Ürün Özelliklerinin Konu Modelleme Yöntemi ile Çıkarılması

Product Aspect Extraction with Topic Model

Ekin Ekinci, Sevinç İlhan Omurca; {ekin.ekinci, silhan}@kocaeli.edu.tr

Öz

Özellik tabanlı duygu analizindeki en önemli ve güç görevlerden biri duyguların ifade edildiği başlık olarak tanımlanan özelliklerin çıkartılmasıdır. İnternetin günlük hayatımızın vazgeçilmez bir parçası haline gelmesi ile birlikte çevrimiçi kullanıcı yorumlarında yaşanan büyük artış, hem otomatik bir yöntemin geliştirilmesi hem de özelliklerin doğru bir şekilde çıkartılmasını gerektirmektedir. Son yıllarda metin madenciliği uygulamalarında büyük önem kazanan konu modelleme yöntemleri ise bu alanda tercih edilmeye başlanmıştır. Büyük boyutlu dokümanlardan denetimsiz bir şekilde gizli yapıyı keşfeden konu modelleme güçlü bir yöntem olarak karşımıza çıkmaktadır. Bu çalışmada kullanıcı yorumlarından ürün özelliklerini çıkarmada en popüler konu modelleme yöntemlerinden biri olan Gizli Dirichlet Ayırımı (GDA) kullanılmıştır. Türkçe otel yorumları üzerinden elde edilen deneysel sonuçlar, GDA'nın özellik çıkarmada başarılı olduğunu göstermiştir.

Anahtar Sözcükler: *Özellik tabanlı duygu analizi, Özellik çıkarma, Konu modelleme, Gizli dirichlet ayırımı*

Abstract

Extraction of aspects, which are defined as titles which expresses sentiments in texts, is one of the most important and challenge task in aspect based sentiment analysis. Along with Internet has become an indispensable part of our daily life, the huge increase in user reviews requires both the development of an automated method and the proper extraction of aspects. In recent years, topic modeling methods which have gained great importance in text mining applications have begun to be preferred in this field. Topic modeling which discovers latent structure from huge documents unsupervisedly emerges as a powerful method.

In this study, Latent Dirichlet Allocation (LDA) which is the most popular topic modeling method is used to extract aspects from user reviews. The experimental results which are obtained from Turkish hotel reviews have shown that LDA is successful in aspect extraction.

Keywords: *Aspect based sentiment analysis, Aspect extraction, Topic modeling, Latent dirichlet allocation*

Gönderim ve kabul tarihi : 15.11.2016 - 01.06.2017

1. Giriş

Web teknolojilerinde yaşanan gelişmeler sonucunda çevrimiçi yorum sitelerinin yaygınlaşması ve bu siteler üzerinden yapılan ürün yorumlarının kullanıcılar üzerinde büyük etki gösterdiğinin anlaşılması ile firmalar müşterinin ürünler hakkındaki düşüncelerini bu ortamlar üzerinden değerlendirmeye başlamışlardır [1, 2]. Bu tür ortamlardaki verilerin sürekli ve önlenemez artışı ve yapılan bir araştırmaya göre genç nüfusun yaklaşık %50'sinin günlük dilde yazılmış olan bu yorumlardan etkilenerek ürün aldıklarının tespit edilmesi de veri analizi açısından önemli bir kaynağın varlığını gözler önüne sermektedir [3]. Yalnız bu denli büyük ve dağınık veri kümesi üzerinden bir çıkarıma varmak için tüm yorumların tek tek okunması mümkün olmamakla birlikte firma ve müşterilerin doğru bir karar vermesi için de otomatik bir sisteme ihtiyacın olduğu açıktır [4].

2000'li yıllarda ortaya çıkan ve günümüzün önemli araştırma alanlarından birisi haline gelen duygu analizi; kişilerin varlıklar, olaylar üzerine fikirlerini, duygularını, değerlendirmelerini, değer biçmelerini, tutumlarını ve hislerini analiz etme işi olarak tanımlanmaktadır [5]. Araştırmacılar, duygu analizi problemini; doküman tabanlı, cümle tabanlı ve özellik tabanlı duygu analizi olarak üçe ayırmaktadırlar. Dokümanda bahsi geçen temel varlık için dokümanı pozitif ya da negatif olarak sınıflandırma doküman tabanlı duygu analizi iken

bunu dokümanda yer alan cümlelerin her biri için gerçekleştirme ise cümle seviyesinde duygu analizi olarak tanımlanmaktadır. Dokümandaki temel varlığın özelliklerini (ürün özellikleri) niteleyen duygu ifadelerine göre bu özellikler için pozitif ya da negatif olarak sınıflandırma ise özellik tabanlı duygu analizi olarak tanımlanmaktadır [6]. Aslına bakılacak olursa doküman ve cümle seviyesindeki analiz arasında temel bir fark bulunmamaktadır. Cümleler dokümanların özet halidir ve her ikisinde de pozitif ya da negatif olma durumuna göre genel bir analiz yapılmaktadır. Ayrıca bir dokümanın negatif olarak sınıflandırılması o dokümanın tümünde olumsuz bir duygudan bahsedildiği anlamına gelmemektedir. Limitli bir analiz yapmamıza neden olan bu yöntemlerde üzerine duygu belirtilen asıl hedef (özellik) belli değildir. Tüm bunlar göz önünde bulundurulduğunda etkili bir duygu analizi gerçekleştirebilmek için ürün özelliklerinin ve bu özellikleri niteleyen duygu ifadelerinin çıkartılmasını sağlayan etkili bir modele ihtiyaç duyulmaktadır. Özellik tabanlı duygu analizinde, özellik ile kastedilen metinlerde duyguların ifade edildiği başlıklar yani; özellik üzerine yorum yapılan temel varlık, bu varlığın özellikleri, alt parçaları ve alt parçalarının özellikleri şeklinde ifade edilebilir [5, 6]. "Personel çok çalışkandı." yorumunda "personel" kelimesi duygu analizindeki özelliğe karşılık gelmektedir.

Bu çalışmada, özellik tabanlı duygu analizi kapsamında otel ile ilgili Türkçe kullanıcı yorumları içerisinde gizli olarak bulunan ürün özelliklerini çıkarmak amacıyla popüler konu modelleme yöntemlerinden biri olan GDA'nın uygulanması hedeflenmiştir.

Çalışmanın geri kalanı; 2. bölümde GDA ile ilgili güncel çalışmalar özetlenmiştir. 3. bölümde GDA ayrıntılı olarak anlatılmıştır. 4. bölümde deneysel çalışma anlatılmış olup, 5. bölümde ise çalışmanın sonuçları değerlendirilmiştir.

2. İlgili Çalışmalar

Pek çok farklı dildeki pek çok farklı metin üzerinden özellik çıkarımı yine pek çok farklı yöntem kullanılarak gerçekleştirilmektedir. Bununla birlikte konu modelleme yöntemlerinin de özellik çıkarımında son yıllarda sıklıkla tercih edildiği görülmektedir. Bu bölümde GDA konu modelleme yöntemi ile özellik çıkarımı yapan çalışmaların bir kısmı özetlenmiştir.

Li vd. [7] dokümandaki belirgin özellikleri çıkarmak amacıyla Sentiment-LDA ve Dependency-Sentiment-LDA olmak üzere iki yöntem önermişlerdir. Önerdikleri yöntemler ile ürün özelliklerini çıkartırken eşzamanlı olarak duygu ifadelerini de çıkartmışlardır. Bu yöntemleri; dokümandaki duygu ifadelerinin temel varlık ile ilişkili olması fikrinden yola çıkarak geliştirmişler ve duygu ifadeleri ve belirgin özellikleri bütün olarak ele alınmıştır. Dependency-Sentiment-LDA'da ayrıca sentiment polaritelerinin belirlenmesini amaçlamışlardır. Wang [8] yarı denetimli bir konu modelleme yöntemi olan Co-LDA yöntemi ile ürün özellikleri ile ilgili pozitif/negatif duyguyu ortaya çıkarmayı amaçlamıştır. Li'nin çalışmasında olduğu gibi duygu ifadelerini ve belirgin özellikleri eşzamanlı olarak modellemiştir. Bu yöntemin yarı denetimli olması uzmanlar tarafından düzgün yazılmış etiketli yorumların veri seti olarak kullanılmasından kaynaklanmaktadır. Co-LDA modeli iki kısımdan oluşmaktadır. Birinci kısım Sentiment-LDA olarak adlandırılmıştır ve duygu ifadelerine yüksek olasılık değeri atamaktadır. İkinci kısım Topic-LDA olarak adlandırılmaktadır ve belirgin özelliklere yüksek olasılık değeri atamaktadır. Jo ve Oh [9] aynı başlık altındaki ürün özelliklerinin yorum içerisinde birbirine yakın oldukları fikrinden yola çıkarak bir cümledeki tüm kelimelerin tek bir ürün özelliği ile ilişkili olduğu yaklaşımını varsayan S-LDA (Sentence-LDA) yöntemini önermişlerdir. Sonra ise bu yöntemin gelişmiş bir hali olan ASUM'u (Aspect Sentiment Unification) geliştirmişlerdir. Bu yöntemde ürün özellikleri ve duygu ifadeleri birlikte modellenerek {özellik, duygu ifadesi} çiftlerinin çıkartılması sağlanmıştır. Bu amaçla duygu ifadelerinin küçük bir kümesinden de yararlanılmıştır. Yöntemler elektronik ve restoran veri kümelerine uygulanarak ürün özellikleri ve {özellik, duygu ifadesi} çiftleri başarılı bir şekilde elde edilmiştir. Xueke vd. [10] GDA tabanlı ve onun eksikliklerini gideren JAS (Joint Aspect/Sentiment) modelini önermişlerdir. Bu yöntem ile ürün özellileri kullanıcı yorumlarından çıkartılırken bu özelliklere ait duygu ifadeleri polariteleri ile birlikte çıkartılarak özellik-bağımlı bir sözlük oluşturulmuştur. Xianghua vd. [11] Çince yazılmış yorumlar üzerinden GDA ile ilk adımda genel konuları çıkartmışlardır. Lokal konuların ve ilişkili duygu ifadelerinin çıkartılmasında ise tüm GDA'yı tüm dokümanda denemek yerine kayan pencere tabanlı bir yöntem önermişlerdir. Ding vd. [12] GDA tabanlı yeni bir hibrid model olan HDP-LDA'yı (Hierarchical Drichlet Process-Latent

Dirichlet Allocation) önermişlerdir. Bu yöntemin klasik GDA'dan farkı belirgin özelliklerin sayısını otomatik olarak belirlemesidir. Moghaddam ve Ester [13], soğuk-başlangıç problemleri üzerinden ürün özelliklerini çıkartmak ve ürün özeliklerine ait oylamayı bulabilmek amacıyla FLDA (factorized LDA) yöntemini önermişlerdir. Soğuk başlangıç; öneri sistemlerinde ürün özelliği ile ilgili yeterli miktarda yorum bulunmamasından kaynaklı bir problemdir. Ürün özelliklerinin ilgili eleman altındaki tekrar sayısı, ilgili elemana ait ürün özelliklerinin tekrar sayısı ve yorum yapan kişinin en çok hangi ürün özelliği için yorum yaptığı modelin varsayımları arasında yer almakta ve ilgili eleman ve yorum yapan kişi üzerinden ürün özellikleri belirlenmektedir. Wang vd. [14] FL-LDA (Fine-grained Labeled LDA) ve UFL-LDA (Unified Fine-grained Labeled-LDA) olmak üzere iki tane yarı denetimli konu modelleme yöntemi ile kullanıcı yorumlarından ürün özelliklerini çıkartmışlardır. FL-LDA yönteminde ürün özelliklerinin çekirdek kümesi modele dahil edilerek bu çekirdek küme ile ilişkili kelimelerin ürün özellikleri olarak etiketlenmesi amaçlanmıştır. UFL-LDA ise FL-LDA'nın gelişmiş bir versiyonu olarak planlanmıştır. Zheng vd. [15] restoran, otel, MP3 çalar ve kamera yorumları üzerinde gerçekleştirdikleri AEP-LDA (Appraisal Expression Pattern-LDA) ile belirgin ürün özelliklerini kullanıcı yorumlarından çıkartmışlardır. Önerdikleri yöntem, cümle seviyesinde olup bir cümleyi oluşturan tüm kelimelerin aynı konuya ait olduğu fikrini esas almaktadır. Yine burada da özellikler ve duygu ifadeleri eşzamanlı olarak çıkartılmışlardır. Lau vd. [16] belirgin özellikleri ve bunlara bağlı duygu ifadelerini yakalamak amacıyla ürünler için bulanık bir ontoloji öğrenme yöntemini geliştirmişlerdir. GDA tabanlı bu yöntem ile hem taksonomiye dayalı ilişkiler hem de taksonomiye dayalı olmayan ilişkiler çıkartılmaktadır. Yin vd. [17] GDA tabanlı DTAS (Dependency - Topic - Affects - Sentiment - LDA) yöntemi ile konuları ve eşzamanlı olarak duygu ifadelerini çıkartmışlardır. GDA'daki temel yaklaşım olan sözcük torbası yaklaşımı önerilen bu yöntemde yok sayılmış olup konuların bir Markov Zincirinden geldiği varsayılmıştır. Ürün özellikleri çıkartılırken iki tane önsel bilgiden yararlanılmıştır: i) eğer bir cümlede isim yoksa bu cümle bir önceki cümle ile aynı konuyu paylaşır ve ii) eğer iki cümlenin konusu aynı ise bu cümlelerin sentimentlerinin de aynıdır. Bagheri vd. [18] dokümanın cümle yapısını esas alan ve GDA'nın bir uzantısı olarak tasarladığı ADM-LDA (Aspect Detection Model-LDA) yöntemini belirgin özellikleri kullanıcı yorumlarından çıkartmak için önermişlerdir. Bu yaklaşımda da Yin'in yaklaşımında olduğu gibi sözcük torbası yaklaşımı yok sayılmış ve özelliklerin bir Markov Zincirinden geldiği ve koşullu bağımsız oldukları varsayılmıştır. Poria vd. [19] GDA'daki kelime dağılımlarının hesaplanmasında yeni bir yöntem geliştirmişlerdir. Önerdikleri Sentic-LDA söz dizilimsel yaklaşım yerine anlamsal yaklaşımı kullanarak özellik tabanlı duygu analizini gerçeklemektedir.

3. Gizli Dirichlet Ayırımı

Konu modelleme yöntemleri; kelimeler üzerinde olasılık dağılımı gösteren konuların rastgele bir araya gelerek dokümanları oluşturduğu esasına dayanmaktadır [20]. Burada konu ile ifade edilmek istenen bir dokümanda tartışılan temel fikirdir yani dokümanın temasıdır. "Twitter'ın gündeminde bugün ne var?", "Veri madenciliği ile ilgili 10 yıl öncesi ve günümüz araştırma konuları ve aralarındaki farklar nelerdir?", "Müşteriler 'A' restoranın hangi yönlerini beğeniyor, hangi yönlerini ise beğenmiyor?" şeklindeki sorulara cevap verme konu modellemenin çalışma alanına girmektedir. Konu modelleme için önerilen algoritmalar istatistiksel yöntemler olup dokümanı meydana getiren kelimeleri analiz ederek bir sonuca varmayı amaçlar. Yöntemler, konuların birbirleri ile olan bağlantısı, zaman içerisinde gösterdikleri değişimleri keşfederken herhangi bir etiketleme adımına ihtiyaç duymazlar [21].

Makine öğrenmesi ve metin madenciliği uygulamalarında büyük önem kazanan ve en temel ve en popüler konu modelleme yöntemlerinden birisi olan Gizli Dirichlet Ayırımı (Latent Dirichlet Allocation-LDA), doküman gibi ayrık verileri modellemek ve dokümanı meydana getiren konuları ortaya çıkarmak için kullanılan üretici grafiksel modeldir [22, 23]. Burada "gizli" dokümanı oluşturan gizli konuları keşfederek dokümanın anlamını bulmayı ifade etmektedir [24]. Üretici model (generative) ile kastedilen ise basit bir olasılıksal süreç ile dokümandaki kelimelerin gizli (rastgele) değişkenler çerçevesinde üretilmesidir yani dokümanın oluşturulmasıdır [20, 24]. Tamamen denetimsiz bir yöntem olan GDA herhangi bir önbilgiye ihtiyaç duymamakla birlikte, kelime torbası yaklaşımına dayalı çalışmaktadır. Kelimelerin doküman içerisindeki yerleşimi göz ardı edilirken, kelimelerin birlikte bulunması bu yöntemde kullanılmaktadır. GDA için üretici model Şekil 1'de verilmiştir.

1. Her konu k için $k \in [1,K]$

a. Kelimelerin konular içerisindeki dağılımını belirle: $\varphi_k \sim Dirichlet(\beta)$

2. Her doküman d için $m \in [1,M]$

a. Konuların doküman içerisindeki dağılımını belirle: $\theta_m \sim Dirichlet(\alpha)$

b. Doküman m'deki her kelime w_n için

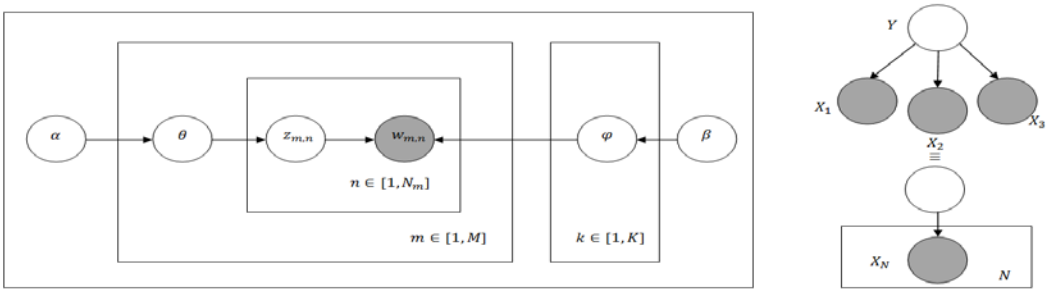
i. Rastgele bir konu belirle $z_{m,n} \sim Mult(\theta_m)$

ii. Rastgele bir kelime seç $w_{m,n} \sim Mult(\varphi_k, z_{m,n})$

Şekil 1. GDA için üretici model

Her doküman, konuların rastgele karışımından meydana gelmekte, dokümanı oluşturan kelimelerin her biri de konuların bir tanesinden seçilmektedir. Konular da sabit bir sözlük içerisindeki kelimelerden olasılık dağılımı göstermektedir. Üretici süreç ilk olarak sözlükteki kelimelerin konular altında örneklenmesi ile başlar. Bir sonraki adımda her konu için konunun dokümanda bulunma olasılığı örneklenir. Dokümanda yer alan her kelime için konu örneklenir ve son olarak ilgili konu için kelime örneklenir. Kelimelerin konular altındaki, konuların da dokümandaki bulunma olasılığı Dirichlet dağılımı ile elde edilir. Dirichlet dağılımı çokterimli değişkenler için eşlenik önsel dağılımdır [25]. Çokterimli değişkenler karşılıklı dışarlanmış mümkün K adet durumdan birini almaktadır.

GDA'nın grafiksel temsilinde ise plate notasyonundan yararlanılmaktadır. Plate notasyonu tekrarlayan yapıları yani aynı tipte birden fazla nesnenin olduğu durumları ifade etmek için kullanılmaktadır. GDA için plate notasyonu ise gözlemlenen verinin rastgele değişkenler ve bu değişkenlerin yönlü kenarlar boyunca yayılımı üzerinden nasıl üretildiğini açıklamaktadır. GDA'ya ait plate notasyonu Şekil 2'de, parametrelerin açıklaması ise Çizelge 1'de verilmiştir. Aslında konu modellemedeki temel amaç; doküman koleksiyonundan konuların çıkartılmasıdır. Bunu yaparken elimizde sadece dokümanlar gözlenebilir durumda olup; kelimelerin konulara atanması, konuların dokümandaki ve kelimelerin konulardaki dağılımları gizlidir. Bu nedenle Şekil 2'de gözlemlenen değişkenler gri renkle temsil edilirken gözlenemeyenler beyaz renk ile temsil edilmiştir.



Şekil 2: GDA için grafiksel model

Çizelge 1. GDA için Grafiksel Model Parametreleri

Parametre	Açıklaması
M	Toplam doküman sayısı
K	Gizli konuların sayısı
V	Sözlükte bulunan toplam kelime sayısı
α	Dirichlet parametresi
β	Dirichlet parametresi
θ	Konuların dokümanlardaki dağılımı
φ	Kelimelerin konulardaki dağılımı
N_m	m. dokümanın boyutu
$z_{m,n}$	m. dokümandaki n. konumda bulunan kelimenin konusu
$w_{m,n}$	m. dokümandaki n. konumda gözlemlenen kelime

Verilen grafiksel modele göre tüm gizli ve gözlemlenen rastgele değişkenlerin birleşik dağılımı Eşitlik 1'de verilmiştir.

$$\left(\prod_{k=1}^{K} p(\varphi_k|\beta)\right)\left(\prod_{m=1}^{M} p(\theta_m|\alpha)\right)\left(\prod_{n=1}^{N} p(z_{m,n}|\theta_m)p(w_{m,n}|z_{m,n},\varphi_k)\right) \quad (1)$$

Yukarda da belirtildiği üzere GDA ile asıl amaçlanan model parametrelerinin elde edilmesidir. Bu amaçla Eşitlik 2'deki sonsal dağılım kullanılmaktadır.

$$p(\beta_{1:K},\theta_{1:M},z_{1:M}|w_{1:M}) = \frac{p(\beta_{1:K},\theta_{1:M},z_{1:M},w_{1:M})}{p(w_{1:M})} \quad (2)$$

Model parametrelerinin elde edilmesi için ise örnekleme yöntemlerinden yararlanılmaktadır. Bu çalışmada Collapsed Gibbs Örnekleme yöntemi kullanılmıştır. Collapsed Gibbs Örneklemenin klasik Gibbs Örneklemeden farkı; kelimelerin konular altındaki olasılıklarını (φ) ve konuların dokümanda bulunma olasılıklarını (θ) dışarlayıp sadece gizli değişken atamaları (z) üzerinden örnekleme gerçekleştirmesidir. Collapsed Gibbs Örnekleme ile belli bir kelimenin hangi konuya atanacağı Eşitlik 3'ten yararlanılarak bulunmaktadır.

$$p(z_i = k|w_i = v,m,\alpha,\beta,.) = \frac{n_{kv-i}+\beta_v}{\sum_{w\in V} m_{w,k}+V\beta}\frac{n_{mk-i}+\alpha}{N_m-1+\alpha K} \quad (3)$$

Eşitlik 3'te w_i, d, α, β ve diğer tüm kelimelerin hangi konuya atanmış oldukları ('.' ile temsil ediliyor) biliniyor iken $z_i = k$ olma olasılığı bulunmaktadır. $-i$ ile $w_i = v$'nin ataması dışarlanmaktadır. Bu işlemler koleksiyonda bulunan tüm dokümanlardaki tüm kelimeler için yapıldıktan sonra φ ve θ değerlerinin Eşitlik 4 ve 5'e göre güncellemesi yapılmaktadır.

$$\theta_{m,k} = \frac{n_{m,k}+\alpha_k}{N_m+\alpha K} \quad (4)$$

$$\varphi_{k,v} = \frac{n_{kv}+\beta_v}{\sum_{v'}^{V}(n_{kv'}+\beta_{v'})} \quad (5)$$

4. Deneysel Çalışma

Bu bölümde GDA'nın veri kümesi üzerine uygulanması ve elde edilen sonuçların niteliksel ve niceliksel olarak değerlendirilmesi anlatılmıştır.

4.1. Veri Kümesi ve Önişleme

Çevrimiçi yorum siteleri ile birlikte ürünler ile ilgili kullanıcı yorumları hem kullanıcılar hem de firmalar için önemli bir kaynak haline gelmiştir. Bu çalışma kapsamında otel ile ilgili kullanıcı yorumları üzerinden ürün özelliklerinin çıkartılması amaçlanmıştır. Türkçe bir veri kümesi[1] oluşturmak için bir web crawler aracılığı ile popüler bir web sitesi olan www.otelpuan.com'dan yararlanılmıştır ve sonuçta etiketli bir küme oluşturulmuştur [6, 26, 27]. Veri kümesine ait ayrıntılı bilgi ve daha önce yapılan çalışmalarda çıkartılan özellik sayısı da Çizelge 2'de yer almaktadır.

Çizelge 2. Veri Kümesine ait Özet Bilgi

Ürün	Yorum Sayısı	Cümle Sayısı
Otel	1000	5364

Elde edilen bu ham veri kümesinden ürün özelliklerinin çıkartılması için yorumların önişleme adımına tabi olması gerekmektedir. Bu amaçla Türkçe doğal dil işleme kütüphanesi olan

[1] http://dx.doi.org/10.13140/RG.2.2.13338.44488

Zemberek'ten yararlanılmıştır [28]. Önişleme adımında ilk olarak yanlış yazılan kelimeler düzeltilmiş, kelimler kök/gövde durumuna getirilip, etkisiz kelimeler, noktalama işaretleri, rakamlar yorumlardan temizlenerek veri kümesinin son hali elde edilmiştir.

4.2. Modelin Uygulanması

Modelin uygulanması aşamasında, Collapsed Gibbs Örnekleme 1000 iterasyonda gerçekleştirilmiştir. Konu modelleme yöntemlerinde konu sayısı bilindiği varsayılarak başta belirlenir. Bu çalışmada K=100 kabul edilmiştir. Ayrıca konu modelleme yöntemleri Dirichlet parametrelerine karşı duyarlı değildir. Bu çalışmada Dirichlet parametreleri simetrik olarak atanmıştır ve $\alpha = \frac{50}{K} = 0.2$ olarak belirlenmişken, $\beta = 0.01$ olarak belirlenmiştir [29]. 20 ile 50 arasında değişen çıkartılan ürün özellikleri ile sonuçlar değerlendirilmiştir.

Sonuçların niteliksel olarak değerlendirilmesi amacıyla çıkartılan ürün özellikleri Çizelge 3'te verilmiştir.

Çizelge 3. LDA ile Çıkartılan Ürün Özellikleri

Etiket	Ürün Özellik
Yiyecek/İçecek	kalite, içecek, yemek, restoran, hizmet, personel, alaka, keyif, dondurma, kutu
İmkan	kalite, otel, imkan, fiyat, personel, yemek, tur, para, ilgi, alaka
Bina	bina, sistem, sorun, restore, restorasyon, mekanik, tasarım, cihaz, daire, detay
Personel	bey, hanım, şef, deniz, personel, aşçı, garson, restoran, müdür, bar
İlgi/Alaka	otel, tatil, ilgi, tercih, alaka, tavsiye, lezzet, yardım, ikram, animasyon

Çizelge 3'te de verildiği üzere her konu için ilk 10 ürün özelliği verilmiştir. Konu etiketleri ise manuel olarak belirlenmiştir.

Sonuçların niceliksel olarak değerlendirilmesi adımında ise Rand Indeks ölçütünden, kesinlik, duyarlılık ve F-ölçümünden yararlanılmıştır. Rand Indeks, kümelemede kullanılan ve 2 küme arasındaki benzerliği bulan bir ölçüttür. Burada Rand Indeks ile önceden çıkartılmış ürün özellikleri ile konu modelleme sonucu elde edilen ürün özelliklerinin benzerlikleri tespit edilerek konu modelleme yönteminin Türkçe veri kümesi üzerindeki başarısının gözlemlenmesi amaçlanmıştır. Kesinlik ile ideal küme ve GDA sonucu oluşan kümenin kesişiminin GDA sonucu oluşan kümeye oranı, duyarlılık ile ideal küme ve GDA sonucu oluşan kümenin kesişiminin ideal kümedeki elemanlara oranı hesaplanmaktadır. F-ölçümü ise kesinlik ve duyarlılık ölçümlerinin harmonik ortalamasıdır. Niceliksel değerlendirme yapılırken Çizelge 4'teki matristen yararlanılmaktadır.

Çizelge 4. Niceliksek Değerlendirme için Kullanılan Matris

	GDA sonucu oluşan küme	GDA sonucu oluşan kümede olmayanlar
İdeal küme	a	b
İdeal kümede olmayanlar	c	d

Burada ideal küme etiketli veri kümesinden elde edilen ürün özelliklerinin kümesidir. Deneyler sonucunda a=155, b=39, c=69 ve d=13021 olarak bulunmuştur. Rand Indeks hesabı Eşitlik 6'ya göre yapılmıştır.

$$Rand = \frac{a+d}{a+b+c+d} \quad (6)$$

Sonuç olarak %99 gibi bir başarı elde edilmiştir. Kesinlik, duyarlılık ve F-ölçümü hesabı da sırasıyla Eşitlik 7, 8 ve 9'a göre yapılmıştır.

$$Kesinlik(p) = \frac{a}{a+c} \quad (7)$$

$$Duyarlılık(r) = \frac{a}{a+b} \quad (8)$$

$$F - ölçümü = \frac{2 \times p \times r}{p+r} \quad (9)$$

Kesinlik %69, duyarlılık %80, F-ölçümü %74 olarak elde edilmiştir. Bu da GDA'nın ürün özelliklerini çıkarmada oldukça başarılı olduğunu göstermektedir.

5. Sonuçlar

Bu çalışmada, otel ile ilgili Türkçe kullanıcı yorumlarından ürün özelliklerinin konu modelleme yöntemi ile çıkartılması hedeflenmiştir. Son yıllarda pek çok disiplin için önem kazanan duygu analizindeki önemli görevlerden biri olan özellik

tabanlı duygu analizinde başarılı bir analiz yapmamız için özelliklerin doğru bir şekilde çıkartılması gerekmektedir.

Denetimsiz bir yöntem olan Konu modelleme de duygu analizi gibi son yıllarda popüler bir yöntem olarak pek çok çalışmada sıklıkla tercih edilmektedir. Bu amaçla; yapılan çalışmada duygu analizi ve konu modelleme yöntemleri birleştirilerek başarılı bir özellik çıkarımı yapılması sağlanmıştır.

6. Kaynakça

[1] Picazo-Vela, S., Chou, S.Y., Melcher, A.J., Pearson, J.M. *Why provide an online review? An extended theory of planned behavior and the role of Big-Five personality traits*, Computers in Human Behavior, 26(4), 2010, pp. 685-696.

[2] Thet, T. T., Na, J. C., Khoo, C. S. G. *Aspect-based sentiment analysis of movie reviews on discussion boards*, Journal of Information Science, 36 (6), 2010, pp. 823-848.

[3] Ha, S.H., Bae, S.Y., Son, L.K. *Impact of Online Consumer Reviews on Product Sales: Quantitative Analysis of the Source Effect*, Applied Mathematics & Information Sciences, 9(2L), 2015, pp. 373-387.

[4] Li, Z., Jing, F., Zhu, X. *Movie Review Mining and Summarization*, In Proceedings of the 15th ACM International Conference on Information and Knowledge Management (CIKM'06), 2006, pp. 43-50.

[5] Liu, B. *Sentiment Analysis and Opinion Mining* Synthesis Lectures on Human Language Technologies, Editör: Hirst, G. Morgan & Claypool, 2012.

[6] Türkmen, H., İlhan Omurca, S., Ekinci, E. *An Aspect Based Sentiment Analysis on Turkish Hotel Reviews*, Girne American University Journal of Social and Applied Sciences, 6, 2016, pp. 9-15.

[7] Li, F., Huang, M., Zhu, X. *Sentiment Analysis with Global Topics and Local Dependency*, In Proceedings of the 24th AAAI Conference on Artificial Intelligence, 2010, pp. 1371-1376.

[8] Wang, W. *Sentiment Analysis of Online Product Reviews with Semi-supervised Topic Sentiment Mixture Model*, In Proceedings of 7th International Conference on Fuzzy Systems and Knowledge Discovery, 2010, pp. 2385-2389.

[9] Jo, Y., Oh, A. *Aspect and Sentiment Unification Model for Online Review Analysis*, In Proceedings of 4th ACM International Conference on Web Search and Data Mining, 2011, pp. 815-824.

[10] Xueke, X., Xueki, C., Songbo, T., Yue, L., Huawei, S. *Aspect-Level Opinion Mining of Online Customer Reviews*, China Communications, 10(3), 2013, pp. 25-41.

[11] Xianghua, F., Guo, L., Yanyan, G., Zhiqiang, W. *Multi-aspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet Lexicon*, Knowledge-Based Systems, 37, 2013, pp. 186-195.

[12] Ding, W., Song, X., Guo, L., Xiong, Z., Hu, X. *A Novel Hybrid HDP-LDA Model for Sentiment Analysis*, In Proceedings of 2013 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, 2013, pp. 329-336.

[13] Moghaddam, S., Ester, M. *The flda model for aspect-based opinion mining: addressing the cold start problem*, In Proceedings of the 22nd International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2013, pp. 909–918.

[14] Wang, T., Cai, Y., Leung, H., Lau, R.Y.K., Li, Q., Min, H. *Product aspect extraction supervised with online domain knowledge*, Knowledge-Based Systems, 71, 2014, pp. 86-100.

[15] Zheng, X., Lin, Z., Wang, X., Lin, K.J., Song, M. *Incorporating appraisal expression patterns into topic modeling for aspect and sentiment word identification*, Knowledge-Based Systems, 61, 2014, pp. 29-47.

[16] Lau, R.Y.K., Li, C., Liao, S.S.Y. *Social analytics: Learning fuzzy product ontologies for aspect-oriented sentiment analysis*, Decision Support Systems, 65, 2014, pp. 80-94.

[17] Yin, S., Han, J., Huang, Y., Kumar, K. *Dependency-Topic-Affects-Sentiment-LDA Model for Sentiment Analysis*, In Proceedings of 2014 IEEE 26th International conference on Tools with Artificial Intelligence, 2014, pp. 413-418.

[18] Bagheri, A., Saraee, M., Jong, F. *Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews*, Knowledge-Based Systems, 52, 2013, pp. 201–213.

[19] Poria, S., Chaturvedi, I., Cambria, E., Bisio, F. Sentic LDA: Improving on LDA with Semantic Similarity *for Aspect-Based Sentiment Analysis*, IJCNN, 2016.

[20] Stevyers, M., Griffiths, T. *Probabilistic Topic Models*, Handbook of Latent Semantic Analysis. Editör: Landauer, T., McNamara, D.S. , Dennis, S., Kintsch W. Erlbaum, 2007.

[21] Blei, D. M. *Probabilistic Topic Models*, Communications of the ACM, 55(4), 2012, pp. 77-84.

[22] Mei, Q., Shen, X., Zhai, C. *Automatic Labeling of Multinomial Topic Models*, In Proceedings of ACM KDD, 2007, pp. 490-499.

[23] Phan, X-H., Nguyen, C-T., Le, D-T., Nguyen, L-M., Horiguchi, S., Ha, Q-T. *A Hidden Topic-Based Framework toward Building Applications with Short Web Documents*, IEEE Transactions on Knowledge and Data Engineering, 23(7), 2011, pp. 961-976.

[24] Jadhav N. *Topic Models for Sentiment analysis: A Literature Survey*, Teknik Rapor, 2014, pp. 1-11.

[25] Bishop, C. M. *Pattern Recognition and Machine Learning*, Editör: Jordan, M., Kleinberg, J., Schölköpf B. Springer, 2006.

[26] Ekinci, E., Türkmen, H., İlhan Omurca, S. *Multi-word Aspect Extraction from User Reviews*, 6th World Conference on Innovatıon and Computer Science (INSODE-2016), 2016.

[27] Türkmen, H., Ekinci, E., İlhan Omurca, S. *A Novel Method for Extracting Feature Opinion Pairs for Turkish* Lecture Notes in Artificial Intelligence Springer, ISBN: 978-3-319-44747-6, 2016, pp. 162-171.

[28] Akın, M. D., Akın, A. A. *Türk Dilleri için Açık Kaynaklı Doğal Dil İşleme Kütüphanesi : Zemberek*, Elektrik Mühendisliği, 431, 2007 pp. 38-44.

[29] Blei, D.M., Ng, A.Y., Jordan, M.I. *Latent dirichlet allocation*, The Journal of Machine Learning Research, 3, 2004. pp. 993-1022.