalışmada farklı tensör modelleri ve iraksaylar ele alınmaktadır. Gerçek veri kümeleri üzerinde gerçekleştirilen deneyler birden fazla kaynaktan gelen verilerin bağlaşımlı tensör ayrışım yöntemi ile ortak analizinin ve simetrik yapıdaki verilerin bağlaşımlı modellere dahil edilmesinin bağlantı tahmin performansını artırmakta olduğunu; ayrıca doğru iraksay ve tensör model seçiminin önemini göstermektedir.

Anahtar Kelimeler—Bağlaşımlı tensör ayrışımı; Bağlantı Tahmini; Eksik veri; Veri Birleştirme; Simetrik Matris.

Abstract—This study deals with the *missing link prediction*, the problem of predicting the existence of missing connections between entities of interest. Link prediction is addressed using coupled analysis of relational datasets represented by several matrices, including symmetric ones and multiway arrays, that will be simply called *tensors*. We propose to use an approach based on probabilistic interpretation of tensor factorisation models, i.e., Generalised Coupled Tensor Factorisation (GCTF), which can simultaneously fit a large class of tensor models to higher-order tensors/matrices with common latent factors using *different loss functions*. In addition, we propose the algorithm for factorization of symmetric matrices. Numerical experiments demonstrate that joint analysis of data from multiple sources via coupled factorisation and integration of symmetric matrices to models improves the link prediction performance and the selection of right loss function and tensor model is crucial for accurately predicting missing links.

Keywords—Coupled tensor factorization, Link prediction, Missing data, Data fusion, Symmetric Matrix.

I. GİRİŞ

Günümüzde, teknolojik ilerlemeler ilişkisel verileri de beraberinde getirmektedir. Örneğin, satış öneri sistemlerinde hangi kullanıcının hangi ürünü almış olduğu bilgisine ek olarak kullanıcıların sosyal ağlardaki arkadaşlık verilerine de

ulaşabilmek mümkündür. Görüldüğü gibi, sosyal ağ verisi bağlantı tahmin problemleri için değerli yan bilgi sağlayabilir ve birçok kaynaktan gelen verileri bir arada analiz etmek tahminin doğruluğunu arttırabilmektedir. Bu çalışmada ilişkisel veri modelleme için bağlantı tahmini problemi ele alınmıştır.

Sosyal ağ analizi ve öneri sistemleri gibi birçok alandaki uygulamalar, bağlantı tahmini problemini iki nesne arasında bir bağlantının varlığı veya yokluğu sonucuna varılması problemi olarak incelemektedir. Örneğin, satış öneri sistemlerinde, bir kullanıcı bir ürünü satın aldığı anda, bu kullanıcı ve ürün arasında bir bağlantı oluştuğu kabul edilir. Bu durumda, kullanıcının ilgilenebileceği düşünülen diğer ürünlerin kullanıcıya önerilmesi gözlemlenmemiş bağlantı tahmini problemi olarak adlandırılır. Ancak, bağlantı tahmini tek bir veriye bağlı olarak yapılırsa, sonucun gerçeğe yakın çıkması muhtemeldir. Bu tarz veri kümelerinde, verinin boyutu büyük olmasına rağmen veriler oldukça seyrek ve resmin tamamının yalnızca çok küçük bir bölümü temsil edilmektedir [1]. Bu nedenle, farklı kaynaklardan gelen ilişkili veriler bağlantı tahmini modellerine dahil edilmektedir [2].

Bağlantı tahmini modellerinde yan bilgiyi ek ilişkisel veri olarak içermenin en etkili yollarından biri, farklı verileri matris kümesi şeklinde göstermektir. Daha sonra, bu matrisler kümesi toplu matris ayrışımı yöntemi kullanılarak bir arada analiz edilebilir [3]. Ancak çoğu uygulamada, matrisler birden çok özneliği gösterebilmede yeterli değildir ve bu nedenle tensör ayrışım modellerine ihtiyaç duyulmaktadır.

Bu çalışmada, bağlantı tahmini problemini matris ve tensör şeklinde yapılmış, simetrik matrisleri de içeren verikümelerini bağlaşımlı analiz ederek ele almaktayız. Bunun için Öklid ve KL iraksayları ile birlikte CANDECOMP/PARAFAC (CP) [4] ve Tucker [5] tensor ayrışım yöntemlerini Genel Bağlaşımlı Tensör Ayrışımı (GBTA) [6] çerçevesi dahilinde kullanılmaktayız. Ayrıca, simetrik matris ayrışımı için kullanılacak güncelleme denklemini de sağlayarak, simetrik matrislerin kullanılmasının bağlantı tahmin performansı üzerindeki etkisini önceki çalışmamızla [7] karşılaştırarak değerlendirmekteyiz.

II. GENELLEŞTİRİLMİŞ BAĞLAŞIMLI TENSÖR AYRIŞIMI

GBTA çerçevesi, Olasılıksal Saklı Tensör Ayrışımı (OSTA) [8] çerçevesinin bağlaşımlı tensör ayrışımı için genişletilmesiyle oluşturulmuş bir çerçevedir. OSTA

çerçevesinde gözlemlenen X tensörü, $Z_\alpha (\alpha = 1, \dots, |\alpha|)$ bileşenlerinin çarpımı cinsinden, aşağıdaki gibi tanımlanmıştır:

$$X(v_0) \approx \hat{X}(v_0) = \sum_{\bar{v}_0} \prod_{\alpha} Z_\alpha(v_\alpha) \quad (1)$$

Burada, X gözlemlenen tensör, \hat{X} model tarafından oluşturulan yaklaşık tensör, Z_α ise tensörü oluşturduğu varsayılan bileşenlerdir (Daha detaylı anlatımı için bkz. [8]).

GBTA çerçevesi OSTA çerçevesini bir adım ileriye taşıyarak birden fazla gözlemlenen tensörün ($X_\nu, \nu : 1 \dots |\nu|$) bulunduğu modellerde eş zamanlı kestirim yapmaya olanak sağlamaktadır. Gözlemlenen her tensör $X_\nu, v_{0,\nu}$ ile gösterilen bir indis kümesine sahiptir ve gözlemlenen tensörlerle bileşenlerin arasındaki ilişki $|\nu| \times |\alpha|$ boyutlu R matrisi ile ilişkilendirilmektedir. Bu matris aşağıdaki gibi tanımlanmaktadır:

$$R^{\nu,\alpha} = \begin{cases} 1 & X_\nu \text{ ve } Z_\alpha \text{ bağılı ise} \\ 0 & X_\nu \text{ ve } Z_\alpha \text{ bağılı değil ise} \end{cases} \quad (2)$$

Daha sonra GBTA çerçevesinde gözlemlenen X_ν tensörleri, Z_α bileşenlerinin çarpımı cinsinden, aşağıdaki gibi tanımlanabilir:

$$X_\nu(v_{0,\nu}) \approx \hat{X}_\nu(v_{0,\nu}) = \sum_{\bar{v}_{0,\nu}} \prod_{\alpha} Z_\alpha(v_\alpha)^{R^{\nu,\alpha}} \quad (3)$$

Gözlemlerin ve bileşenlerin negatif olmadığı modellerde, X_ν tensörlerini gözlemledikten sonra Z_α bileşenlerin kestirimi için gerekli algoritma [6]'da sunulmuştur. Bu algoritma için (4)'te tanımlanan güncelleme denklemi yinelemeli olarak uygulanmalıdır.

$$Z_\alpha \leftarrow Z_\alpha \circ \frac{\sum_{\nu} R^{\nu,\alpha} \Delta_{\alpha,\nu}(M_\nu \circ \hat{X}_\nu^{-p} \circ X_\nu)}{\sum_{\nu} R^{\nu,\alpha} \Delta_{\alpha,\nu}(M_\nu \circ \hat{X}_\nu^{1-p})} \quad (4)$$

Bu denklemde p kullanılacak iraksayı belirlemekte olup $p = 0, 1$ değerleri sırasıyla β iraksayının özel durumları olan Öklid ve Kullback-Leibler iraksaylarına denk gelmektedir. Ayrıca, M_ν, X_ν 'nin değerine göre tanımlanmış ikili maskedir ve X_ν tensöründeki gözlemlenmiş elemanlar için değeri 1'e, gözlemlenemeyen elemanlar için de 0'a eşittir.

$$\Delta_{\alpha,\nu}(A) = \sum_{v_{0,\nu} \cap \bar{v}_\alpha} A(v_{0,\nu}) \prod_{\alpha' \neq \alpha} Z_{\alpha'}(v_{\alpha'})^{R^{\nu,\alpha'}} \quad (5)$$

(5)'te verilen Δ_α fonksiyonu gözlemlenen indisler üzerinde tanımlanmaktadır ve argümanı A tensörüdür. (4)'da verilen güncelleme denklemleri hesaplanırken $\Delta(\cdot)$ fonksiyonu $A = M_\nu \circ \hat{X}_\nu^{-p} \circ X_\nu$ ve $A = M_\nu \circ \hat{X}_\nu^{1-p}$ için hesaplanmaktadır.

Simetrik Matris Ayırışımı: Bu bölümde simetrik yapıya sahip matrislerin GCTF çerçevesi ile ayırışımının nasıl gerçekleştirilebileceğini göstereceğiz. Simetrik yapıdaki matrisler bileşenlerinin çarpımı cinsinden aşağıdaki gibi tanımlanmaktadır:

$$X(i_1, i_2) \approx \hat{X}(i_1, i_2) = \sum_r Z(i_1, r)Z(i_2, r) \quad (6)$$

Görüldüğü gibi X matrisine aynı bileşenin çarpımı şeklinde yaklaşmaktadır. Simetrik matrisi oluşturan bu bileşene ulaşmak için yinelemeli olarak uygulanması gereken güncelleme denklemi aşağıda verilmektedir.

$$Z \leftarrow Z \circ \frac{(X/\hat{X}^p)Z}{\hat{X}^{1-p}Z} \quad (7)$$

Ancak, bu çalışmada amacımız (7)'de verilen denklemin GCTF çerçevesi ile çözümünün sağlanabilmesi ve simetrik matrislerin bağımlı modellere dahil edilebilmesidir. Bunun için R matrisinde, modeldeki her bir indis için bir sütun tanımlanmalıdır. Bu da simetrik matrisi oluşturan bileşenler için R matrisinin iki farklı indis içermesi anlamına gelmektedir. Böylece (4)'te verilen güncelleme denkleminin simetrik matrislerin ayırışımında da kullanılabilmesi sağlanmaktadır. Bölüm III'te verilen modeller, simetrik matrislerin GCTF ile ayırışımını daha iyi açıklamaktadır.

Ayrıca, bazı deneylerimizde simetrik matrisleri oluşturan bileşenler için kullandığımız güncelleme denkleminin sayısal olarak sabit sonuç vermediğini ve yakınsamayı engelleyen salınımlı davranış göstermekte olduğunu gördük. Salınımlı davranışı engellemek amacıyla, bu bileşenler için, $\epsilon = 0.02$ iken (4)'ün aşağıda verilen sönümlendirilmiş versiyonunu kullandık.

$$Z_\alpha \leftarrow Z_\alpha \circ \left(\frac{\sum_{\nu} R^{\nu,\alpha} \Delta_{\alpha,\nu}(M_\nu \circ \hat{X}_\nu^{-p} \circ X_\nu)}{\sum_{\nu} R^{\nu,\alpha} \Delta_{\alpha,\nu}(M_\nu \circ \hat{X}_\nu^{1-p})} \right)^{1-\epsilon} \quad (8)$$

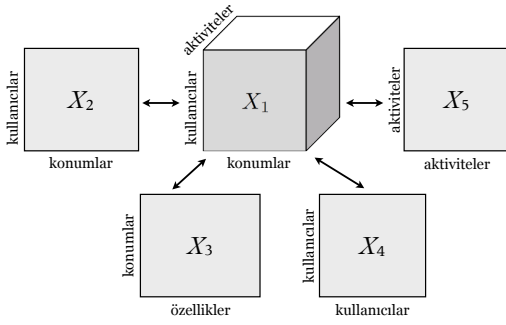
III. BAĞLANTI TAHMİNİ İÇİN TENSÖR AYRIŞIM MODELLERİ

Bu bölümde bağlantı tahmini problemi için farklı bağımlı tensör ayırışımı modelleri ve iraksaylar kullanarak çözüm sunuyoruz. Bu modellerdeki temel amacımız gözlemlenen elemanları 0 ve 1 lerden oluşan (1-varlığı bilinen bağlantı ve 0-yokluğu bilinen bağlantı olmak üzere) gözlem tensöründeki (X_1) gözlemlenmemiş bağlantıların yan bilgi sağlayan tensör ve matrisleri (X_2, X_3, X_4, X_5) kullanarak doğru bir şekilde tahmin edilmesidir. Bu zor bir bağlantı tahmini problemidir çünkü X_1 tensöründeki bağlantıların %1'inden daha az bir kısmı gözlemlenmiştir veya tensörün bazı dilimleri tamamen gözlemlenmemiş olabilmektedir. Böyle durumlarda düşük mertebeli tensör ayırışım yöntemleri kayıp bağlantı tahmininde başarısız olmaktadır. GBTA çerçevesini kullanarak, uygulamaya yönelik farklı bağımlı tensör ayırışım modellerini kolaylıkla oluşturabiliriz. Bu modellerin seçimi uygulamaya ve kullandığımız veri kümesine bağlı olduğundan, öncelikle kullanacağımız verikümesini kısaca açıklayacağız.

UCLAF verikümesi¹ [9] GPS verisinden elde edilmiş kullanıcı, konum ve aktivite bilgilerinden oluşmaktadır. İlk olarak, kullanıcı-konum-aktivite üçlemeleri arasındaki ilişkiler X_1 tensörünü oluşturmada kullanılır. Bu tensörün bir elemanı $X_1(i, j, k)$ i kullanıcısının j konumunu ziyaret edip orada k aktivitesini gerçekleştirip gerçekleştirmediğini göstermektedir. GPS verisinden alınan bilgiler 168 anlamlı konum olarak kümelenebilir ve kullanıcıların bu konumlarda yaptıkları yorumlar 5 farklı aktivite (yemek, alışveriş, film, spor, eğlence) olarak değerlendirilmiştir. Sonuç olarak bu verikümesi 164 kullanıcı, 168 konum ve 5 aktiviteden oluşturulmuştur (Daha detaylı anlatımı için bkz. [9]). Ayrıca toplanan GPS verisi kullanıcı-konum tercihleri, konum-özellik bilgisi, kullanıcı-kullanıcı benzerliği ve aktivite-aktivite bağıntısı ek bilgilerini de içermektedir. Bu ek bilgiler sırasıyla X_2, X_3 ve simetrik yapıda X_4, X_5 matrisleri ile gösterilmektedir.

Bu modelde kullanıcı-konum tercih matrisi I kullanıcı sayısı ve J konum sayısı olmak üzere $I \times J$ boyutuna

¹<http://www.cse.ust.hk/~vincentz/aaai10.uclaf.data.mat>



Şekil 1: UCLAF verikümesi

sahiptir ve konum indisini göstermek için j yerine m indisini kullandığımızdan $X_2(i, m)$ elemanlarını içermektedir. Bunun nedeni X_1 ve X_2 nin farklı konum değerlerini ölçmesidir: X_1 i kullanıcısının j konumunda gerçekleştirdiği aktiviteyi gösterirken, X_2 i kullanıcısının m konumuna ziyaret etme sıklığını ve orada belli bir zamandan fazlasını geçirdiğini göstermektedir. X_1 ve X_2 verilerindeki konum indisleri, j ve m arasındaki ilişki ortak kullanıcı bileşeni üzerinden sağlanmaktadır. Bunlara ek olarak, her bir özellik değerinin, $X_3(j, n)$, j konumundaki ilgi çekici nokta (müze, restoran, ... vs) oranına karşılık geldiği konum-özellik matrisi bulunmaktadır. $J \times N$ boyutlu bu matrisi kullanarak konumların benzerlikleri hakkında bilgi edinebiliriz. Son olarak, kullanıcı-kullanıcı benzerlikleri ile $I \times I$ boyutlu X_4 matrisini ve aktivite-aktivite bağıntıları ile de $K \times K$ boyutlu X_5 matrisini oluşturuyoruz. Bu verikümesinde 18 kullanıcı hiçbir konum ve aktivite bilgisine sahip değildir. Bu nedenle modellerde diğer 146 kullanıcının bilgilerini kullanmaktayız. Sonuç olarak, deneylerimizde kullanıcı sayısı $I = 146$, konum sayısı $J = 168$, aktivite sayısı $K = 5$ ve konum özellik sayısı $N = 14$ tür.

X_1 tensöründeki gözlemlenmemiş bağlantıları tahmin etmek için iki farklı bağımlı tensör modeli sunuyoruz. Her iki modelde de negatif olmayan ayrışım problemlerimiz için Öklid ve KL ıraksaylarını kullanmaktayız. İlk ayrışım modelimizde, X_1 tensörünü X_2 , X_3 ve simetrik X_4 , X_5 matrisleri ile analiz ederek bağımlı ayrışım yaklaşımını CP tensör ayrışım yöntemine uygulamaktayız. Bu yaklaşım bize aşağıdaki modeli vermektedir:

$$\hat{X}_1(i_1, j, k_1) = \sum_r A(i_1, r)B(j, r)C(k_1, r) \quad (9)$$

$$\hat{X}_2(i_1, m) = \sum_r A(i_1, r)D(m, r) \quad (10)$$

$$\hat{X}_3(j, n) = \sum_r B(j, r)E(n, r) \quad (11)$$

$$\hat{X}_4(i_1, i_2) = \sum_r A(i_1, r)A(i_2, r) \quad (12)$$

$$\hat{X}_5(k_1, k_2) = \sum_r C(k_1, r)C(k_2, r) \quad (13)$$

Bu modelde ortak bileşenleri olan beş gözlemlenmiş tensörümüz vardır ve modele ait R matrisi aşağıdaki gibidir:

$$R = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{bmatrix} \quad (14)$$

Modeldeki tensörler beş farklı bileşenden oluşur ancak simetrik matrisleri oluşturan bileşenler için farklı iki indeks tanımlandığından; A için i_1, i_2 ve C için k_1, k_2 olmak üzere, R matrisi yedi sütundan oluşmaktadır. Bu şekilde simetrik X_4 matrisinin yalnızca A bileşeninden ve simetrik X_5 matrisinin yalnızca C bileşeninden oluşması sağlanmakta ve bu bileşenlerinin değerini bulmak için (8)'de verilen güncelleme denklemi kullanılabilir. X_1, X_2 ve $X_4, A(i, r)$ elemanlarından oluşan ortak A bileşenine sahiptir ve A 'nın her bir satırı $A(i, :)$, i kullanıcısının $|r|$ boyutlu 'tercih' alanındaki saklı konumu olarak düşünülebilir. Aynı zamanda $B(j, r)$ elemanlı B bileşeni ise j konumunun aynı tercih alanındaki saklı konumunu göstermektedir. Sonuç olarak i kullanıcısının j konumunda k aktivitesini gerçekleştirmiş olması için $A(i, r)B(j, r)$ değerinin en az bir r için büyük bir değere sahip olması gerekmektedir. Konum bileşeni B ise X_1 dışında konum-özellik matrisi X_3 tarafından da etkilenmektedir.

İkinci ayrışım modelimizde ise, aynı şekilde, bağımlı ayrışım yaklaşımını Tucker tensör ayrışım yöntemine uygulamakta ve aşağıdaki modeli elde etmekteyiz:

$$\hat{X}_1(i_1, j, k_1) = \sum_{p,q,r} A(i_1, p)B(j, q)C(k_1, r)G(p, q, r) \quad (15)$$

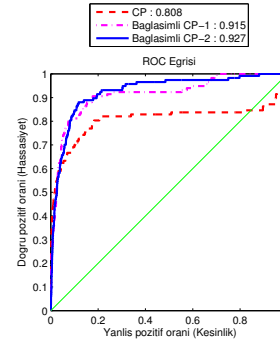
$$\hat{X}_2(i_1, m) = \sum_p A(i_1, p)D(m, p) \quad (16)$$

$$\hat{X}_3(j, n) = \sum_q B(j, q)E(n, q) \quad (17)$$

$$\hat{X}_4(i_1, i_2) = \sum_p A(i_1, p)A(i_2, p) \quad (18)$$

$$\hat{X}_5(k_1, k_2) = \sum_r C(k_1, r)C(k_2, r) \quad (19)$$

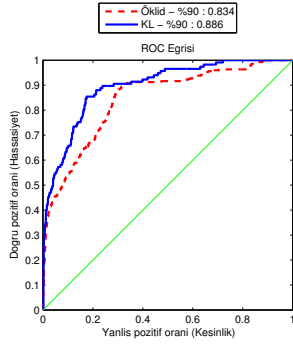
Bu modelde, (9)'den farklı olarak i kullanıcısının j konumunda k aktivitesini gerçekleştirmiş olması $\sum_{p,q} A(i, p)B(j, q)D(p, q, r)$ değeri ile belirlenir ve ilk modele göre daha esnek bir yapıya sahiptir.



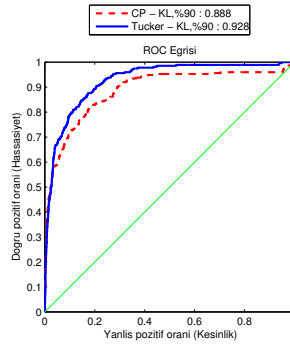
Şekil 2: CP ve Bağımlı(CP) modelleri - Öklid, 80%

IV. SONUÇLAR

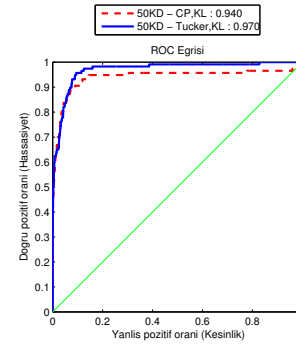
Bu bölümde, modellerimizi değerlendirmek için düzenlediğimiz çeşitli deneyleri ve sonuçlarını sunuyoruz. İlk önce bağımlı modellerin bağlantı tahmini problemdeki üstünlüğünü gösterebilmek için standart CP ve Tucker tensör ayrışım modelleri ile kendi bağımlı modellerimizi karşılaştırdık. Bu deneylerde Öklid ve KL ıraksaylarını kullandık ve sırasıyla verinin $\{40, 60, 80, 90, 95\}$ 'ini rastgele



Şekil 3: Bağlaşımli (CP), Öklid ve KL ıraksayları, 90%



Şekil 4: Bağlaşımli CP ve Tucker - KL, 90%



Şekil 5: Bağlaşımli CP ve Tucker, KL, 50 kayıp dilim

maskeledik. Bütün durumlarda bağlaşımli modeller standart modellerden daha iyi sonuç vermektedir. Şekil 2 Öklid ıraksayını kullandığımız ve verinin %80'inin gözlemlenmemiş olduğu durumda standart ve bağlaşımli CP modellerinin karşılaştırılmasını göstermektedir. Daha önceki çalışmamızda [7] yer alan ilk bağlaşımli CP modelinde (Bağlaşımli CP-1) simetrik yapıya sahip X_4 ve X_5 matrisleri kullanılmamıştır. Görüldüğü gibi bu matrislerin modele dahil edilmesi (Bağlaşımli CP-2) ve özellikle gözlenmemiş verinin yüksek olduğu durumda ek bilginin kullanılması tahmin performansını artırmaktadır. ıraksay seçiminin tahmin performansı üzerindeki etkisini göstermek için farklı gözlemlenmemiş veri miktarlarında bağlaşımli CP ve Tucker modelleri ile deneyler gerçekleştirdik. Bütün durumlarda KL ıraksayının tahmin performansı, Öklid ıraksayından yüksek sonuçlar vermektedir. Şekil 3 verinin %90'ının gözlemlenmediği durumda Öklid ve KL ıraksaylarının bağlaşımli CP ve Tucker modelleri üzerindeki performanslarını göstermektedir. Son olarak, tensör ayrışım modelinin seçiminin tahmin performansı üzerindeki etkisini göstermek için bağlaşımli CP ve Tucker modellerinin performanslarını karşılaştırmaktayız. Şekil 4 Tucker modelinin bu verikümesi için CP modelinden daha uygun olduğunu göstermektedir. Bunun nedeni ise Tucker modelinin CP modeline göre daha esnek bir yapıya sahip olması ve verinin yapısal bilgisini ortaya çıkarmada daha başarılı olmasıdır.

Körleme Başlangıç Problemi: Bağlantı tahmininde sıkça karşılaşılan problemlerden biri de yeni bir kullanıcının uygulamaya dahil olmasıdır. Bu durumda X_1 tensöründe yeni kullanıcıya ait hiçbir bilgi yoktur ve bu da X_1 tensöründe gözlemlenmemiş bir dilim oluşmasına neden olur. Bu matris ayrışım probleminde tamamı gözlemlenmemiş satır veya sütunlara denk gelmektedir. Düşük mertebeli tensör ayrışım modelleri bu problemi çözememektedir. Bağlaşımli modellerimiz X_1 tensöründe herhangi bir kullanıcıya ait hiçbir girdi bulunmadığı durumlarda bile ek bilgilerin yardımı ile bağlantı tahminini gerçekleştirebilmektedir. Şekil 5, 50 kullanıcının bilgilerinin gözlemlenmemiş olduğu durumda bağlaşımli CP ve Tucker modellerinin KL ıraksayı ile gerçekleştirdiği performansı göstermektedir.

Bu çalışmada ulaştığımız sonuçları önceki çalışmamızla [7] kıyasladığımızda, tüm deneysel düzenlemelerde tahmin performansının arttığını görmekteyiz.

V. VARGILAR

Bu çalışmada matris ve tensör şeklinde gösterilebilen ilişkili verilerde bağlantı tahmini problemi için bağlaşımli tensör ayrışımı yöntemlerine ek olarak simetrik matrislerin ayrışımı için de yeni bir matris ayrışım yöntemi sunulmuştur. Bu konuda varolan çalışmalar ya sadece tek bir tensör ayrışım yöntemi ya da tek bir ıraksay kullanılarak geliştirilmiştir. Ancak, modellerin GBTA çerçevesi ile tanımlanması sayesinde farklı tensör modelleri, simetrik matrisler ve ıraksaylar çoklu verikümelerinin ayrışımında kolaylıkla kullanılabilir. Gerçek GPS verisi üzerinde gerçekleştirmiş olduğumuz deneyler bağlaşımli modellerin ve simetrik matris kullanımının bağlantı tahmin performansını artırdığını; tensör ayrışım yöntemi ve ıraksay seçiminin de bağlantı tahmin doğruluğunda ne kadar büyük etkisi olduğunu açık bir şekilde göstermektedir.

KAYNAKÇA

- [1] Clauset, A., Moore, C., Newman, M.: Hierarchical structure and the prediction of missing links in networks. *Nature* **453** (2008)
- [2] Hasan, M.A., Zaki, M.J.: A survey of link prediction in social networks. In Aggarwal, C.C., ed.: *Social Network Data Analytics*. Springer US (2011) 243–275
- [3] Singh, A.P., Gordon, G.J.: Relational learning via collective matrix factorization. In: *KDD'08*. (2008)
- [4] Harshman, R.: Foundations of the parafac procedure: Models and conditions for an “explanatory” multi-modal factor analysis. *UCLA Working Papers in Phonetics* **16** (1970)
- [5] Tucker, L.R.: Implications of factor analysis of three-way matrices for measurement of change. In Harris, C.W., ed.: *Problems in measuring change*. University of Wisconsin Press, Madison WI (1963) 122–137
- [6] Yılmaz, Y.K., Cemgil, A.T., Simsekli, U.: Generalised coupled tensor factorisation. In: *The Neural Information Processing Systems (NIPS)*. (2011)
- [7] Ermis, B., Acar, E., Cemgil, A.T.: Link prediction via generalised coupled tensor factorisation. *CoRR* (2012)
- [8] Yılmaz, Y.K., Cemgil, A.T.: Probabilistic latent tensor factorization. In: *LVA/ICA*. (2010) 346–353
- [9] Zheng, V.W., Cao, B., Zheng, Y., Xie, X., Yang, Q.: Collaborative filtering meets mobile recommendation: A user-centered approach. In: *AAAI*. (2010)