

**DUYGU ANALİZİNDE FİKİR HEDEFİ ÇIKARIMI İÇİN ALANLAR ARASI
BİR MODEL TASARIMI**

Muhammet Yasin PAK

Doktora Tezi

Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Serkan GÜNAL

Eskişehir

Eskişehir Teknik Üniversitesi

Lisansüstü Eğitim Enstitüsü

Ekim 2021

ÖZET

DUYGU ANALİZİNDE FİKİR HEDEFİ ÇIKARIMI İÇİN ALANLAR ARASI BİR MODEL TASARIMI

Muhammet Yasin PAK

Bilgisayar Mühendisliği Anabilim Dalı

Eskişehir Teknik Üniversitesi, Lisansüstü Eğitim Enstitüsü, Ekim 2021

Danışman: Prof. Dr. Serkan GÜNAL

Fikir hedefi çıkarımı, metin verisinde duygu analizinin temel görevlerinden biridir. Son yıllarda, fikir hedefi çıkarımı için bağımlılık ayrıştırıcı tabanlı yaklaşımlar yaygın olarak incelenmiştir. Ancak, bağımlılık ayrıştırıcıları, dil ve dilbilgisi kısıtlamaları ile sınırlıdır. Bu yüzden, bu tez çalışmasında, bilinmeyen alanlardaki İngilizce ürün yorumlarından alanlar arası fikir hedefi çıkarımı için, böyle kısıtları olmayan, sıralı desen temelli bir kural madenciliği modeli önerilmiştir. Böylece, fikir hedefleri çıkarılırken yorumların alanını bilmeye gerek kalmamaktadır. Önerilen model, literatürde sıklıkla karıştırılan fikir hedefi ve özellik kavramları arasındaki farkı da ortaya koymaktadır. Model, iki aşamadan oluşmaktadır. İlk aşamada, hedef alana ait yorum içerisindeki özellikler, kaynak alanlardan otomatik olarak üretilen kurallar kullanılarak çıkarılmaktadır. Ayrıca, kaynak alanlardaki özellikler hedef alana transfer edilmektedir. Bu aşamada ek olarak, özellik çıkarma performansını arttırmak için özellik budama uygulanmaktadır. İkinci aşamada, fikir hedefi çıkarımı için otomatik olarak üretilen kurallar kullanılarak, ilk aşamada çıkarılan özellikler arasından fikir hedefleri çıkarılmaktadır. Önerilen model, farklı alanlardaki veri kümeleri üzerinde değerlendirilmiş ve literatürdeki modellerle karşılaştırılmıştır. Deneysel sonuçlar, bilinmeyen alanlardaki yorumların fikir hedeflerinin, önceki çalışmalara kıyasla daha yüksek doğrulukla çıkarılabildiğini göstermiştir.

Anahtar Sözcükler: Fikir hedefi çıkarımı, Özellik çıkarımı, Duygu analizi, Metin madenciliği.

ABSTRACT

DESIGN OF A MODEL FOR CROSS-DOMAIN OPINION TARGET EXTRACTION IN SENTIMENT ANALYSIS

Muhammet Yasin PAK

Department of Computer Engineering

Eskişehir Technical University, Institute of Graduate Programs, October 2021

Supervisor: Prof. Dr. Serkan GÜNAL

Opinion target extraction is one of the core tasks in sentiment analysis on text data. In recent years, dependency parser-based approaches have been commonly studied for opinion target extraction. However, dependency parsers are limited by language and grammatical constraints. Therefore, in this thesis, a sequential pattern-based rule mining model, which does not have such constraints, is proposed for cross-domain opinion target extraction from English product reviews in unknown domains. Thus, knowing the domain of reviews while extracting opinion targets becomes no longer a requirement. The proposed model also reveals the difference between the concepts of opinion target and aspect, which are commonly confused in the literature. The model consists of two stages. In the first stage, the aspects of reviews are extracted from the target domain using the rules automatically generated from source domains. The aspects are also transferred from the source domains to a target domain. Moreover, aspect pruning is applied to further improve the performance of aspect extraction. In the second stage, the opinion target is extracted among the aspects extracted at the former stage using the rules automatically generated for opinion target extraction. The proposed model was evaluated on several benchmark datasets in different domains and compared against the literature. The experimental results revealed that the opinion targets of the reviews in unknown domains can be extracted with higher accuracy than those of the previous works.

Keywords: Opinion target extraction, Aspect extraction, Sentiment analysis, Text mining.

TEŞEKKÜR

Tez çalışmam boyunca bilgi ve tecrübeleriyle desteğini hiçbir zaman esirgemeyen danışman hocam Sayın Prof. Dr. Serkan GÜNAL'a katkılarından dolayı en derin şükranlarımı sunarım.

Çalışmalarımın başlangıcından sonuna kadar tezimin daha iyi bir noktaya gelmesi için değerli düşünce ve yorumlarıyla katkı sağlayan ve sabırla beni dinleyerek yol gösteren tez izleme jürisinde yer alan hocalarım Doç. Dr. Alper Kürşat UYSAL'a ve Doç. Dr. Semih Ergin'e çok teşekkür ederim.

Sevgi ve desteğini her zaman yüreğimde hissettiğim can yoldaşım, Sevgili Eşim Ayşin PAK'a, hayatımıza neşe ve mutluluk veren Sevgili Oğlum Yiğit Emre PAK'a en derin sevgilerimi sunarım.

Destekleriyle ve dualarıyla bizi yalnız bırakmayan Sevgili Annem'e ve Babam'a sevgi ve saygılarımı sunarım.

Muhammet Yasin PAK

ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ

Bu tezin bana ait, özgün bir çalışma olduğunu; çalışmamın hazırlık, veri toplama, analiz ve bilgilerin sunumu olmak üzere tüm aşamalarında bilimsel etik ve kurallara uygun davrandığımı; bu çalışma kapsamında elde edilen tüm veri ve bilgiler için kaynak gösterdiğimi ve bu kaynaklara kaynakçada yer verdiğimi; bu çalışmanın Eskişehir Teknik Üniversitesi tarafından kullanılan “bilimsel intihal tespit programı”yla tarandığını ve hiçbir şekilde “intihal içermediğini” beyan ederim. Herhangi bir zamanda, çalışmamla ilgili yaptığım bu beyana aykırı bir durumun saptanması durumunda, ortaya çıkacak tüm ahlaki ve hukuki sonuçları kabul ettiğimi bildiririm.

Muhammet Yasin PAK

İÇİNDEKİLER

Sayfa

ÖZET	iii
ABSTRACT.....	iv
TEŞEKKÜR	v
ETİK İLKE VE KURALLARA UYGUNLUK BEYANNAMESİ.....	vi
İÇİNDEKİLER	vii
TABLolar DİZİNİ.....	ix
ŞEKİLLER DİZİNİ.....	x
SİMGELER VE KISALTMALAR DİZİNİ.....	xi
1. GİRİŞ.....	1
1.1. Tanımlar.....	7
1.2. Tezin Amacı ve Hedefleri	8
1.3. Tezin Organizasyonu	8
2. İLGİLİ ÇALIŞMALAR.....	9
3. DUYGU ANALİZİ	14
3.1. Doküman Düzeyinde Duygu Analizi	16
3.2. Cümle Düzeyinde Duygu Analizi	18
3.3. Özellik Düzeyinde Duygu Analizi	18
3.3.1.Özellik duygu sınıflandırma.....	19
3.3.2.Özellik çıkarımı	20
4. ÖNERİLEN MODEL.....	23
4.1. Özellik Çıkarımı	23
4.1.1.POS etiketleme	23
4.1.2.Ön işleme.....	26
4.1.3.Fikir kelimelerinin etiketlenmesi	26
4.1.4.Etiket desenlerinin ve aday özelliklerin çıkarılması	27
4.1.5.Özellik budama.....	29
4.1.6.Kuralların uygulanması.....	29
4.2. Fikir Hedefi Çıkarımı	31

4.3. Otomatik Kural Üretimi.....	32
4.3.1.Özellik çıkarımı için otomatik kural üretimi.....	33
4.3.2. Fikir hedefi çıkarımı için otomatik kural üretimi.....	34
4.4. Katkılar	36
5. DENEYSEL ÇALIŞMA.....	37
5.1. Veri Kümeleri	37
5.2. Deneysel Parametreler	37
5.3. Değerlendirme Ölçütleri	39
6. DENEY SONUÇLARI VE TARTIŞMA.....	40
7. SONUÇ	49
KAYNAKÇA	51
ÖZGEÇMİŞ	57

TABLolar DİZİNİ

Sayfa

Tablo 1.1. Fikir hedefleri verilmiş örnek yorumlar	8
Tablo 4.1. POS etiketleri ve açıklamaları	25
Tablo 4.2. Örnek bir yorum için POS etiketleme, ön işleme ve fikir kelimelerinin etiketlenmesi	27
Tablo 4.3. Örnek yorumlar ve etiket desenleri	28
Tablo 4.4. Etiket grupları ve genel etiket desenleri	28
Tablo 4.5. Tüm kaynak alanlar için örnek budama kelimeleri	29
Tablo 4.6. "memory card" aday özelliği için dizilerin üretilmesi	30
Tablo 4.7. "memory card" aday fikir hedefi için dizilerin üretilmesi	32
Tablo 4.8. D4 hedef alan olmak üzere, D1-D3 kaynak alanları için bazı örnek kurallara ait dizilerin frekansları	34
Tablo 4.9. Özellik çıkarımı için üretilen örnek bir kuralın destek ve güven değerlerinin hesaplanması	34
Tablo 4.10. Fikir hedefi çıkarımı için üretilen örnek bir kuralın destek ve güven değerlerinin hesaplanması	36
Tablo 5.1. Veri kümeleri ve özellikleri	38
Tablo 6.1. Önerilen model ile elde edilen en iyi sonuçlar (F1-skor)	41
Tablo 6.2. Özellik ve fikir hedefi çıkarımı için en iyi sonuçları sağlayan parametreler	41
Tablo 6.3. Her bir hedef alan için fikir hedefi çıkarımının karşılaştırmalı sonuçları (F1-skor)	47
Tablo 6.4. Fikir hedefi çıkarımında önerilen modelle ilgili çalışmaların karşılaştırılması	48

ŞEKİLLER DİZİNİ

Sayfa

Şekil 4.1. Önerilen modelin genel şeması.....	24
Şekil 4.2. $n=4$ için test kurallarıyla birlikte kural kümeleri için genel model ve aday özellik için özellik çıkarma için karşılık gelen kurallar	31
Şekil 6.1. D1 hedef alanı için en iyi sonuçları sağlayan parametrelerle özellik çıkarımı ve fikir hedefi çıkarımında minimum destek değerleri için F1-skorları.....	42
Şekil 6.2. D2 hedef alanı için en iyi sonuçları sağlayan parametrelerle özellik çıkarımı ve fikir hedefi çıkarımında minimum destek değerleri için F1-skorları.....	43
Şekil 6.3. D3 hedef alanı için en iyi sonuçları sağlayan parametrelerle özellik çıkarımı ve fikir hedefi çıkarımında minimum destek değerleri için F1-skorları.....	43
Şekil 6.4. D4 hedef alanı için en iyi sonuçları sağlayan parametrelerle özellik çıkarımı ve fikir hedefi çıkarımında minimum destek değerleri için F1-skorları.....	44
Şekil 6.5. D5 hedef alanı için en iyi sonuçları sağlayan parametrelerle özellik çıkarımı ve fikir hedefi çıkarımında minimum destek değerleri için F1-skorları.....	44
Şekil 6.6. D6 hedef alanı için en iyi sonuçları sağlayan parametrelerle özellik çıkarımı ve fikir hedefi çıkarımında minimum destek değerleri için F1-skorları.....	45
Şekil 6.7. D7 hedef alanı için en iyi sonuçları sağlayan parametrelerle özellik çıkarımı ve fikir hedefi çıkarımında minimum destek değerleri için F1-skorları.....	45

SİMGELER VE KISALTMALAR DİZİNİ

NB	: Basit Bayes (Naïve Bayes)
SVM	: Destek Vektör Makinesi (Support Vector Machine)
DT	: Karar Ağacı (Decision Tree)
kNN	: k-En Yakın Komşu (k-Nearest Neighbor)
MLP	: Çok Katmanlı Algılayıcı (Multi-Layer Perceptron)
HMM	: Saklı Markov Modeli (Hidden Markov Model)
CRF	: Koşullu Rastgele Alanlar (Conditional Random Fields)
PLSA	: Olasılıksal Gizli Anlam Analizi (Probabilistic Latent Semantic Analysis)
LDA	: Gizli Dirichlet Ataması (Latent Dirichlet Allocation)
PMI	: Noktasal Karşılıklı Bilgi (Pointwise Mutual Information)
POS	: Konuşmanın Bölümü (Part of Speech)
GRU	Kapılı Tekrarlayan Birimler (Gated Recurrent Units)
RNN	Tekrarlayan Sinir Ağı (Recurrent Neural Network)
UI	: Tanımlanamayan (Unidentified)

1. GİRİŞ

Her insan çevresindeki diğer insanlarla sözlü veya yazılı olarak iletişime ihtiyaç duyar ve bu nedenle yazının icadından bu yana geçen binlerce yıl içerisinde farklı dillerde birçok yazı ve metin, insanlar arasında aktarılmıştır. Paylaşılan metin miktarı yıllar içerisinde katlanarak artmış ve günümüzde hayal edilemeyecek büyüklüklere ulaşmıştır. İletişim teknolojilerinin ve özellikle internetin ortaya çıkışı ile birlikte paylaşılan veri, sosyal medyanın da dahil olmasıyla devasa büyüklüklere ulaşmıştır. Bu kadar büyük bir veri kaynağının tamamının sadece insan gücüyle analiz edilip anlamlı bilgilerin çıkarılması mümkün gözükmemektedir. Gelişen bilgisayar teknolojileri ve özellikle makine öğrenmesi ve yapay zekâ gibi yöntemler, günümüzde yapısal olmayan yapıda bulunan bu metinlerin yapısal hale getirilmesinde sıklıkla kullanılmaktadır. Böylece, büyük miktarda metnin otomatik olarak analizi yapılabilmekte ve insanların uzun süreler içerisinde elde edebileceği bilgi kısa sürede çıkarılabilmektedir. Veri madenciliğinin bir parçası olan metin madenciliğinde, ele alınan veri yapısal olmayan bir yapıda bulunmakta ve doğal dilde yazılmış metinlerden oluşmaktadır. Bu nedenle, metin madenciliği ele alınması gereken farklı bir çalışma alanı olarak ortaya çıkmaktadır. Metin madenciliği alanı ise zaman içerisinde farklı çalışma alanlarına ayrılmıştır. Analiz edilen metnin çeşitli biçimlerde olması farklı alanların ortaya çıkmasına neden olmuştur. Metinleri içerdikleri bilgi bakımından objektif ve sübjektif olarak iki kategoride ele almak mümkündür. Objektif bilgi içeren metinler herkes tarafından doğrulanabilen ve kanıtlanabilen gerçek olgu ve verileri içermektedir. Sübjektif bilgi içeren metinler ise kişiden kişiye değişen metinler olup kişiye ait duygu, düşünce ve yorumları içerir.

2000’li yılların başından itibaren internet kullanıcı sayıları artmaya başlamış ve bunun sonucunda insanların içerik yayınlatabildikleri, duygu ve düşüncelerini aktarabildikleri ortamlar ortaya çıkmıştır. İnsanlar birçok farklı konudaki görüşlerini sosyal medya platformlarında paylaşmakta, aynı zamanda e-ticaret sistemlerinden aldıkları ürünlere ait beğeni ve eleştirilerini belirtebilmektedirler. Paylaşılan bu yorum ve düşüncelerin analizi, insanların karar verme süreçlerinde kolaylıklar sağlamaktadır. Örneğin, bir kişi bir hizmet veya ürün satın alacağı zaman, aynı ürünü kullanan kişilerin paylaşmış olduğu yorumları okuyarak fikir sahibi olabilmektedir. Sadece ürün alırken değil aynı zamanda herhangi bir konuda karar verirken diğer insanların düşüncelerine başvurulmaktadır. Yorumların analizi tüketici tarafından önemli olmakla birlikte üretici

konumunda olan insanlar için de büyük fayda sağlamaktadır. Herhangi bir ürünü üreten firma, kullanıcıların bu ürünle ilgili yaptığı yorumlara bakarak daha sonra yapabileceği geliştirmelerle alakalı fikir sahibi olabilmektedir. İnsanların kullanımına sunulan her bir hizmet benzer şekilde kullanıcıların fikirlerine ihtiyaç duymaktadır. Analizi yapılacak bu metinler duygu ve düşünceler içerdiğinden sübjektif bilgiler içermektedir. Son yıllarda bu sübjektif bilgi içeren metinlerin analizine ihtiyaç duyulduğundan yeni bir çalışma alanı ortaya çıkmıştır. Fikir madenciliği (opinion mining) veya duygu analizi (sentiment analysis) olarak ifade edilen çalışma alanı; insanların duygu ve düşüncelerini içeren, herhangi bir konu hakkında yazılmış sübjektif bilgilerin analiz edilmesini amaçlar. Yapısal olmayan bir biçimde olan yorum metinleri duygu analizi sonucunda yapısal bir biçime dönüştürülmeye çalışılarak analizi ve yönetilmesi kolay bilgiler haline getirilir.

Metinlerde duygu analizi problemi genel olarak doküman düzeyinde, cümle düzeyinde ve özellik düzeyinde olmak üzere üç farklı düzeyde gerçekleştirilir (Liu B. , 2012). Doküman düzeyinde yapılan analiz, duygu içeren bir metni bir bütün olarak ele alarak, metnin genelinde hâkim olan duyguyu ortaya çıkarmayı amaçlar. Metindeki duygu genellikle pozitif, negatif ve nötr olmak üzere üç sınıftan biriyle gösterilir. Cümle düzeyinde duygu analizinde ise, metindeki her bir cümle ayrı ayrı ele alınır. Metnin genelinde duygulardan biri diğerlerine göre baskın olsa da her bir cümle farklı duygu ifade edebilmekte, ayrıca metin içerisinde objektif bilgiler içeren, ancak duygu içermeyen cümleler olabilmektedir. Cümle düzeyinde analiz, cümleleri ayrı bir şekilde ele alarak doküman düzeyinde analize göre daha anlamlı sonuçlar ortaya çıkarabilir. Cümle içerisinde bir konu hakkında genel bir duygu ifade edilebileceği gibi, konunun farklı yönleri hakkında da yorumlar bulunabilmektedir. Örneğin; bir ürün hakkında yapılan yorumda ürünün geneliyle ilgili bir duygu ifade edilebileceği gibi, ürünün farklı özellikleriyle ilgili duygular cümle içerisinde birlikte bulunabilmektedir. Aynı cümle içerisinde bazı özellikler hakkında pozitif yorumlar bulunurken, bazıları hakkında da negatif ifadeler olabilmektedir. Bu durumda, cümle düzeyinde yapılan analizler de yeterli olmamaktadır. Özellik (aspect) düzeyinde yapılan analiz, hakkında yorum yapılan varlığın (ürün, servis veya herhangi bir konu) her bir özelliği için ayrı ayrı analizinin yapılmasını amaçlar (Liu B. , 2017). Böylelikle, detaylı bir analiz ortaya konularak bir ürünün sadece beğenilip beğenilmediği değil, aynı zamanda hangi özelliklerinin sevilip hangi özelliklerinin sevilmediği ortaya çıkarılmış olur. Örneğin, “Bu telefonu gerçekten çok beğendim” cümlesinde “telefon” için genel bir duygu belirtildiği halde, “telefonun

kamerası oldukça kaliteli, ancak şarjı hemen bitiyor” cümlesinde telefonun “kamera” özelliği için pozitif, şarj (batarya) özelliği için ise negatif duygu belirtilmiştir.

Özellik düzeyinde analiz, fikir hedefi çıkarımı (opinion target extraction) ve duygu sınıflandırma olmak üzere iki alt görevden oluşur. Fikir hedefi çıkarımı, cümle içerisinde hakkında duygu ifade edilen fikir hedeflerinin çıkarılmasını amaçlar. Duygu sınıflandırma aşamasında ise çıkarılan fikir hedeflerinin her biri için ifade edilen duygu; pozitif, negatif veya nötr olarak sınıflandırılır. Fikir hedefi çıkarımı ve özellik çıkarımı kavramları literatürde genellikle aynı anlamlarda kullanılmaktadır. Ancak, özellik ve fikir hedefi kavramları farklı şeyleri ifade etmektedir. Fikir hedefi kavramı özellik kavramının bir alt kümesidir. Yorum içerisinde, bir varlığa ait özellik hakkında duygu ifade edilmişse bu özellik, fikir hedefi olarak tanımlanır. Ancak, objektif bir cümlede geçen bir özellik hakkında kişisel bir yorum veya duygu bulunmuyorsa bu özellik, fikir hedefi olarak nitelendirilmez. Sonuç olarak bir özelliğin fikir hedefi olup olmadığına karar vermek için cümle içerisindeki kullanımına bakmak gerekmektedir. Örneğin, “batarya”, “kamera” ve “ekran” telefon varlığına ait özelliklerdir. “Telefon oldukça güzel bir ekrana sahip” cümlesinde “ekran” özelliği hakkında bir duygu ifade edildiğinden “ekran” özelliği bu cümlede aynı zamanda bir fikir hedefidir. “Telefonun ekranı 7 inçtir” cümlesinde ise “ekran” özelliğinden bahsedilmesine rağmen cümle nesnel bir ifade içerdiğinden burada “ekran” özelliği bir fikir hedefi değildir.

Fikir hedefi çıkarımı problemi veya yaygın kullanımıyla özellik çıkarımı problemi son zamanlarda yoğun bir şekilde çalışılan bir konudur (Rana & Cheah, 2016) (Tubishat, Norisma, & Abushariah, 2018). Fikir hedefi kavramı çalışmalarda farklı kelimelerle ifade edilmektedir. Fikir hedefine karşılık olarak bazı çalışmalar öznitelik (feature) (Jin, Ho, & Srihari, 2009) (Zhang, Liu, Lim, & O’Brien-Strain, 2010), bazı çalışmalar özellik (Marcacini, Rossi, Matsuno, & Rezende, 2018) (Ray & Chakrabarti, 2020) veya nesne özniteliği (Li, ve diğerleri, 2010) kavramlarını, bazıları (Qiu, Liu, Bu, & Chen, 2011) ise özellik ve fikir hedefi kavramlarını birbirlerinin yerine kullanmaktadırlar. Ayrıca, varlık ve özellik kavramları farklı şeyleri ifade etmesine rağmen çalışmaların birçoğu bu iki kavramı özellik kavramı altında toplayarak ele almaktadır. Ancak, cümle içerisinde varlığın özelliği fikir hedefi olabileceği gibi, varlık tek başına da fikir hedefi olabilir. Özellik ve varlık kavramlarını ayrı olarak değerlendiren çalışmalar da literatürde mevcuttur (Shu, Liu, Xu, & Kim, 2016). Sonuç olarak, özellik çıkarma ve fikir hedefi

çıkarmı farklı problemleri çözmeyi amaçlamaktadır. Birçok çalışma bu iki problemi aynı olarak ele alsa da son zamanlarda farklı olduğunu ifade eden çalışmalar olduğu görülmektedir (Wang, He, Jiang, Huang, & Qiu, 2020). Bu çalışmalara göre, özellik çıkarmı bir metin veya metin koleksiyonu içerisinde belirli bir varlığa ait özelliklerin listesini çıkarmayı amaçlamaktadır. Fikir hedefi çıkarımının hedefi ise metin içerisindeki her bir cümle için fikir hedeflerini çıkarmaktır. Bu tez çalışmasında ise özellik çıkarmı, hakkında duygu içerip ifade edilip edilmemesine bakılmaksızın bir cümle içerisindeki tüm özelliklerin çıkarılması işlemi olarak ele alınmaktadır. Fikir hedefi çıkarımı ise özellik çıkarmı aşamasında çıkarılan özelliklerin içerisinde hakkında duygu ifade edilen fikir hedeflerinin çıkarılması işlemidir. Cümle içerisinde çıkarılması beklenen özellikler, çoğu zaman cümle içerisinde açık (explicit) bir şekilde ifade edilse de bazen örtülü (implicit) olarak bulunur. Örneğin, “Telefon oldukça güzel bir ekrana sahip” ve “Telefonun ekranı 7 inçtir” cümlelerinde “ekran” özelliği cümle içerisinde açıkça görüldüğünden açık özellik olarak tanımlanır. “Bu telefon çok pahalı” cümlesinde ise telefonun “fiyat” özelliği hakkında duygu belirtilmesine rağmen, bu özellik açıkça ifade edilmemektedir. Bunun yerine “pahalı” kelimesi içerisinde örtülü olarak bulunmaktadır. Bu tür özellikler ise örtülü özellik olarak ele alınır. Literatürdeki çalışmaların birçoğu açık özelliklerin çıkarılması problemine odaklanmıştır. (Rana & Cheah, 2016). Bu tez çalışmasında da açık özellikler üzerinde çalışılmıştır.

Fikir hedefi çıkarımı için literatürde gözetimli (supervised), gözetimsiz (unsupervised) ve yarı gözetimli (semi-supervised) yaklaşımlar kullanılmaktadır. Gözetimli yaklaşımlar daha yüksek performans gösterdiğinden, diğer yaklaşımlara göre daha çok tercih edilmektedir. Ancak, gözetimli yaklaşımların başarısı etiketli verilerin varlığına bağlıdır. Dolayısıyla başarılı bir fikir hedefi çıkarımı için, üzerinde çalışılan alan (domain) için fikir hedefleri manuel olarak etiketlenmiş verilere ihtiyaç vardır. Burada alan olarak ifade edilen şey, ürün yorumlarının yazılmış olduğu, kamera, restoran vb. kategorilerdir. Verilerin manuel olarak etiketlenmesi zaman alıcı ve zahmetli bir süreç olmaktadır. Ayrıca fikir hedefi çıkarımı yapılacak metin, birçok farklı alana ait olabilir. Örneğin, sadece ürün yorumları üzerinde fikir hedefi çıkarımı yapılmak istendiğinde bile, birçok farklı kategoride ürün olduğu görülmektedir. Ürün yorumlarındaki fikir hedeflerinin başarılı bir şekilde çıkarılabilmesi için her bir kategori veya alan için ayrı ayrı etiketli verilerin elde edilmesi gerekmektedir. Fikir hedefi çıkarımı problemi, sadece ürün yorumlarını değil, birçok farklı alandan metni hedef aldığından, olası her bir alan

için etiketli veri toplama problemi ortaya çıkmaktadır. Bu problemin çözümü için sadece etiketli veriler kullanan gözetimli yaklaşımlar yerine, tamamen etiketsiz verilerin kullanılmasını hedefleyen gözetimsiz yaklaşımlar veya az miktarda etiketli veriyle birlikte etiketsiz verilerin bir arada kullanılmasını amaçlayan yarı gözetimli yaklaşımlar, literatürde bulunmaktadır. Gözetimsiz veya yarı gözetimli yaklaşımların kullanılması etiketli veriye olan ihtiyacı azaltarak, farklı alanlar için kullanılabilecek alan bağımsız fikir hedefi çıkarma modelleri oluşturmayı kolaylaştırmaktadır. Etiketli verilerin kullanımının azaltılması için diğer bir yol ise yeni bir etiketli veri oluşturmak yerine var olan etiketli verilerin kullanılmasıdır. Gözetimli yaklaşımlarda, yeni bir alana ait metin için fikir hedefi çıkarımı yapılmak istendiğinde bu yeni alana ait etiketli verilerin manuel olarak hazırlanması gerekmektedir. Ancak halihazırda bulunan diğer alanlara ait etiketli verilerin kullanılabilmesi, yeni bir alandaki metin için ihtiyaç duyulan etiketli veri miktarını azaltmaktadır. Alanlar arası öğrenme aktarımı modelleri (Cross domain transfer learning models) bir alandaki etiketli verilerin başka bir alana aktarılması problemini ele almaktadır. Alanlar arası öğrenme aktarımı problemi, kaynak (source) alanlardaki etiketli verilerin hedef (target) alana transfer edilmesini amaçlamaktadır. Burada hedef alan; fikir hedefi çıkarımının yapılacağı alanı ifade ederken, kaynak alan; etiketli verilerin kullanıldığı hedef alan dışındaki diğer alanları ifade eder. Hedef alana ait az miktarda etiketli veri olabileceği gibi, herhangi bir etiketli verinin bulunmadığı durumlar da olabilmektedir. Etiketli verinin bulunmadığı durumlarda hedef alana ait etiketsiz veriler, yapılacak transfer işlemi için faydalı bir kaynak olarak kullanılmaktadır. Ele alınması gereken diğer bir olasılık ise hedef alana ait alan bilgisinin bulunmadığı durumlardır. Böyle bir durumda hedef alana ait herhangi bir bilgi olmadığından, hedef alana ait etiketli veya etiketsiz herhangi bir verinin kullanılması da mümkün olmayacaktır. Bu noktada duygu analizinin günümüzdeki olası kullanım alanlarından bahsetmek yerinde olacaktır. Duygu analizinin sıklıkla kullanıldığı alanlardan biri ürün yorumlarının analiz edilmesidir. Bir e-ticaret sitesindeki ürün yorumları, otel rezervasyonu yapan sistemlerdeki otel yorumları ve buna benzer yorumlar ürün yorumlarına örnek gösterilebilir. Otel yorumları farklı oteller için benzerlik gösterse de, e-ticaret sistemlerindeki ürünler farklı alanlarda olabildiğinden bu ürünlere ait yorumların da analizi farklı olmaktadır. Bu durumda her bir farklı alana ait etiketli verilere ihtiyaç duyulmaktadır. Duygu analizinin diğer bir kullanım alanı ise sosyal medyada paylaşılan yorumların analizidir. Bu yorumlar ürün yorumlarından farklı olarak tek bir alanda

yazılan yorumlar yerine farklı alanlardan oluşan yorumlardan oluşmaktadır. Dolayısıyla bu tür durumlarda hedef alana ait bilgi bulunmadığından, etiketli veya etiketsiz verilerin toplanması imkânsız hale gelmektedir. Bu problemin çözümü için tamamen gözetimsiz şekilde çalışan bir yöntem kullanılması gerekmektedir. Diğer bir çözüm ise farklı kaynak alanlara ait etiketli veya etiketsiz verilerin hedef alan için kullanılabilmesini sağlayan yöntemler geliştirilmesidir. Bu tez çalışmasının amacı, hedef alanın bilinmediği durumlarda fikir hedefi çıkarımı için bir yaklaşım ortaya koymaktır.

Fikir hedefi çıkarımı problemi için önerilen alanlar arası öğrenme aktarımı temelli yaklaşımlar kaynak alanlara ait etiketli verileri kullanarak elde edilen bilgiyi hedef alana aktarmaktadır. Özellik ve fikir kelimeleri (opinion words) arasındaki sözdizimsel (syntactic) ilişkiler transfer edilebilmektedir (Wang & Pan, 2019). Ayrıca, farklı alanlarda kullanılan ortak fikir kelimelerinin veya özelliklerin, kaynak alanlardan hedef alana doğrudan transfer edilmesi mümkün olmaktadır (Shu, Xu, & Liu, 2017). Bu tez çalışması, açık özellikler için alanlar arası öğrenme temelli fikir hedefi çıkarımı için bir yaklaşım önermektedir.

Literatürdeki fikir hedefi çıkarımı yaklaşımlarında özellikler ile fikir kelimeleri arasındaki ilişkiden sıklıkla faydalanılmaktadır. Fikir kelimeleri ile ilişkisi olan özellikler fikir hedefi olarak çıkarılmaktadır. Çalışmaların birçoğu fikir hedefi ve fikir kelimesi arasındaki ilişkiyi çıkarmak amacıyla cümle üzerinde bağımlık ayrıştırması (dependency parsing) işlemi uygulamaktadır. Bu işlem sayesinde kelimeler arasındaki ilişkiler başarıyla çıkarılmaktadır, ancak gramer yapısı düzgün olmayan cümlelerde aynı başarıyı gösterememektedir. Sıklıkla kullanılan diğer bir işlem ise Konuşmanın Bölümü (POS-ing. Part of Speech) etiketleme işlemidir. POS etiketleme işleminde cümle içerisindeki her bir kelimeye isim, fiil, sıfat, bağlaç gibi bir etiket atılarak kelimenin cümle içerisindeki türü belirlenir. POS etiketleme işlemi, hemen hemen tüm fikir hedefi çıkarımı yaklaşımlarında kullanılmaktadır.

Bu çalışmada, fikir hedefi çıkarımı için gözetimli kural madenciliği temelli alanlar arası model önerilmiştir. N gram temelli sıralı kural madenciliği kullanılarak ürün yorumlarındaki desenler analiz edilmiş ve fikir hedefi çıkarımı için uygulanmıştır. Analiz edilecek yorumun alanının bilinmediği varsayılmaktadır. Bu durumda hedef alan için etiketli veya etiketsiz veri kullanmak mümkün olmamaktadır ve dolayısıyla sadece kaynak alanlardaki veriler kullanılabilir. Önerilen modelde, özellik kavramı fikir

hedefi kavramından ayrıldığından mevcut kural tabanlı yaklaşımlarda tek aşamada gerçekleştirilen fikir hedefi çıkarımı, bu çalışmada iki aşamada gerçekleştirilmektedir. İlk aşamada, kaynak alandaki etiketli verilerden üretilen özellik çıkarma kuralları kullanılarak özellikler yorum içerisinde çıkarılır. Ortak özellikler hedef alandan kaynak alana transfer edilir. Ayrıca, özellik budama işlemi uygulanır. İkinci aşamada ise, ilk aşamada çıkarılan özellikler içerisinde fikir hedefleri çıkarılır. Bunun için kaynak alandan elde edilen fikir hedefi kuralları uygulanır. Önerilen model, 7 farklı veri kümesi üzerinde gerçekleştirilen deneyler ile değerlendirilmiş ve literatürdeki diğer yaklaşımlarla karşılaştırılmıştır. Önerilen model, fikir hedefi çıkarımı konusunda önceki çalışmaların sonuçlarına göre daha başarılı olmuş, özellikle isim öbeklerinden oluşan fikir hedefleri için en iyi sonuçlar elde edilmiştir.

1.1. Tanımlar

Giriş kısmında kısaca bahsedilen bazı kavramlar bu bölümde örneklerle daha detaylı bir şekilde açıklanmıştır. Duygu bir varlığın kendisi hakkında ifade edilebileceği gibi onun bir özelliği hakkında da ifade edilebilir. "the hotel was amazing" yorumunda, "hotel" bir varlıktır ve duygu varlığın kendisi hakkındadır. "the food of the hotel was delicious" yorumunda ise "food" bir özelliktir ve duygu varlığın özelliği hakkındadır. Özellik çıkarma çalışmalarında varlık genellikle dikkate alınmaz. Özellik terimi hem varlık hem özellik terimlerinin her ikisini birden içerir (Rana & Cheah, 2016).

Özellik çıkarımı probleminde, çıkarılan özellik yorum içerisinde açık bir şekilde geçiyorsa bu özellik açık özellik olarak, aksi halde örtülü özellik olarak ifade edilir. "the food of the hotel was delicious" yorumunda "food" açık özelliktir. "The hotel was very expensive" yorumunda ise "price" örtük özelliktir. Açık özellikler genellikle isim ve isim öbeklerinden oluşur (Liu B. , 2012). Bir yorum bir veya birden fazla isim veya isim öbeği içerebilir ama onlardan bazıları özelliktir. Bir veya daha fazla özellik hakkında duygu ifade edilebilir, ancak bazı özellikler hakkında duygu ifade edilmeyebilir. Özellik hakkında duygu ifade ediliyorsa bu özellik "fikir hedefi" olarak adlandırılır. Fikir hedefleri etiketlenmiş örnek yorumlar Tablo 1.1'de verilmiştir. 1. yorumda "price" ve "service" hem özellik hem fikir hedefi olarak tanımlanmaktadır. "Zoom" ve "optical zoom" yorum 2'deki özellikler iken, "optical zoom" fikir hedefi olarak belirtilmiştir. Yorum 3'te "lens", "viewfinder", "angle" ve "wide-angle" isim ve isim öbekleridir.

Tablo 1.1. *Fikir hedefleri verilmiş örnek yorumlar*

(1) price, service ##the price is reasonable although the service is poor.
(2) optical zoom ##the optical zoom is awesome
(3) viewfinder ##the lens is visible in the viewfinder when the lens is set to the wide-angle
(4) [Fikir hedefi yok] ##i want to buy a phone with good photo quality

Bu isim ve isim öbeklerinden “lens” ve “viewfinder” özellikler iken, “viewfinder” özelliği, hakkında duygu ifade edildiğinden fikir hedefidir. Yorum 4’te "phone", "photo" and "photo quality" isim ve isim öbekleri özellik olarak bulunurken, yorum içerisinde herhangi bir fikir hedefi bulunmamaktadır çünkü herhangi bir duygu ifade edilmemektedir. Sonuç olarak, yorum içerisinde özellik kümesi isim ve isim öbekleri kümesinin bir alt kümesi, fikir hedefi kümesi ise özellik kümesinin bir alt kümesidir.

1.2. Tezin Amacı ve Hedefleri

Bu tezdeki araştırma amaçları şu şekilde listelenebilir:

- İngilizce ürün yorumlarından fikir hedeflerinin çıkarılması amacıyla otomatik kural üretimi için sıralı desen temelli bir model tasarlamak,
- Bilinmeyen alanlarda fikir hedefi çıkarımı için alanlar arası bir yaklaşım adapte etmek,
- Özellik ve fikir hedefi kavramları arasındaki farkı ortaya çıkarmak,
- Özellik çıkarımı için kural üretimi sürecinde fikir hedefi olmayan özelliklerin kullanımını sağlamak.

1.3. Tezin Organizasyonu

Bu tez, organizasyon açısından toplam yedi bölümden oluşmaktadır. Birinci bölümde çalışmanın genel anlatımı, temel kavramlar, amacı ve hedefleri anlatılmaktadır. İkinci bölümde ise konuyla ilgili yapılan çalışmalar anlatılmıştır. Duygu analizi hakkında temel bilgiler üçüncü bölümde verilmiştir. Dördüncü bölümde fikir hedefi çıkarımı için önerilen model sunulmuştur. Beşinci bölümde deneysel çalışma kapsamında kullanılan veri kümeleri, deneysel parametreler ve değerlendirme ölçütleri açıklanmıştır. Altıncı bölümde deneysel sonuçlar sunulmuş ve değerlendirilmiştir. Sonuç bölümü olan yedinci bölümde ise tezin genel bir değerlendirilmesi yapılmıştır.

2. İLGİLİ ÇALIŞMALAR

Fikir hedefi çıkarımı çalışmaları etiketli veri kullanımına bağlı olarak, gözetimsiz, gözetimli ve yarı-gözetimli olarak gruplanabilir. Gözetimsiz modeller sadece etiketsiz veri kullanarak fikir hedeflerini çıkarmayı amaçlarken, gözetimli modeller etiketli veri kullanır. Yarı-gözetimli modeller ek olarak, fikir kelime listesi veya özellik ve fikir kelimesi sözlükleri kullanabilir. Ayrıca, manuel olarak oluşturulan kurallar veya web arama motorları gibi dış kaynaklar kullanılabilir.

Konu modelleme gözetimsiz bir öğrenme modelidir. Bu modelde her bir dokümanın özellik ve fikir kelimelerinin her ikisini kapsayan konuların karışımından oluştuğu varsayılır (Li & Lei, 2021). Brody ve Elhadad (Brody & Elhadad, 2010) özellikleri ve özellik özgü fikir kelimelerini tanımlamak için gözetimsiz yerel konu modelini önermişlerdir. García ve arkadaşları (García-Pablos, Cuadros, & Rigau, 2018) çoklu alan ve çoklu dil için konu modelleme temelli hemen hemen gözetimsiz bir özellik tabanlı duygu analizi sistemi önermişlerdir. Önerdikleri sistemde her bir alan için bir özellikte birlikte pozitif ve negatif fikir kelimelerinden oluşan çekirdek kelimeleri kümesi kullanmışlardır. Ekinci ve İlhan Omurca (Ekinci & İlhan Omurca, 2020) özellik tabanlı duygu analizi için anlamsal bilgiyi modele dahil ederek Concept-LDA isimli bir model geliştirmişlerdir. Konu modelleri yorumdaki fikir hedeflerini çıkartmak yerine büyük dokümanlardan özellikleri çıkarmak ve gruplamak için daha uygundur.

Frekans temelli modeller sıklıkla kullanılan isimlerin genellikle önemli özellikler olduğunu varsayan gözetimsiz modellerdir. Hu ve Liu (Hu & Liu, 2004) sık geçen özellikleri komşu fikir kelimelerini bulmak için kullanmışlardır. Daha sonra sık geçmeyen özellikleri ise fikir kelimelerini kullanarak çıkarmışlardır. Diğer bir çalışmada (Popescu & Etzioni, 2007), Noktasal Karşılıklı Bilgi (PMI- ing. Pointwise Mutual Information) skoru her bir özellik için hesaplanmıştır. Bazı aday özellikleri budayarak frekans temelli modellerin kesinliğini (precision) arttırmak amaçlanmıştır. Özellikler ve fikir kelimeleri arasındaki ilişkiyi ortaya çıkaran sözdizimsel bağımlılıkları kullanarak fikir kelimelerini çıkarmak için bazı kurallar manuel olarak tanımlanmıştır. Li ve arkadaşları (Li, Zhou, & Li, 2015) fikir hedefi çıkarımının performansını arttırmak için PMI-IR ile frekans temelli model önermişlerdir. Deneyler Çince ürün yorumları üzerinde gerçekleştirilmiştir.

Gözetimsiz kural tabanlı modeller özellikler ve fikir kelimeleri arasındaki ilişkiyi kullanarak kural çıkarmayı amaçlar. Herhangi bir fikir kelimesi ile ilişkili olan özellikler fikir hedefi olarak çıkarılmaya çalışılır. Çift yayılım, sözdizimsel bağımlılıkları kullanarak fikir kelimeleri ve özelliklerin her ikisinin eş zamanlı olarak çıkarılmasını sağlayan ön yükleme (bootstrapping) temelli bir modeldir (Qiu, Liu, Bu, & Chen, 2011). Bu modelin en önemli avantajı ön yükleme işlemini başlatmak için sadece başlangıç fikir sözlüğüne ihtiyaç duymasıdır. Bu model çekirdek liste kullanımından dolayı yarı-gözetimli bir model olarak görülebilir. Fikir kelimeleri ve fikir hedefleri çekirdek fikir kelimeleri kullanılarak çıkarılır. Daha sonra çıkarılan kelimeler bir sonraki fikir kelimesi ve fikir hedefi çıkarımı için kullanılır. Bu süreç eklenecek yeni bir kelime kalmayana kadar devam eder. Ancak, ön yükleme temelli modeller hata yayılım problemine neden olur (Liu, Xu, & Zhao, 2012). İlk iterasyonlarda yanlış bir şekilde çıkarılan fikir hedefi ve fikir kelimeleri, sonraki iterasyonlarda muhtemelen yeni hatalara neden olur. Sonuç olarak, bu hatalar ileriki iterasyonlarda kademeli olarak birikir. Ayrıca, sözdizimsel temelli modellerin başarısı bağıllık ayrıştırıcısının başarısına bağlıdır. Gramer hataları olan yorumlar ayrıştırma işleminde başarısız olabilmektedir. Poria ve arkadaşları (Poria, Cambria, Ku, Gui, & Gelbukh, 2014) sağ duyu (common-sense) bilgisi ve sözdizimsel bağımlılıkları kullanarak açık ve örtük özellikleri tespit etmek için tamamen gözetimsiz kural tabanlı bir yaklaşım önerir. Fikir hedefleri manuel olarak oluşturulmuş bağımlılık kuralları ve fikir sözlüğü kullanılarak çıkarılır. Kang ve Zhou (Kang & Zhou, 2017) parça-bütün ilişkisini ve yoruma özgü desenleri birleştirerek nesnel özelliklerin tanımı ile çift yayılımı (Qiu, Liu, Bu, & Chen, 2011) genişletmişlerdir. Birliktelik kural madenciliği (Association rule mining) (Hai, Chang, & Kim, 2011) ve kelime tabanlı çeviri modeli (Word-based translation model) (Liu, Xu, & Zhao, 2012), özellik ve fikir kelimeleri arasındaki ilişkiyi çıkarmak için kullanılan diğer gözetimsiz modellerdir.

Dizi modelleri gözetimlidir ve fikir hedefi çıkarımı problemini dizi etiketleme problemi olarak ele alırlar. Bunun nedeni, özellik ve fikir kelimelerinin bir yorumda dizi içerisinde ve genellikle bağımsız olmalarıdır (Zhang & Liu, 2014). Saklı Markov Modeli (HMM- ing. Hidden Markov Model) (Jin, Ho, & Srihari, 2009) ve Koşullu Rastgele Alanlar (CRF- ing. Conditional Random Fields) (Li, ve diğerleri, 2010) dizi modellerinin örnekleridir. CRF, düzgün bir şekilde çalışabilmesi için büyük miktarda özniteliğe ihtiyaç duyar ve başarısı yorumların gramer yapısı olarak doğruluğuna bağlıdır (Ray & Chakrabarti, 2020). Gözetimli kural tabanlı modeller fikir hedefi çıkarımı için ayrıca

kullanılır. Yorumlardaki fikir hedefleri etiketlendikten sonra kurallar üretilir. Kural tabanlı olmaları açısından bu modeller gözetimsiz kural tabanlı yaklaşımlara benzer. Ancak etiketleme işleminde çekirdek kelimeler yerine etiketli veri kullanılmaktadır. Bu modeller gözetimli olduğundan dizi modellerine benzer. Hu ve Liu (Hu & Liu, 2006) pros ve cons yorumlardan fikir hedeflerini çıkarmak amacıyla dil desenlerini bulmak için gözetimli sıralı desen madenciliği modelini önermişlerdir. Pros ve cons yorumlar özellik ve genellikle fikir kelimesi içeren kısa ifadelerdir. Bu çalışma sıralı desenlerin kullanıldığı ilk çalışmadır. Yorumların duygu içerdiği varsayıldığından, yorum içerisindeki özellikler fikir hedefleri olarak çıkarılır. Yorum POS etiketleyici tarafından etiketlendikten sonra fikir hedefleri özellik etiketiyle etiketlenir. N-gram yaklaşımı 3-gram (POS etiketleriyle birlikte 3 kelime) için kısa parçalar üretmek amacıyla kullanılır. Diğer bir çalışmada (Liu B. , 2007), sınıf sıralı desen madenciliği modeli, diğer bir ifadeyle, sıralı kural madenciliği dil desenlerini bulmak için uygulanır. Kurallar minimum destek (support) değerine göre seçilir ve bu kurallar test yorumundaki özellikleri tanımlamak için kullanılır. Uygulanan model n-gram biçiminde kurallar üretmesine rağmen kural içerisinde boşluklara izin verir. En yüksek güven değerini sağlayan kural seçilir. Rana ve Cheah (Rana & Cheah, 2019) fikir hedefi çıkarımı için sıralı desen madenciliği temelli gözetimli bir model önermişlerdir. Çıkarma işlemi üç adımda gerçekleştirilir. Öncelikle, yorum içerisindeki fikir hedefleri “aspect” etiketiyle, fikir kelimeleri ise “opinion” etiketiyle etiketlenir. Daha sonra, sıralı desenler PrefixSpan algoritması kullanılarak üretilir. İkinci aşamada, sıralı desenler analiz edilerek özellik ve fikir kelimeleri arasındaki ilişkiyi ifade eden fikir hedefi kuralları manuel olarak tanımlanır. Son olarak, açık özellikler fikir hedefi kuralları kullanılarak çıkarılır. Samha ve arkadaşları (Samha, Li, & Zhang, 2014) ürün özelliklerini tanımlayan sıklık etiketlerinden oluşan sıklık kümelerini çıkarmak için bir yaklaşım sunmuştur. Bazı çalışmalar (Htay & Lynn, 2013) (Maharani, Widyantoro, & Khodra, 2015) (Asghar, Khan, Zahra, Ahmad, & Kundi, 2019) (Rana & Cheah, 2018) ürün özellikleri ve fikir kelimelerini çıkarmak için manuel desenlerden yararlanmışlardır.

Son zamanlarda, derin öğrenme modelleri fikir hedefi çıkarımı için de sıklıkla kullanılmaktadır (Do, Prasad, Maag, & Alsadoon, 2019) (Alharbi, Taileb, & Kalkatawi, 2021). Fikir hedefi çıkarımı sisteminin başarılı bir şekilde gerçekleştirilmesi amacıyla, özellikle derin öğrenme modellerinde belirli bir alan ve dilde eğitim için manuel olarak etiketlenmiş verilere ve dile özel kaynaklara ihtiyaç duyulur. Geleneksel makine öğrenimi

algoritmaları ile karşılaştırıldığında, hesaplama kaynakları ve işlem zamanı derin öğrenme ile ilgili diğer önemli konulardır (Al-Smadi, Qawasmeh, Al-Ayyoub, Jararweh, & Gupta, 2018).

Son yıllarda, fikir hedefi çıkarımında etiketli veriye olan ihtiyacı azaltmak için alan adaptasyonu ile ilgili birçok çalışma yapılmıştır. Alanlar arası transfer öğrenme çalışmalarının birçoğu kaynak alanlara ait etiketli veri kullanırken hedef alana ait etiketsiz veri kullanır (Wang & Pan, 2019) (Wang & Pan, 2018) (Wang & Pan, 2020) (Li, ve diğerleri, 2019). Bazı çalışmalarda (Marcacini, Rossi, Matsuno, & Rezende, 2018) hedef alana ait herhangi bir veri kullanılmaz. Alanlar arası transfer öğrenme gözetimli yaklaşımlar için etiketli verinin transfer edilmesine karşılık gelir (Liu, Shi, Ji, & Jia, 2019). Marcacini ve arkadaşları (Marcacini, Rossi, Matsuno, & Rezende, 2018) heterojen ağ temelli (heterogeneous network-based) bir model sunmaktadırlar. Bu modelde kaynak alana ait etiketli özellikler, hedef alana ait dilsel öznitelikler ve etiketsiz özellikler heterojen bir ağdaki düğümler olarak eşleştirilir. Fikir hedefi çıkarımı, hedef alan bilinmediğinden hedef alan için diğer kaynak alanları kullanılarak gerçekleştirilir. Diğer bir yaklaşımda (Jakob & Gurevych, 2010), elle oluşturulmuş öznitelikler (POS etiketleri ve bağımlılıklar) ile doğrusal-zincir CRF modeli alanlar arası ayarlarda fikir hedefi çıkarımı için uygulanır. Diğer bir çalışmada (Ding, Yu, & Jiang, 2017), alanlar arası Tekrarlayan Sinir Ağı (RNN- ing. Recurrent Neural Network) modeli önerilmiştir. Model, manuel olarak tanımlanmış kurallardan oluşan yardımcı görevleri bütünleştirir. Wang ve Pan (Wang & Pan, 2018), etiket gürültüsünü azaltmak için yardımcı görevde bir otomatik kodlayıcı içeren Kapı Özyinelemeli Geçitler (GRU- ing. Gated Recurrent Units) ile alan adaptasyonu için bağımlılık ağacı tabanlı bir RNN önermektedir. Wang ve Pan (Wang & Pan, 2020) önceki çalışmayı (Wang & Pan, 2018) genişletir ve hem kelime özelliklerini hem de sözdizimsel baş ilişkilerini girdi olarak alan koşullu bir etki alanı çekişmeli ağını (adversarial network) entegre eder. Başka bir çalışmada (Pereg, Korat, & Wasserblat, 2020), BERT modeliyle birleştirilmiş bir öz dikkat mekanizmasına (a self-attention mechanism) dış dilsel bilgiyi dahil etmek için bir model önerilmiştir. Diğer bir alan adaptasyon yaklaşımı, gelecekteki öğrenmelere yardımcı olmak için geçmişte öğrenilen bilgi veya bilgilerden yararlanmayı amaçlayan yaşam boyu öğrenmedir (Shu, Liu, Xu, & Kim, 2016) (Shu, Xu, & Liu, 2017). Alanlar arası yaklaşımlar yalnızca etiketlenmiş verileri hedef alana transfer ederken, yaşam boyu öğrenme yaklaşımları

kaynak alanlardaki etiketlenmemiş verileri kullanabilir (Hai Z. , Chang, Kim, & Yang, 2013) (Quan & Ren, 2014) (Chen, Mukherjee, & Liu, 2014).

Dragoni ve arkadaşlarının çalışmasında (Dragoni, Federici, & Rexha, 2019), dilbilgisi bağımlılıklarının yapısı özellikler ve duygular arasındaki bağlantıları çıkarmak için analiz edilir. Özellik çıkarımı için bir dizi kural uygulanır. Luo ve arkadaşları (Luo, Huang, & Zhu., 2019), müşteri yorumlarından öne çıkan özelliklerin çıkarılmasıyla ilgili yeni bir problem tanımlamaktadır. Ayrıca, öne çıkan özelliklerin çıkarılması için yeni bir değerlendirme veri kümesi oluşturulmaktadır. Al-Smadi ve arkadaşları (Al-Smadi, Al-Ayyoub, Jararweh, & Qawasmeh, 2019) Arapça otel yorumlarının özellik tabanlı duygu analizi üzerine bir araştırma sunar. Çeşitli sınıflandırıcılar ile birlikte gözetimli öğrenme yaklaşımları fikir hedefi çıkarımı için çalıştırılır.

3. DUYGU ANALİZİ

Duygular neredeyse tüm günlük faaliyetlerimizin merkezinde yer almakta ve davranışlarımızı etkilemektedir. Başkalarının dünyayı algılama ve değerlendirme şekli bizim düşüncelerimizi ve yaptığımız seçimleri önemli ölçüde etkilemektedir. Bu nedenle bir konu hakkında karar vermemiz gerektiğinde başkalarının fikirlerine başvururuz. Bu durum sadece bireyler için değil, kurumlar için de geçerli olmaktadır.

Duygu kavramı ile birlikte, düşünce, fikir, davranış gibi kavramlar duygu analizinin, diğer bir deyişle, fikir madenciliğinin temel kavramlarıdır. İnternette sosyal medyanın ortaya çıkması ve gelişmesiyle, bloglar, forumlar, sosyal ağlar vb. sıklıkla kullanılan platformlar haline gelmiştir. Bu platformlarda paylaşılan veriler ile birlikte duygu içeren büyük miktarlarda veri elde edilmektedir. 2000’li yıllardan bu yana, duygu analizi doğal dil işleme konuları arasında popüler olan konular arasında yer almaktadır. Duygu analizi ayrıca veri madenciliği, metin madenciliği ve web madenciliği içerisinde sıklıkla karşımıza çıkmaktadır. Duygu analizi, insanların bir varlık veya onun bir özelliği hakkında ifade ettiği duygu ve düşünceleri analiz eden çalışma alanıdır. Burada varlık bir ürün veya servis olabileceği gibi, herhangi bir konu da olabilir. Tez içerisinde duygu analizi ve fikir madenciliği kavramlarının her ikisi de birbirinin yerine kullanılacaktır.

Duygu analizinin birçok çalışma alanı vardır. Şirketler ürün ve servisleri ile ilgili müşterilerden bilgi almak istemektedirler. Müşteriler ise bir ürünü satın almadan önce daha önce satın almış müşterilerin yorumlarına ihtiyaç duymaktadırlar. Geçmişte insanlar sadece çevresindeki insanların fikirlerine ulaşabilmekteydi. Şirketler ise müşterilerin fikirlerini alabilmek için anket vb. yöntemlere başvurmaktaydı. Bu yöntemler ancak sınırlı sayıda kişiye ulaşabilmeyi sağlıyordu. Günümüzde sosyal medyanın büyümesiyle insanların fikirlerini paylaşabileceği ortamlar ortaya çıkmıştır. Böylece, insanların veya şirketlerin karar verme süreçlerinde kullanabileceği verilere ulaşımı kolaylaştırmıştır. Bir ürün satın almak isteyen bir kişi ürünle ilgili birçok yoruma başvurabilmektedir. Şirketler ise anket vb. çalışmalar yapmadan birçok kişiye ulaşma imkânı elde etmiştir. Ancak, farklı kaynaklardan gelebilen yüksek miktarda yorumun manuel olarak analiz edilmesi ve özetlenmesi zor olmaktadır. Duygu analizi sistemlerinin gerekliliği bu noktada ortaya çıkmaktadır.

Duygu analizini detaylı bir şekilde incelemeden önce belirli kavramları ele almak gerekli olacaktır. Fikir tanımını yapmadan önce örnek bir yorum üzerinden analizler gerçekleştirilecektir.

Yazan: Ali Yazar

Tarih: 20 Nisan 2021

“(1) Altı ay önce bir Canon kamera satın aldım (2) Oldukça güzel bir ürün (3) Görüntü kalitesi muhteşem (4) Pil ömrü de yeterince uzun. (5) Ancak eşim onun kendisi için çok ağır olduğunu söylüyor.”

Canon ürünü hakkında yapılan yorum toplamda 5 cümleden oluşmaktadır. Yorum Canon kamera hakkında pozitif veya negatif olmak üzere birden fazla düşünce içermektedir. Cümlelere sırayla bakmak gerekirse, ilk cümlede herhangi bir düşünce bulunmamaktadır. İkinci cümle kamera hakkında genel bir pozitif düşünce ifade etmektedir. Üçüncü cümlede kameranın görüntü kalitesi hakkında, dördüncü cümlede ise pil ömrü hakkında pozitif düşünce belirtilmektedir. Beşinci cümlede ise kameranın ağırlığı konusunda negatif düşünce ifade edilmektedir. Bütün bu cümleler ışığında belirli çıkarımlar yapmak mümkündür. Bir düşüncenin iki temel bileşeni vardır. Bunlar düşüncenin hedefi ve duygusudur. Düşüncenin hedefi, hakkında düşünce ifade edilen herhangi bir varlık veya varlığın bir özelliği olabilmektedir. Örneğin, ikinci cümlede düşüncenin hedefi Canon kamera iken, üçüncü cümlede düşünce hedefi, kameranın görüntü kalitesidir. Duygu ise pozitif, negatif veya nötr olarak ifade edilir. Duygu ayrıca 1-5 aralığı vb. ölçeklerdeki sayısal değerler ile de temsil edilebilir.

Bir yorumdaki düşünceler tek bir kişi tarafından ifade edilebileceği gibi, birden fazla kişi de söz konusu olabilir. Örneğin, yukarıdaki yorumda (2), (3) ve (4) cümlelerindeki düşünceler Ali Yazar isimli yazara aitken (5) cümlesindeki düşünce, yazarın eşine ait bir düşüncedir. Düşüncenin sahibi olan kişi, fikir sahibi (opinion holder) olarak ifade edilir. Diğer bir bilgi ise tarih bilgisidir. Yorumun güncelliğini göstermesi açısından önemli bir bilgidir.

Bütün bu analizler ışığında düşüncenin tanımını yapmak gerekirse, düşünce 5 bileşenden oluşan yapısal bir biçimde ifade edilebilir. Düşünce, $(e_i, a_{ij}, s_{ijkl}, h_k, t_l)$ şeklinde gösterilebilir. Burada e_i , varlığın ismi; a_{ij} , e_i nin özelliğini, h_k düşünce sahibini, t_l yorum zamanını, s_{ijkl} e_i varlığına ait a_{ij} özelliği ile ilgili ifade edilen duyguyu belirtir. Eğer duygu

belirli bir özellik için değil de varlığın kendisi içinse ağı özelliği “GENERAL” olarak ifade edilir.

Duygu analizi doküman, cümle ve özellik olmak üzere 3 düzeyde ele alınmaktadır. Doküman düzeyinde amaç, dokümandaki genel duyguyu pozitif veya negatif olarak sınıflandırmaktır. Örneğin, bir ürün yorumundan ürün hakkındaki genel duygu çıkarılır. Her bir dokümanın tek bir varlık hakkında duygu ifade ettiği varsayılır. Bu düzey ayrıca doküman düzeyinde duygu sınıflandırma olarak da ifade edilir. Cümle düzeyinde analizde ise her bir cümle ayrı ayrı ele alınır. Cümlelerin pozitif, negatif veya nötr duygu içerip içermediği belirlenir. Cümle düzeyinde analiz, öznellik sınıflandırma (subjectivity classification) işlemiyle yakından ilişkilidir. Öznellik sınıflandırma işlemi ile cümleler nesnel veya öznel cümle olarak sınıflandırılır. Öznel cümleler öznel görüş ve düşünceleri ifade ederken, nesnel cümleler kesin ve gerçek bilgileri ifade eder. Doküman ve cümle düzeyinde analizde duygular sınıflandırılmasına rağmen fikrin hedefi net olarak belirlenmemektedir. Örneğin, bir ürün yorumunda ürün hakkında bir duygu belirlenirken, ürünün hangi özelliklerinin sevilip sevilmediği analiz edilmemektedir. Özellik düzeyinde analiz ise ürünün her bir özelliği için ayrı ayrı analiz yapmaktadır.

3.1. Doküman Düzeyinde Duygu Analizi

Duygu sınıflandırma duygu analizi içerisinde sık çalışılan konulardan biridir. Bir yorum dokümanını pozitif veya negatif olarak sınıflandırmayı amaçlar. Diğer bir ismi doküman düzeyinde duygu sınıflandırma veya analizdir. Bu analizde dokümanın tamamı üzerindeki genel duygu tespit edilmeye çalışılır, yani özellik olarak “GENERAL” kabul edilir. Dokümanın tek bir varlık hakkında olduğu ve tek bir düşünce sahibi tarafından yazıldığı varsayılır. Bu varsayım ürün yorumları için geçerli olmaktadır. Ürün yorumlarında genellikle tek bir kişi tarafından tek bir ürün hakkında yorum yapılmaktadır. Ancak blog vb. yerlerde yapılan yorumlarda tek bir yorum birden fazla ürün hakkında olabileceğinden doküman düzeyinde duygu sınıflandırma uygulanamaz.

Duygu sınıflandırma için gözetimli veya gözetimsiz yaklaşımlar uygulanmaktadır. Gözetimli yaklaşımlarda duygu sınıflandırma genel olarak 2 sınıflı bir sınıflandırma problem olarak ele alınır. Bazı çalışmalar ise nötr sınıfını da dahil etmektedirler. Ürün sitelerinde yer alan ürün yorumları eğitim için kullanılmaktadır. Bu yorumların yanında genellikle 1-5 aralığında belirli oylar verilmektedir. 4-5 puan alan yorumlar pozitif yorumlar, 1-2 puan alan yorumlar negatif yorumlar, 3 puan alan yorumlar ise nötr

yorumlar olarak değerlendirilmektedir. Duygu sınıflandırma problemi aslında klasik metin sınıflandırma problemlerine benzemektedir. Belirli konu başlıklarını sınıf olarak ele alarak sınıflandırma yapılan bir metin sınıflandırma probleminde konu ile ilgili kelimeler önemli olmaktadır. Duygu sınıflandırmada ise pozitif veya negatif düşünce belirten kelimeler önemlidir. İyi, kötü, muhteşem, korkunç gibi kelimeler bu tür kelimelere örnek verilebilir. Duygu sınıflandırma problemi metin sınıflandırma problemi olarak ele alındığında, Basit Bayes (NB- ing. Naïve Bayes), Destek Vektör Makineleri (SVM- ing. Support Vector Machines) gibi herhangi bir sınıflandırma modeli kullanılabilir. Kullanılacak özniteliklerin belirlenmesi, diğer metin sınıflandırma problemlerinde olduğu gibi duygu sınıflandırma problemi için de temel aşamalardan biridir. Kullanılan örnek öznitelikler şu şekildedir:

- *Terimler ve frekansları*: Kelimeler bireysel olarak veya n -gram, yani n tane sıralı kelimenin bir arada kullanılması, şeklinde öznitelik olarak kullanılmaktadır. Bazı durumlarda kelime sıraları da dikkate alınmaktadır. Kelimeler terim ağırlıklandırma yöntemleri kullanılarak ağırlıklandırılarak kullanılmaktadır.
- *POS etiketleri*: Her bir kelimenin sözcük türü de diğer bir önemli özniteliktir. Örneğin, sıfatlar duygu belirtmede önemli bir işarettir ve bu yüzden ayrı bir öznitelik olarak kullanılmaktadır. Kelimeler gibi sözcü türleri de bireysel olarak veya n -gram şeklinde kullanılabilir.
- *Duygu ifade eden kelimeler*: Bir dilde iyi ve kötü duyguları ifade eden belli başlı kelimeler vardır. Bu kelimeler manuel veya otomatik olmak üzere farklı şekillerde elde edilebilir. Duygu ifade eden kelimeler genellikle sıfat ve zarflardan oluşmakla birlikte isim veya deyimlerden de oluşabilir.
- *Duygu dönüştürücüler (sentiment shifters)*: Her dilde cümlelerin olumlu veya olumsuz anlamını değiştiren bazı kelimeler vardır. İngilizce’de “not” kelimesi Türkçe’de “değil” kelimesi en çok kullanılan duygu dönüştürücülerdir. Bu kelimelerin ayrı bir öznitelik olarak ele alınması önemlidir.

Gözetimli modellerde olduğu gibi gözetimsiz modellerde de duygu ifade eden kelimelerin kullanımı oldukça önemlidir. Gözetimsiz yaklaşımlar anlatılırken en temel çalışmalardan biri üzerinden örneklendirilecektir. (Turney, 2002). Bu yaklaşım 3 temel adımdan oluşmaktadır. İlk adımda ardışık kelime grupları belirli desenlere göre çıkarılır. Örneğin bu desenlerden biri ilk kelimenin sıfat, ikinci kelimenin isim olma durumudur.

“this piano produces beautiful sounds” örnek cümlesinde “beautiful sounds” kelime grubu desenlere göre çıkarılır. İkinci adımda PMI skoru hesaplanır. “beautiful sounds” kelime grubu ile “excellent” kelimesi arasındaki skor ve “poor” kelimesi arasındaki skor hesaplanır. Üçüncü adımda ise skorlar arasındaki fark hesaplanarak kelime grubunun duygu skoru belirlenmiş olur.

3.2. Cümle Düzeyinde Duygu Analizi

Cümle düzeyinde analizde her bir cümle ayrı ayrı analiz edilir. Ancak, doküman düzeyinde analiz ile cümle düzeyinde analiz arasında temel bir fark yoktur çünkü cümle dokümanın küçük şekli olarak düşünülebilir. 5 bileşenden oluşan gösterim burada kullanılmaz, çünkü cümle düzeyinde sınıflandırma ara adımdır. Fikir hedefi olan özellik ve varlıkların bilinmediği durumlarda cümlelerin pozitif veya negatif olarak sınıflandırılması sınırlı bir kullanıma sahiptir. Ancak özellik ve varlıkların bilinmesi durumunda cümle düzeyinde sınıflandırma anlamlı olmaktadır. Cümle düzeyinde duygu sınıflandırma problemi, 3 sınıflı sınıflandırma problemi veya 2 ayrı sınıflandırma problemi olarak ele alınmaktadır. İlk problem bir cümlelerin duygu içerip içermediğine göre sınıflandırılmasıdır ki, bu sınıflandırma öznel sınıflandırma olarak ifade edilir. İkinci problem ise duygu içeren bir cümle olarak tespit edilen cümlelerin pozitif veya negatif olarak sınıflandırılması problemidir.

3.3. Özellik Düzeyinde Duygu Analizi

Doküman düzeyinde veya cümle düzeyinde analiz uygulamalar için yeterli olmamaktadır çünkü bu düzeylerde fikir hedefi çıkarılamamakta ve fikir hedefleri için duygu ataması yapılamamaktadır. Her bir dokümanda tek bir varlık hakkında değerlendirme yapıldığı varsayılsa bile dokümanın pozitif duygu ifade etmesi, varlığın bütün özelliklerinin de pozitif olduğu anlamına gelmez. Daha detaylı bir analiz için öncelikle özelliklerin keşfedilmesi daha sonra ise her bir özellik için pozitif veya negatif duygunun belirlenmesi gerekir. Daha önce düşüncenin (e_i , a_{ij} , s_{ijkl} , h_k , t_l) şeklinde gösterilebildiğini görmüştük. Özellik düzeyinde analizde buradaki her bir bileşen çıkarılmaya çalışılır. Biz burada bu görevlerden ikisi olan özellik çıkarma ve duygu sınıflandırma aşamalarına odaklanacağız.

- *Özellik duygu sınıflandırma:* Bu görevde tespit edilen özelliklere ait duyguların pozitif, negatif veya nötr olarak sınıflandırılmasını sağlamaktır. Örnek cümlelere

bakıldığında, ilk cümlede “ses kalitesi” özelliğine ait düşünce pozitif, ikinci cümlede ise “GENERAL” olarak pozitif olarak sınıflandırma yapılmaktadır.

- *Özellik Çıkarımı:* Cümlede duygu ifade edilen özellikleri yani fikir hedeflerini çıkarmayı amaçlar. Örneğin, “Telefonun ses kalitesi oldukça iyi” cümlesinde “telefon” varlığına bağlı “ses kalitesi” özelliği görülmektedir. “Bu telefonu çok sevdim” cümlesinde ise telefon varlığına ait bir özellikten bahsedilmediğinden fikir hedefi “GENERAL” olarak kabul edilir.

3.3.1. Özellik duygu sınıflandırma

Özellik duygu sınıflandırma yaklaşımları gözetimli öğrenme yaklaşımları ve sözlük (lexicon) temelli yaklaşımlar olmak üzere ikiye ayrılır.

Gözetimli öğrenme yaklaşımları için, cümle düzeyinde kullanılan öğrenme temelli modeller uygulanabilir. Diğer bir çalışmada (Wei & Gulla, 2010), hiyerarşik sınıflandırma modeli önerilmiştir. Ancak buradaki temel mesele, her bir duygu ifadesinin kapsamının nasıl belirleneceği, yani cümle içerisinde her bir özelliğe ait kısmın nasıl belirleneceğidir. Temel yaklaşım bağımlılıkları belirlemek için bağıllık ayrıştırıcısı kullanmaktır. Örneğin, (Jiang, Yu, Zhou, Liu, & Zhao, 2011) çalışmasında bağıllık ayrıştırıcısını, sınıflandırma için özellik bağımlı öznitelikler kümesi üretmek için kullanılmıştır. Bir başka yaklaşımda (Boiy & Moens, 2009) bağıllık ağacında hedef özelliğe göre öznitelik pozisyonuna bağlı olarak her bir öznitelik ağırlıklandırılır. Gözetimli öğrenme etiketli verilerin varlığına bağlıdır. Bir alana ait etiketli verilerle eğitilmiş sınıflandırıcılar bir başka alanda başarısız olmaktadır.

Sözlük temelli yaklaşımlar gözetimsiz yaklaşımlardır ve birçok alana başarıyla uygulanabilir. Bu yaklaşımlar fikir sözlüğü, duygu kuralları ve bazı durumlarda bağıllık ağacı kullanırlar. Ayrıca duygu dönüştürücüleri, ama cümlecikleri (but clauses) ve duyguyu değiştiren diğer yapılar ele alınır. Sözlük temelli yaklaşımları örneklemek için burada örnek bir model sunulacaktır (Ding, Liu, & Yu, 2008). Bu model 4 adımdan oluşmaktadır. Model İngilizce diline göre tasarlandığından örnek cümle İngilizce olarak verilmiştir. Varlık ve özelliklerin bilindiği varsayılmaktadır.

1. *Fikir kelimelerini işaretler:* Cümle içerisindeki fikir kelimeleri fikir sözlükleri kullanılarak tespit edilip işaretlenir. Pozitif kelimeler [+1] ile negatif kelimeler [-1] ile işaretlenir. Örneğin, “the voice quality of this phone is not good, but the

battery life is long.” cümlesinde “good” kelimesi pozitif olarak işaretlenir. Cümlelerin son hali “the voice quality of this phone is not good [+1], but the battery life is long.” şeklindedir. Burada “long” kelimesi sözlükte yer almadığından işaretlenmez. Bu kelime bağlama bağlı (context dependent) fikir kelimesi olarak kabul edilir.

2. *Duygu dönüştürücülerini uygula:* Duygu dönüştürücülerini duyguyu değiştiren kelime veya ifadelerdir. İngilizcede “not, never, none, nobody, nowhere, neither, cannot” en yaygın kullanılan duygu dönüştürücüleridir. Örnek cümlede bulunan “not” kelimesi pozitif olan good [+1] kelimesini negatife çevirmiştir. Cümlelerin son hali, “the voice quality of this phone is not good [-1], but the battery life is long.” şeklindedir.
3. *Ama cümleciklerini ele al:* Karşıt kelimeler de cümle içerisindeki duygunun değişmesine neden olur. İngilizcede en sık kullanılan karşıt kelime olarak “but” kullanılır. Karşıt kelimedenden önce belirtilen duygu ile karşıt kelimedenden sonra belirtilen duygu birbirine zıttır. Her iki taraftan birinin duygusu bilinmediği durumlarda bilinmeyen tarafın duygusu tespit edilebilir. Örnek cümlede “but” ifadesinden önce negatif duygunun bulunması, “but” ifadesinden sonra pozitif duygu olacağı anlamına gelir. Burada “long” kelimesinin duygusu pozitif olarak tespit edilir. Cümlelerin son hali, “the voice quality of this phone is not good [-1], but the battery life is long [+1].” şeklindedir.
4. *Duyguları belirle:* Her bir özellik için duyguyu tespit edebilmek amacıyla duygu skorları hesaplanır. Skor hesaplanırken özelliğin cümledeki her bir fikir kelimesine uzaklığı dikkate alınır. Fikir kelimelerinin skorlarının fikir kelimelerinin özelliğe olan uzaklığına bölümleri toplamı, özellik için genel skoru vermektedir. Genel skorun işaretine göre özelliğe ait duygu pozitif veya negatif olarak belirlenmektedir.

3.3.2. Özellik çıkarımı

Özellik çıkarımı problemi bilgi çıkarımı problemi olarak ifade edilebilmektedir. Ancak duygu analizi probleminde cümle içerisinde duygu ifadeleri bulunmaktadır ve her bir duygu ifadesinin bir hedefi vardır. Bu hedefler cümle içerisinde özellikler olarak çıkarılmaktadır. Duygu ifadeleri pozitif veya negatif duyguyu ifade edebildiği gibi örtük özellikleri de belirtebilmektedir. Örneğin, “Bu araç çok pahalı” cümlesinde “pahalı” kelimesi “ücret” özelliğini ifade ettiğinden örtük özelliktir. “Telefonun kamerası oldukça

kaliteli” cümlesinde ise “kamera” özelliği açık özellik olarak görülmektedir. Bu tezde açık özellikler üzerinde çalışılmaktadır. Açık özelliklerin çıkarılmasında 4 temel yaklaşım bulunmaktadır. Bunlar; sık geçen isim ve isim ifadelerine bağlı çıkarım, duygu ve hedef ilişkilerini kullanarak çıkarım, gözetimli öğrenme kullanarak çıkarım ve konu modelleme (topic modeling) kullanarak çıkarımdır.

Sık geçen isim ve isim ifadelerine bağlı çıkarım yaklaşımları verilen bir alanda büyük miktarda yorum içerisinde isim ve isim öbeklerini özellik olarak çıkarmayı amaçlar (Hu & Liu, 2004). İsim ve isim öbekleri POS etiketleme işlemi kullanılarak tanımlanır. İsim ve isim öbeklerinin geçme sıklıkları hesaplanarak belirli sıklıkta olanlar özellik olarak seçilir. Buradaki temel düşünce insanların yorumlarda sık kullandığı ifadelerin önemli özellikleri ortaya çıkardığı düşüncesidir. Nadir olarak geçen isim ve isim öbekleri ise ya daha az önemli özellikleri ya da özellik olmayan kelimeleri göstermektedir. Diğer bir çalışmada (Popescu & Etzioni, 2007) bu algoritmanın geliştirilmesi amaçlanır. Bu amaçla bir varlıkla ilişkisi olmayan isim öbeklerini kaldırmayı dener. PMI kullanarak isim öbekleri ile varlık ifadeleri arasında skor hesaplar. Örneğin “camera” varlığı için varlık ifadelerine örnek olarak “of camera”, “camera has”, “camera comes with” verilebilir. Sonuç olarak PMI skorları çok düşük olan isim öbekleri elenmiş olur.

Diğer bir yaklaşımda ise duygu ve hedef ilişkileri kullanılmaktadır. İfade edilen duyguların bir hedefi olduğundan duygu ve hedef arasında bir ilişki vardır. Bu ilişki fikir hedefi olan özelliklerin çıkarılması için kullanılabilir çünkü fikir kelimeleri çoğu zaman tespit edilebilir. Bu model sık geçmeyen özelliklerin çıkarılmasında kullanılmıştır (Hu & Liu, 2004). Bir cümle sık geçen bir özelliğe sahip değilse ancak bazı fikir kelimelerine sahipse her bir fikir kelimesine en yakın isim ve isim öbekleri fikir hedefi olarak çıkarılır. Örneğin; “The software is amazing.” cümlesinde “amazing” kelimesinin fikir kelimesi olduğu bilindiğinden “software” özellik ve aynı zamanda fikir hedefi olarak çıkarılır. Bu ilişki temelli model, dokümanlardaki önemli özellikleri çıkarmak için yararlıdır çünkü herhangi bir duyguyla ilişkisi olmayan özellikler kaldırılarak önemli özellikler tespit edilebilir. Diğer bir çalışmada (Zhuang, Jing, & Zhu, 2006), bağlılık ayrıştırıcısı bağlılık ilişkilerini tanımlamak için kullanılır. Başka bir çalışmada ise, bağlılık fikri fikir kelimeleri ve özelliklerin birlikte çıkarılmasını sağlayan çift yayılım (double propagation) modeline genişletilir (Qiu, Liu, Bu, & Chen, 2011).

Bilgi çıkarımı için geçmişte birçok gözetimli öğrenme temelli algoritma önerilmiştir. Bunların çoğunu sıralı öğrenme (sequential learning) temelli algoritmalar oluşturur. Bu gözetimli modeller etiketli verilere ihtiyaç duyar. Cümleler üzerinde özellik, özellik değil şeklinde etiketleme (labeling) işlemi yapılır. En çok kullanılan sıralı öğrenme modelleri HMM (Rabiner, 1989) ve CRF (Lafferty, McCallum, & Pereira, 2001)'dir. Diğer bir çalışmada (Jin, Ho, & Srihari, 2009) özellik ve duygu ifadelerini çıkarmak amacıyla belirli desenleri ortaya çıkarmak için sözcük türetilmiş (lexicalized) HMM uygulanmıştır. Diğer bir çalışmada (Jakob & Gurevych, 2010) alan bağımsız bir çıkarım için farklı alanlardan veriler kullanılarak CRF ile sistem eğitilmiştir. Diğer bir yöntem ise sıralı desen madenciliği (sequential pattern mining) ile üretilmiş sıralı desen kuralları (sequential pattern rules) kullanarak özellik çıkarmaktır (Liu, Hu, & Cheng, 2005) (Jindal & Liu, 2006). Diğer bir çalışmada bir sınıflı SVM (one-class SVM) adında kısmi gözetimli öğrenme modeli kullanmışlardır (Yu, Zha, Wang, & Chua, 2011).

Özellik çıkarımında diğer bir yaklaşım konu modelleme kullanılmasıdır. İstatistiksel konu modelleri büyük miktardaki metin dokümanları içerisinden konuları keşfetme modeli olarak ortaya çıkmaktadır. Konu modelleme gözetimsiz bir öğrenme modelidir. Bu modelde her bir dokümanın konuların bir karışımından oluştuğu, her bir konunun ise kelimelerin bir olasılıksal dağılımından oluştuğu varsayılmaktadır. Konu modellemede iki temel model vardır, bunlar Olasılıksal Gizli Anlam Analizi (Probabilistic Latent Semantic Analysis) (PLSA) (Hofmann, 1999) ve Gizli Dirichlet Ataması (Latent Dirichlet allocation) (LDA) (Blei, Ng, & Jordan, 2003)'dir. Konu modellemedeki konular duygu analizinde özelliklere karşılık gelmektedir. Bu yüzden konu modelleme özelliklerin çıkarılmasında kullanılabilir. Ancak konular özellik ve fikir kelimelerinin her ikisini birden kapsamaktadır ve duygu analizi için bu ikisinin ayrılmaya ihtiyacı vardır. Ayrımı gerçekleştirmek için temel konu modelleme yaklaşımlarının genişletilmesi gerekmektedir.

4. ÖNERİLEN MODEL

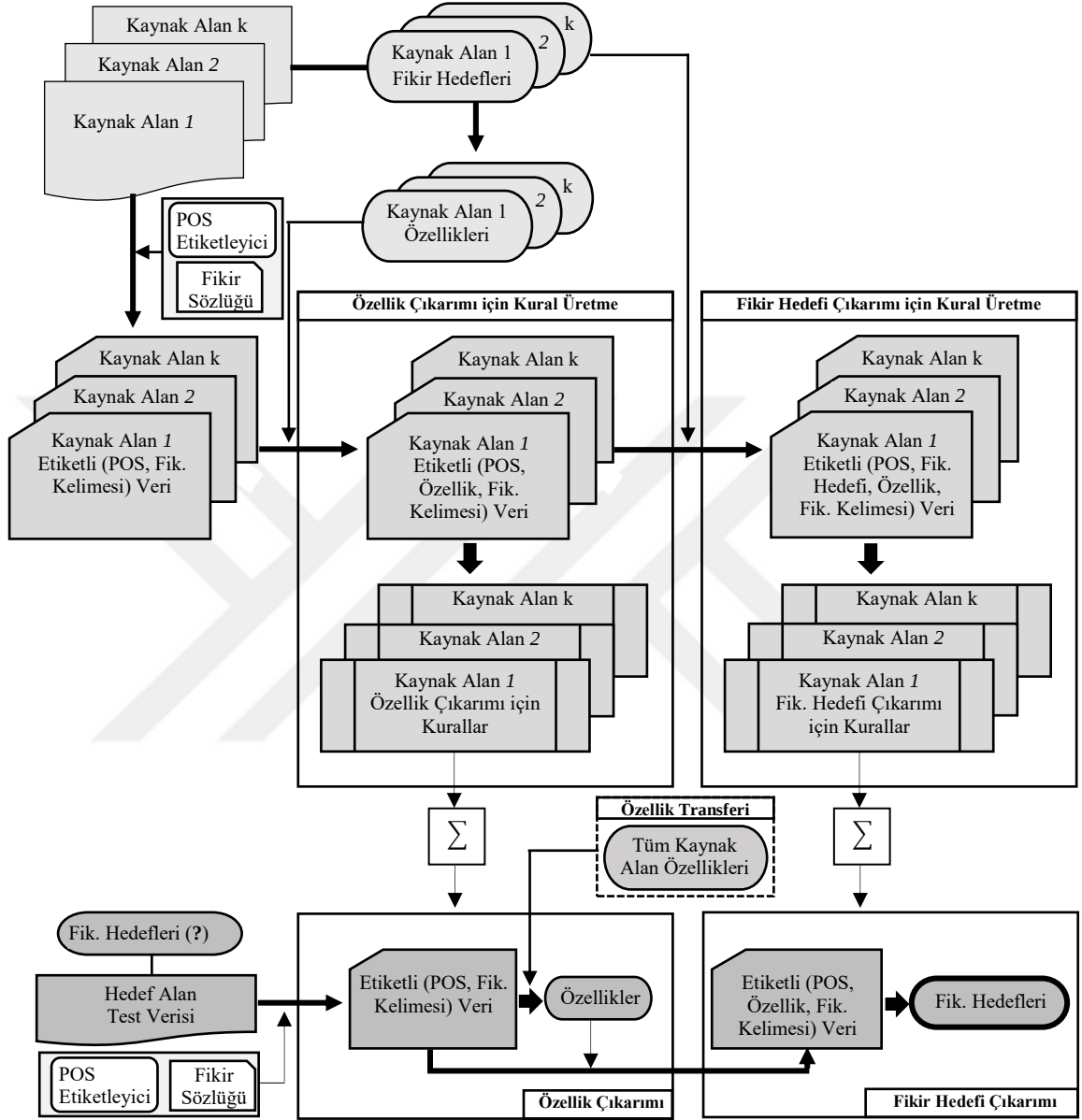
Fikir hedefi çıkarımı problemi, özellik çıkarımı ve fikir hedefi çıkarımı olmak üzere iki aşamada gerçekleştirilir. Özellik çıkarımı aşamasında, hedef alana ait verilen bir yorumdaki özellikler belirli özellik çıkarma kuralları kullanılarak çıkarılır. Bu aşamada ayrıca, kaynak alanlara ait özellikler hedef alana transfer edilir. Fikir hedefi çıkarımı aşamasında fikir hedefi çıkarımı kuralları kullanılarak ilk aşamada çıkarılan özellikler arasından fikir hedefleri belirlenir. Özellik çıkarma ve fikir hedefi çıkarma kuralları kaynak alanlar kullanılarak otomatik olarak üretilir. Daha sonra bu kurallar arasından destek (support) ve güven (confidence) değerleri kullanılarak seçilme işlemi yapılır. Önerilen model için genel bir şema Şekil 4.1’de gösterilmiştir. Özellik çıkarımı, fikir hedefi çıkarımı, özellik ve fikir hedefi çıkarımı için kural üretimi sonraki alt bölümlerde detaylıca açıklanmaktadır.

4.1. Özellik Çıkarımı

Bu aşamada yorum üzerinde öncelikle POS etiketleme işlemi uygulanır. Daha sonra bazı ön işleme (preprocessing) adımları uygulanır ve fikir kelimeleri etiketlenir. Yorum içerisindeki aday (candidate) özellikler etiket desenlerine (tag patterns) göre belirlenir ve özellik çıkarma kuralları kullanılarak her bir aday özellik için özellik skoru hesaplanır. En yüksek skora sahip olan aday özellikler özellik olarak çıkarılır. Aday özellikler, kaynak alanlardan transfer edilen özellikler arasında ise, bu özellikler skoruna bakılmaksızın özellik olarak kabul edilir.

4.1.1. POS etiketleme

POS etiketleyici metni analiz ederek metin içerisindeki her bir kelime için bir etiket ataması yapar. Böylece binlerce kelimeden oluşan kelime uzayı sınırlı sayıda etiketle temsil edilmiş olmaktadır. Bu işlem ile kelimeler gruplandırılarak kelimeler arasındaki ilişkilerin çıkarılması kolaylaştırılmaktadır. POS etiketleri ve açıklamaları Tablo 4.1’de verilmiştir. Ancak etiketleme işlemi bazı durumlarda istenilmeyen bilgi kayıplarına neden olabilmektedir. Örneğin, edatlar ve yan cümle bağlaçları metin içerisinde farklı rollere sahip olmalarına rağmen aynı “IN” etiketiyle temsil edilmektedir. Bilgi kaybını önlemek amacıyla, “IN” ve “DT” etiketleri için etiketleme işlemi uygulanmamıştır. Örnek bir yorum ve etiketlenmiş versiyonu Tablo 4.2a’da verilmiştir.



Şekil 4.1. Önerilen modelin genel şeması

Tablo 4.1. POS etiketleri ve açıklamaları

Etiket	Açıklama	Örnek	Etiket	Açıklama	Örnek
CC	Birleştirme bağlaçları (Coordinating conjunction)	and, or, but	PRP\$	İyelik zamiri (Possessive pronoun)	my, our, their
CD	Nicel sayılar (Cardinal number)	one, 2, 19.95	RB	Zarf (Adverb)	so, not, very
DT	Belirteçler (Determiner)	the, a, this	RBR	Karşılaştırma Zarfı (Comparative adverb)	better, more, less
EX	Varlık anlamı olan there (Existential there)	there	RBS	Üstünlük zarfı (Superlative adverb)	Most
FW	Yabancı kelime (Foreign Word)	i	RP	Edat (Particle)	up, out, of
IN	Edat veya Yantümce bağlacı (Preposition or subordinating conjunction)	with, of, from, if, because	SYM	Sembol (Symbol)	*, @, ok
JJ	Sıfat (Adjective)	good, other, few	TO	Karşı, için (To)	To
JJR	Karşılaştırma sıfatları (Comparative adjectives)	more, better, higher	UH	Ünlem (Interjection)	oh, eh, yes
JJS	Üstünlük sıfatları (Superlative adjectives)	most, best, biggest	VB	Fiil, yalın hal (Verb, base form)	be, have, want
LS	Liste ögesi işaretçisi (List item marker)	i	VBD	Fiil, geçmiş zaman (Verb, past tense)	was, did, went
MD	Kip (Modal)	may, will, can	VBG	İsim fiil veya şimdiki sıfat fiil (Verb, gerund or present participle)	being, looking
NN	İsim (tekil veya sayılamayan) (Noun, singular or mass)	everything, half, pm	VBN	Geçmiş sıfat fiil (Verb, past participle)	been, had, visited
NNS	İsim (çoğul) (Noun, plural)	prices, items, people	VBP	3. Tekil dışında şimdiki fiil (Verb, non-3rd person singular present)	have, do, are, know
NNP	Tekil özel isim (Singular proper noun)	friday, paris	VBZ	3. Tekil şimdiki fiil (Verb, 3rd person singular present)	has, does, is, comes
NNPS	Çoğul özel isim (Plural proper noun)	Americans	WDT	Soru belirleyici (Wh-determiner)	which, that, what
PDT	Önbelirteç (Predeterminer)	half, all, such	WP	Soru zamiri (Wh-pronoun)	what, who
POS	İyelik son eki (Possessive ending)	's	WP\$	İyelik soru zamiri (Possessive wh-pronoun)	Whose
PRP	Kişi zamiri (Personal pronoun)	we, you, they	WRB	Soru zarfı (Wh-adverb)	when, where, how

4.1.2. Ön işleme

Kelimeler arasındaki ilişkilerin çıkarılması için kelimeler sıralı ve ardışık bir biçimde ele alınmaktadır. İlişkilerin daha etkili bir şekilde çıkarılması için bazı ön işleme adımları uygulanmaktadır. Örnek yorum için ön işleme adımları Tablo 4.2b-2e’de sunulmuştur. Uygulanan ön işleme adımları şu şekildedir:

Parantezlerin Ele Alınması: İki ardışık kelime arasındaki parantezler ilişkinin çıkarılmasını zorlaştırmaktadır. Bu nedenle, parantezler kaldırılmış ve parantez içerisindeki ifadeler yorumun sonuna eklenmiştir.

Olumsuz İfadelerin Kaldırılması: Olumsuz ifadeler özellik çıkarımını etkilemediğinden yorum içerisinden kaldırılmaktadır. Öncelikle aşağıda belirtilen dönüşümler uygulanmakta, daha sonra ise “not” ve “n’t” ifadeleri kaldırılmaktadır.

{wo, MD} {n’t, RB} → {will, MD}	{does, VBZ} {not/n’t, RB} {operate, VB} → {operate, VBZ}
{ca, MD} {n’t, RB} → {can, MD}	{does, VBZ} {not/n’t, RB} {have, VB} → {has, VBP}
{do, VB} {not/n’t, RB} → {}	{did, VBD} {not/n’t, RB} {work, VB} → {work, VBD}

Deyimsel Fiillerin Ele Alınması: Birden fazla kelimeden oluşan deyim fiiller tek bir kelimeye dönüştürülür.

Sözcük Gruplarının Ele Alınması: Sözcük grupları, manuel olarak oluşturulmuş sözcük grupları listesi kullanılarak tek bir kelimeye dönüştürülür. Sözcük gruplarına örnek olarak; "in fact", "a few days ago", "i have to say" verilebilir.

4.1.3. Fikir kelimelerinin etiketlenmesi

Yorum içerisinde bulunan fikir kelimeleri, fikir kelime listesi kullanılarak “OP” etiketiyle etiketlenir. Fikir kelime listesi oluşturulurken fikir sözlüğü kullanılmaktadır (Hu & Liu, 2004). Fikir kelimesinin etiket (tag) bilgisi etikete (label) eklenmektedir çünkü farklı etikete (tag) sahip fikir kelimeleri yorum içerisinde farklı rollere sahip olmaktadır. Örneğin “good” kelimesi Tablo 4.2f’de gösterildiği gibi, POS etiketleme işleminde "good / JJ" şekline, fikir kelimesi etiketleme işleminde ise "good / JJ / OP-JJ" şekline dönüştürülmektedir.

Tablo 4.2. Örnek bir yorum için POS etiketleme, ön işleme ve fikir kelimelerinin etiketlenmesi

Örnek Yorum	“In fact, memory card (1 gb) is not good for a new phone today, it let down me.”
(a) POS Etiketleme	{in/in} {fact/NN} {/,} {memory/NN} {card/NN} { (/ -LRB-} {1/CD} {gb/NN} {)/-RRB-} {is/VBZ} {not/RB} {good/JJ} {/,} {for/for} {a/a} {new/JJ} {phone/NN} {today/NN} {it/PRP} {let/VBD} {down/RP} {me/PRP} {./}
(b) Parantezlerin ele alınması	{in/in} {fact/NN} {/,} {memory/NN} {card/NN} {is/VBZ} {not/RB} {good/JJ} {/,} {for/for} {a/a} {new/JJ} {phone/NN} {today/NN} {it/PRP} {let/VBD} {down/RP} {me/PRP} {./} {1/CD} {gb/NN}
(c) Olumsuz ifadelerin kaldırılması	{in/in} {fact/NN} {/,} {memory/NN} {card/NN} {is/VBZ}{good/JJ} {/,} {for/for} {a/a} {new/JJ} {phone/NN} {today/NN} {it/PRP} {let/VBD} {down/RP} {me/PRP} {./} {1/CD} {gb/NN}
(d) Deyimsel fiillerin ele alınması	{in/in} {fact/NN} {/,} {memory/NN} {card/NN} {is/VBZ} {good/JJ} {/,} {for/for} {a/a} {new/JJ} {phone/NN} {today/NN} {it/PRP} {let down/VBD} {me/PRP} {./} {1/CD} {gb/NN}
(e) Sözcük gruplarının ele alınması	{in~fact, in~NN} {/,} {memory/NN} {card/NN} {is/VBZ} {good/JJ} {/,} {for/for} {a/a} {new/JJ} {phone/NN} {today/NN} {it/PRP} {let down/VBD} {me/PRP} {./} {1/CD} {gb/NN}
(f) Fikir kelimelerinin etiketlenmesi	{in~fact, in~NN} {/,} {memory/NN} {card/NN} {is/VBZ} {good/JJ/OP-JJ} {/,} {for/for} {a/a} {new/JJ} {phone/NN} {today/NN} {it/PRP} {let down/VBD} {me/PRP} {./} {1/CD} {gb/NN}
POS etiketlenmiş ve fikir kelimeleri etiketlenmiş yorum	{in~fact, in~NN} {/,} {memory/NN} {card/NN} {is/VBZ} {good/JJ/OP-JJ} {/,} {for/for} {a/a} {new/JJ} {phone/NN} {today/NN} {it/PRP} {let down/VBD} {me/PRP} {./} {1/CD} {gb/NN}

4.1.4. Etiket desenlerinin ve aday özelliklerin çıkarılması

Yorum içerisindeki fikir hedefleri ve özellikler çeşitli etiketlerle (tags) temsil edilmektedirler. Bu etiketler etiket desenleri olarak adlandırılır. Tablo 4.3’te örnek yorumlara ait fikir hedefleri ve karşılık gelen etiket desenleri verilmiştir. Örneğin, yorum 1’de “viewfinder” fikir hedefine ait etiket deseni “NN” iken, yorum 2’de “use” için etiket deseni “VB” dir. Fikir hedefleri ayrıca sözcük öbekleri şeklinde bulunabilmektedir.

Tablo 4.3. Örnek yorumlar ve etiket desenleri

(1) <i>viewfinder</i> : NN ## the lens is visible in the viewfinder when the lens is set to the wide angle
(2) <i>use</i> : VB ## the camera is very easy to use.
(3) <i>made</i> : VBN ## it is made of plastic.
(4) <i>learning</i> : VBG ## learning how to use it will not take very long.
(5) <i>lens cover</i> : NN NN ## the lens cover is surely loose.
(6) <i>digital zoom</i> : JJ NN ## the digital zoom is a little blurry.
(7) <i>canon powershot g3</i> : NN NN NN ## i recently purchased the canon powershot g3.

Yorum (5), (6) ve (7) için etiket desenleri sırasıyla “NN NN”, “JJ NN” ve “NN NN NN” olarak belirlenmektedir. Bir etiket deseni kaynak alanların tamamında ortak olarak bulunuyorsa genel etiket deseni olarak seçilir. Genel etiket desenleri ayrıca yorum içerisindeki rollerine bağlı olarak etiket gruplarına ayrılmaktadır. Bu çalışmada kullanılan etiket grupları ve genel etiket desenleri Tablo 4.4’te listelenmektedir. Tablo 4.2’de POS etiketleme ve etiketleme işlemi yapılmış olan yorum içerisinde genel etiket desenlerine karşılık gelen kelime ve sözcük öbekleri aday özellikler olarak çıkarılır:

NN: memory, card, phone, gb, today; NN NN: memory card; JJ NN: new phone;
VBZ: is; JJ: new;

Tablo 4.4. Etiket grupları ve genel etiket desenleri

[NN]	
NN: camera, picture, sound, feel, look	JJ NN: digital camera, manual mode, photo quality
NN NN: picture quality, lens cover	NN NN NN: word processing program
[VB]	
VB: use, feel, look, cost	VBZ: looks, performs, runs
[JJ]	
JJ: remote, sound, outside	

4.1.5. Özellik budama

Aday özellikler belirli etiket desenlerine göre çıkarılmaktadır ancak bunlardan bazıları gereksiz ve yanlış olabilmektedir. Özellik budama işlemi, bu tür aday özelliklerin kaldırılmasını amaçlamaktadır. Örneğin, Tablo 4.2’deki örnek yoruma bakıldığında “today”, “new phone”, “new” ve “is” kelimeleri aday özellikler içerisinde yanlış veya gereksiz kelimelere örnek verilebilir. Budama kelime listesi her bir hedef alan için ayrı ayrı elde edilir. Kaynak alanlara ait POS etiketlenmiş, fikir kelimeleri ve özellikleri etiketlenmiş veriler budama için kullanılır. Öncelikle her bir kaynak alan için aday özellikler çıkarılır. Bir aday özellik kaynak alanların çoğunda ortak olarak bulunuyorsa ve bu aday özellik herhangi bir kaynak alanda özellik değilse, budama kelime listesine eklenir. Sonuç olarak, çıkarılan bir aday özellik veya onun alt kümesi aday kelime listesinde bulunuyorsa, bu aday özellik aday özellikleri içerisinde kaldırılmış olur. Yukarıdaki örnekte, “today”, “new”, “is” kelimeleri ve “new phone” kelimesinin alt kümesi olan “new” kelimesi budama kelime listesi içerisinde olduklarından bu kelimeler kaldırılır. Bütün kaynak alanlarda ortak olan budama kelime listesinin bazı örnekler Tablo 4.5’te listelenmiştir.

Tablo 4.5. *Tüm kaynak alanlar için örnek budama kelimeleri*

is	way	buy	have	same	find	first	point	regular	different
be	new	end	much	week	only	other	makes	average	something
do	big	etc	able	real	fact	today	anyone	purchase	understand

4.1.6. Kuralların uygulanması

Bir yorum birden fazla aday özellik içerebilir, ancak bunlardan bir veya birkaçı özellik olabilir. Belirli kurallar kullanılarak her bir aday özellik için özellik skoru hesaplanmaktadır. En yüksek skorlara sahip aday özellikler özellik olarak çıkarılmaktadır. Özellik çıkarımı için kural üretimi Bölüm 4.3.1’de açıklanmaktadır. Özellik skoru hesaplanacak olan her bir aday özellik için test kuralları çıkarılır. Öncelikle POS etiketleyici tarafından etiketlenmiş ve fikir kelimeleri etiketlenmiş yorum kelime ve etiket düzeyinde olacak şekilde kelime dizilerine dönüştürülür. Skoru hesaplanacak aday özellik “A-CAND” etiketiyle etiketlenir. Örnek olarak “memory card” aday özelliğinin skorunu hesaplayalım. Öncelikle “memory card” “A-CAND” etiketiyle etiketlenir. Daha sonra kelime ve etiket düzeyinde diziler Tablo 4.6’da gösterildiği gibi üretilir. Belirli bir

n değeri için, “A-CAND” etiketini içeren n uzunluğundaki diziler aday özellik için test kurallarıdır. Test kurallarının uygulanması için genel bir model sunulmuştur. Bu modele göre, kelime düzeyinde kurallar etiket düzeyinde kurallara göre daha önceliklidir. Daha uzun kural uzunluğuna sahip olan kurallar ise daha kısa kurallara göre daha önceliklidir. Bu önceliklere göre kural kümeleri oluşturulmuş ve bu kural kümelerinin uygulanma sırası belirlenmiştir. Kural kümelerinin uygulanması için sunulan model $n=4$ için Şekil 4.2’de gösterilmiştir. Şekil 4.2’de ayrıca “memory card” aday özelliği için örnek test kuralları ve karşılık gelen özellik çıkarma kuralları sunulmuştur. Özellik çıkarma kuralları oluşturulurken aday özelliğin yorum içerisinde karşılık gelen etiket desenine karşılık gelen etiket grubu <ASPECT> etiketi içerisine yazılmaktadır.

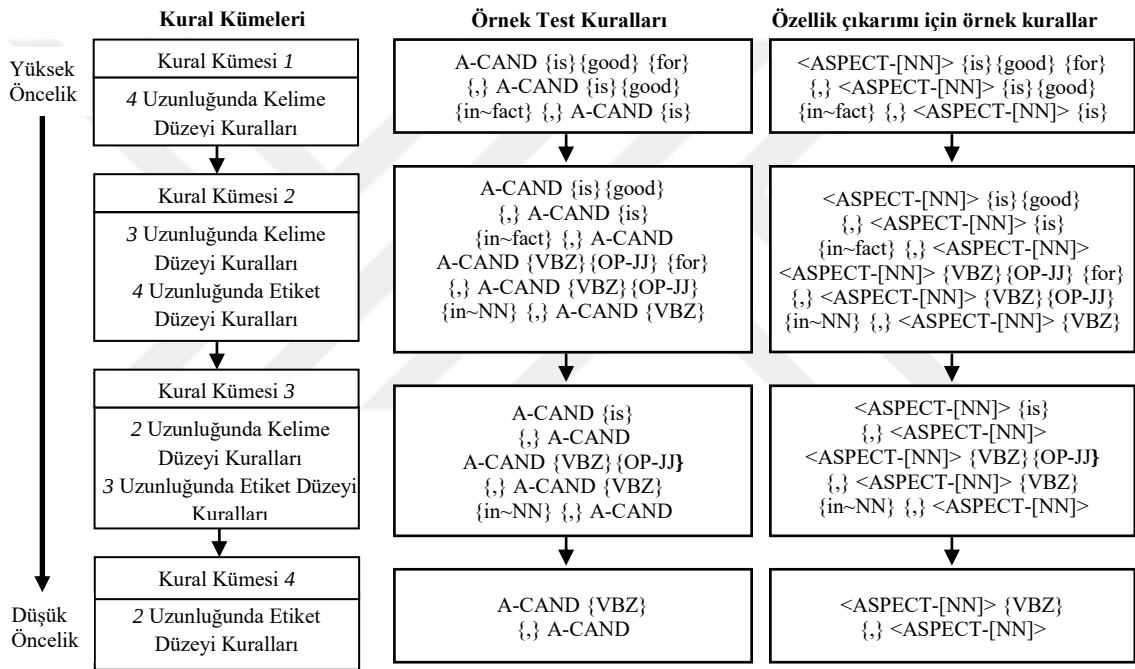
Tablo 4.6. “memory card” aday özelliği için dizilerin üretilmesi

Örnek Yorum	“In fact, memory card (1 gb) is not good for this phone, it let down me.”
POS Etiketlenmiş ve Fikir Kelimeleri	{in~fact, in~NN} {./,} {memory card/NN NN/A-CAND} {is/VBZ} {good/JJ/OP-JJ} {./,} {for/for} {a/a} {new/JJ} {phone/NN} {today/NN}
Etiketlenmiş Yorum	{it/PRP} {let down/VBD} {me/PRP} {./,} {1/CD} {gb/NN}
Kelime Düzeyi Dizi	{in~fact} {,} A-CAND {is} {good} {for} {a} {new} {phone} {today} {,} {it} {let~down} {me} {,} {1} {gb}
Etiket Düzeyi Dizi	{in~NN} {,} A-CAND {VBZ} {OP-JJ} {for} {a} {JJ} {NN} {NN} {,} {PRP} {VBD} {PRP} {./,} {CD} {NN}

Aday özellik için özellik skorunun hesaplanması şu şekilde gerçekleşmektedir: Test kurallarına karşılık gelen kurallar kaynak alanlar kullanılarak özellik çıkarma için üretilen kurallarla eşleştirilir. $n=4$ için örneklendirecek olursak, eşleştirme işlemi Kural Kümesi 1’den başlar ve Kural Kümesi 4’e kadar devam eder. Kural kümesi 1’deki herhangi bir kural üretilen kuralla eşleşmezse işlem Kural Kümesi 2’den devam eder ve bu şekilde ilerler. Eşleşen kural kümesindeki kuralların güven değerlerinin maksimum değeri, aday özellik için özellik skoru olarak tanımlanır. Kural kümesindeki herhangi bir kuralla eşleşme sağlanmazsa, aday özelliğin özellik skoru 0 olarak kabul edilir.

Her bir aday özellik için özellik skoru hesaplanır. Özellik çıkarma için kullanılacak olan minimum özellik skoru, özellik skorlarının maksimumu ile belirli bir tolerans oranı çarpılarak elde edilir. Örnek yorum için çıkarılan aday özellikler için hesaplanan skorların şu şekilde olduğunu varsayalım: {memory card: 0.90, memory:

0.85, phone:0.75, card: 0.70}. Tolerans oranı 0.90 olarak seçilirse, minimum özellik skoru $0.90 \times 0.90 = 0.81$ olarak hesaplanır. Minimum özellik skorundan daha yüksek skorlu aday özellikler, özellik olarak seçilir. Örnek yorumda, "memory card" ve "memory" özellik olarak seçilir. Ayrıca, kaynak alandaki özellikler hedef alana transfer edilir. "phone" aday özelliğinin özellik skoru minimum özellik skorundan düşük olmasına rağmen kaynak alandan transfer edildiği için özellik olarak seçilir. Sonuç olarak, "memory card", "memory" ve "phone" aday özellikleri özellik çıkarma aşamasında örnek yorum için özellik olarak çıkarılmaktadır.



Şekil 4.2. $n=4$ için test kurallarıyla birlikte kural kümeleri için genel model ve aday özellik için özellik çıkarma için karşılık gelen kurallar

4.2. Fikir Hedefi Çıkarımı

Verilen bir yorumdaki aday özelliklerden bazıları özellik çıkarma aşamasında özellik olarak çıkarılır. Bir yorumda en az bir fikir hedefi olduğu varsayılmaktadır. Buna göre, bir özellik çıkarılmışsa, bu özellik fikir hedefi olarak ele alınır. Eğer yorum birden fazla özelliğe sahipse, fikir hedefi çıkarma aşamasında bu özellikler arasından fikir hedefi/hedefleri seçilir. Her bir özellik bir fikir hedefi olmak zorunda değildir, çünkü her bir özellik için duygu ifade edilmeyebilir. POS etiketleme ve ön işleme adımları özellik

çıkarma aşamasındakiyle aynı şekilde gerçekleştirilir. Önceki aşamada çıkarılan özellikler özellik etiketleme işleminde “ASPECT” etiketiyle ve etiket grubuyla etiketlenir. Ancak, bir özellik başka bir özelliğin alt kümesi ise bir çakışma meydana gelmektedir. Örneğin, “memory” özelliği “memory card” özelliğinin alt kümesidir. Bu yüzden her iki özelliği aynı anda etiketlemek mümkün değildir. Çözüm olarak üst küme olan özellik “memory card” etiketleme için seçilmiştir. Fikir kelimelerinin etiketlenmesi özellik çıkarma aşamasında olduğu şekliyle gerçekleştirilir. Örnek olarak, “memory card” aday fikir hedefinin fikir hedefi skorunu hesaplayalım. Kelime ve etiket düzeyinde kelime dizileri üretilir ve aday fikir hedefi olan “memory card” OT-CAND etiketiyle etiketlenir. Diğer özellikler ise ASPECT etiketiyle etiketlenir. Test yorumu için üretilen diziler Tablo 4.7’de sunulmuştur. n uzunluğundaki diziler özellik çıkarma aşamasına benzer şekilde elde edilirler ve bu diziler Şekil 4.2’deki modele göre test kuralları olarak kullanılır. Aday fikir hedefi skorları, fikir hedefi çıkarımı için üretilen kurallar kullanılarak hesaplanır. Ayrıca fikir hedefi yorum içerisinde sıklıkla fikir kelimesine yakın konumlanmaktadır. Bu varsayıma göre, aday fikir hedefinin fikir kelimelerine yakınlığına göre fikir hedefi skoru değeri belirlenen bir miktarda (örneğin 0.10) artırılır. Fikir kelimelerine yakınlığı dikkate alınarak ve fikir hedefi çıkarım kurallarını kullanarak her bir fikir hedefi adayı için fikir hedefi skoru hesaplanır. Sonuç olarak, en yüksek skora sahip aday fikir hedefleri gerçek fikir hedefi olarak çıkarılır.

Tablo 4.7. "memory card" aday fikir hedefi için dizilerin üretilmesi

POS Etiketlenmiş ve Etiketlenmiş Yorum	{in~fact/in~NN} {/,} {memory card/NN NN/OT-CAND} {is/VBZ} {good/JJ/OP-JJ} {for/for} {a/a} {new/JJ} {phone/NN/ASPECT-[NN]} {today/NN} {/,} {it/PRP} {let~down/VBD} {me/PRP} {/,} {1/CD} {gb/NN}
Kelime Düzeyi Dizi	{in~fact} {/,} OT-CAND {is} {good} {for} {a} {new} ASPECT-[NN] {today} {/,} {it} {let~down} {me} {/,} {1} {gb}
Etiket Düzeyi Dizi	{in~NN} {/,} OT-CAND {VBZ} {OP-JJ} {for} {a} {JJ} ASPECT-[NN] {NN} {/,} {PRP} {VBD} {PRP} {/,} {CD} {NN}

4.3. Otomatik Kural Üretimi

Özellik çıkarma ve fikir hedefi çıkarma için kurallar, belirli ölçütlere göre üretilir ve seçilir. Alan bağımsız bir model oluşturmak için bir kuralın en az 2 alanda görülmesi

gerekmektedir. Destek ve güven değerleri her bir kural için hesaplanır. Kurallar seçilirken kullanılan minimum destek ve güven değerleri deneysel olarak belirlenir.

4.3.1. Özellik çıkarımı için otomatik kural üretimi

Özellik çıkarma kurallarını elde etmek için kaynak alanlara ait POS etiketlenmiş, fikir kelimeleri etiketlenmiş ve özellikleri etiketlenmiş yorumlara ihtiyaç duyulmaktadır. Önişleme ve POS etiketleme işlemlerinden sonra fikir kelimeleri etiketlenmektedir. Daha sonra, özellikler “ASPECT” etiketiyle etiketlenir ve “ASPECT” etiketinin yanına özelliğe ait etiket grup bilgisi de eklenir. Bir kaynak alana ait özelliklerin listesi her bir yoruma ait fikir hedeflerinin bir araya getirilmesiyle oluşur. Sonuç olarak, etiketlenmiş yorumlar kelime ve etiket düzeylerinde dizilere çevrilir. Daha sonra, “ASPECT” etiketini içeren n uzunluğunda diziler ve etiket desenlerini içeren n uzunluğunda diziler, kelime ve etiket düzeyinde diziler içerisinde elde edilir. Her bir kaynak alan için tüm yorumlardaki her bir n uzunluğundaki dizinin kaç defa geçtiği hesaplanır. “ASPECT” etiketini içeren her bir n uzunluğundaki dizi özellik çıkarımı için bir aday kuraldır. Özellik çıkarımı için kural üretiminde destek ve güven değerlerinin hesaplanması için gerekli formüller TD hedef alanı için aşağıda verilmiştir:

$TP = \{tp_1, tp_2, \dots, tp_g\}$, R kuralı için etiket desenleri kümesi ve $TG = \{tg_1, tg_2, \dots, tg_h\}$ etiket grupları kümesi olsun. Burada her bir etiket grubu tg_i etiket desenlerinin alt kümesidir. $SD = \{SD_1, SD_2, \dots, SD_k\}$ kaynak alanların kümesi ve TD özellik çıkarımı için hedef alan olsun.

$S_A = \{\text{ASPECT-TG}, (\text{kelime|etiket})\}$, diziyi oluşturan elemanlar kümesi olsun ve n -uzunluğunda özellik dizisi s_A , yalnızca bir ASPECT-TG etiketi ve bir veya daha fazla kelime veya etiket (tag) içeren elemanların sıralı listesidir ve $n \geq 2$ olmak üzere $\langle i_{A1}i_{A2} \dots i_{An} \rangle$ ile gösterilir. $S_T = \{\text{TG}, (\text{kelime|etiket})\}$, diziyi oluşturan elemanlar kümesi olsun ve n -uzunluğunda etiket dizisi s_T , yalnızca bir TG etiketi ve bir veya daha fazla kelime veya etiket (tag) içeren elemanların sıralı listesidir ve $n \geq 2$ olmak üzere $\langle i_{T1}i_{T2} \dots i_{Tn} \rangle$ ile gösterilir. Özellik dizisinde ASPECT-TG etiketinin pozisyonu ile etiket dizisinde TG etiketinin pozisyonu eşit olsun ve bu dizilerdeki etiket dışındaki elemanlar aynı olsun.

R_A , TD için özellik çıkarımı için kural olsun. Buna göre, R_A kuralı için destek ve güven değerleri Denklem 1 ve 2'ye göre hesaplanır:

$$destek(R_A) = \frac{\sum_{j=1}^k Frekans(s_A) \text{ in } SD_j}{\sum_{j=1}^k \#Cümle \text{ in } SD_j} \quad (1)$$

$$güven(R_A) = \frac{\sum_{j=1}^k \left\{ \frac{Frekans(s_A)}{Frekans(s_T)} \text{ for } SD_j \right\}}{k} \quad (2)$$

Sayısal bir örnek olarak, D4 hedef alan olmak üzere, D1-D3 kaynak alanları için bazı örnek kurallara ait dizilerin frekansları Tablo 4.8’de verilmiştir. Kaynak alanlara ait cümle sayıları 100, 200, 500 olarak varsayılmıştır. Özellik çıkarımı için üretilen bir kuralın destek ve güven değerleri Tablo 4.9’da gösterildiği şekilde hesaplanabilir.

Tablo 4.8. D4 hedef alan olmak üzere, D1-D3 kaynak alanları için bazı örnek kurallara ait dizilerin frekansları

Dizi	D1	D2	D3
the [NN] is	11	32	80
the ASPECT-[NN] is	10	30	75
the OP. TARGET-[NN] is	9	25	50

Tablo 4.9. Özellik çıkarımı için üretilen örnek bir kuralın destek ve güven değerlerinin hesaplanması

Örnek Kural 1: the <ASPECT-[NN]> is

Etiket Grubu: NN

Etiket Desenleri: {NN, NN NN, JJ NN, NN NN NN}

$$destek(\text{Örnek Kural 1}) = \frac{10 + 30 + 75}{100 + 200 + 500} = 0.144$$

$$güven(\text{Örnek Kural 1}) = \frac{10/11 + 30/32 + 75/80}{3} = 0.928$$

4.3.2. Fikir hedefi çıkarımı için otomatik kural üretimi

Kaynak alanlara ait POS etiketlenmiş, fikir kelimeleri etiketlenmiş ve fikir hedefi etiketlenmiş yorumlar kullanılarak fikir hedefi kuralları çıkarılır. Etiketli verilerdeki her bir yorum için fikir hedefleri manuel olarak çıkarılmaktadır. Ön işleme, POS etiketleme ve fikir kelimelerinin etiketlenmesi tamamlandıktan sonra yorum içerisindeki fikir hedefleri “OP.TARGET” etiketiyle etiketlenir. Etiketleme işlemi yapılırken ayrıca “OP.TARGET” etiketinin yanına fikir hedefinin etiket grubu da eklenir. Özellikler de

özellik çıkarımı aşamasındakine benzer şekilde “ASPECT” etiketiyle etiketlenir. Diziler üretildikten ve n uzunluğundaki diziler elde edildikten sonra “OP.TARGET” veya “ASPECT” etiketlerini içeren n uzunluğundaki dizilerin sayısı tüm kaynak alanlar için hesaplanır. Fikir hedefi çıkarımı için kural üretiminde destek ve güven değerlerinin hesaplanması için gerekli formüller TD hedef alanı için aşağıda verilmiştir:

$TP = \{tp_1, tp_2, \dots, tp_g\}$, R kuralı için etiket desenleri kümesi ve $TG = \{tg_1, tg_2, \dots, tg_h\}$ etiket grupları kümesi olsun. Burada her bir etiket grubu tg_i etiket desenlerinin alt kümesidir. $SD = \{SD_1, SD_2, \dots, SD_k\}$ kaynak alanların kümesi ve TD fikir hedefi çıkarımı için hedef alan olsun.

$S_{OT} = \{OP.TARGET-TG, (kelime|etiket)\}$, diziyi oluşturan elemanlar kümesi olsun ve n -uzunluğunda özellik dizisi s_{OT} , yalnızca bir OP. TARGET-TG etiketi ve bir veya daha fazla kelime veya etiket (tag) içeren elemanların sıralı listesidir ve $n \geq 2$ olmak üzere $\langle i_{OT1} i_{OT2} \dots i_{OTn} \rangle$ ile gösterilir. $S_A = \{ASPECT-TG, (kelime|etiket)\}$, diziyi oluşturan elemanlar kümesi olsun ve n -uzunluğunda özellik dizisi s_A , yalnızca bir ASPECT-TG etiketi ve bir veya daha fazla kelime veya etiket (tag) içeren elemanların sıralı listesidir ve $n \geq 2$ olmak üzere $\langle i_{A1} i_{A2} \dots i_{An} \rangle$ ile gösterilir. Özellik dizisinde OP. TARGET-TG etiketinin pozisyonu ile etiket dizisinde ASPECT-TG etiketinin pozisyonu eşit olsun ve bu dizilerdeki etiket dışındaki elemanlar aynı olsun.

R_{OT} , TD için fikir hedefi çıkarımı için kural olsun. Buna göre, R_{OT} kuralı için destek ve güven değerleri Denklem 3 ve 4’e göre hesaplanır:

$$destek(R_{OT}) = \frac{\sum_{j=1}^k Frekans(s_{OT}) \text{ in } SD_j}{\sum_{j=1}^k \#Cümle \text{ in } SD_j} \quad (3)$$

$$güven(R_{OT}) = \frac{\sum_{j=1}^k \left\{ \frac{Frekans(s_{OT})}{Frekans(s_A)} \text{ for } SD_j \right\}}{k} \quad (4)$$

Sayısal bir örnek olarak, D4 hedef alan olmak üzere, D1-D3 kaynak alanları için bazı örnek kurallara ait dizilerin frekansları Tablo 4.8’de verilmiştir. Kaynak alanlara ait cümle sayıları 100, 200, 500 olarak varsayılmıştır. Fikir hedefi çıkarımı için üretilen bir kuralın destek ve güven değerleri Tablo 4.10’da gösterildiği şekilde hesaplanabilir.

Tablo 4.10. Fikir hedefi çıkarımı için üretilen örnek bir kuralın destek ve güven değerlerinin hesaplanması

Örnek Kural 2: <i>the <OP.TARGET-[NN]> is</i>
Etiket Grubu: <i>NN</i>
Etiket Desenleri: <i>{NN, NN NN, JJ NN, NN NN NN}</i>
$destek(\text{Örnek Kural 2}) = \frac{9 + 25 + 50}{100 + 200 + 500} = 0.105$
$güven(\text{Örnek Kural 2}) = \frac{9/10 + 25/30 + 50/75}{3} = 0.800$

4.4. Katkılar

Önerilen modelin literatüre birçok katkısı bulunmaktadır. Öncelikle, önceki çalışmaların birçoğunda karıştırılan özellik ve fikir hedefi kavramları açıkça tanımlanmıştır. Bu tanımlamalara dayanarak fikir hedefi çıkarımı problemi, özellik çıkarımı ve fikir hedefi çıkarımı olmak üzere iki aşamada ele alınmıştır. Literatürdeki çalışmaların çoğu yorumlardaki sadece fikir hedeflerini etiketlerken, hakkında duygu ifade edilmeyen özellikleri dikkate almamaktadır. Bizim çalışmamızda ise yorumlardaki özelliklerin kullanımı sağlanmış, ayrıca kaynak alanlardaki özellikler özellik çıkarma aşamasında hedef alana transfer edilmiştir. Önceki çalışmalardan farklı olarak, POS etiketleme işleminde bazı etiketler dahil edilmemiş, böylece etiketlemenin neden olduğu bilgi kaybı azaltılmıştır. Yorum içerisindeki fikir kelimelerinin, özelliklerin ve fikir hedeflerinin etiketlenmesi yapılırken, tanımlanmış etiket grupları da dikkate alınmıştır. Bu sayede kelimenin cümledeki kullanımına göre etiketler özelleştirilmiştir.

5. DENEYSEL ÇALIŞMA

Önerilen modelin özellik ve fikir hedefi çıkarımındaki başarısını değerlendirmek amacıyla deneysel çalışmalar gerçekleştirilmiştir. Bu bölümde, deneylerde kullanılan veri kümeleri detaylı özellikleriyle açıklanmıştır. Ayrıca, deneyler için parametre değerleri ve değerlendirme için kullanılan ölçütler belirtilmiştir.

5.1. Veri Kümeleri

Deneyler, literatürde sık kullanılan 7 veri kümesi üzerinde gerçekleştirilmiştir (Hu & Liu, 2004b; Pontiki et al., 2014). Bu veri kümelerine ait özellikler Tablo 5.1’de özetlenmiştir. Bu çalışma özellik çıkarma problemini açık özellikler için ele aldığından örtülü özelliğe sahip cümleler kaldırılmıştır. Cümle içerisinde doğrudan bulunamayan fikir hedefleri tanımlanamayan (unidentified) fikir hedefi olarak ifade edilmiş ve “UI” şeklinde kısaltılmıştır. Tanımlanamayan fikir hedefleri için örnek yorumlar aşağıda verilmiştir:

- *picture quality##downloading pictures is quick and a breeze, and the quality is astounding.*
- *picture, ease of use##it takes excellent pics and is very easy to use, if you read the manual.*
- *no-off button##the on-off button feels somewhat flimsy, and has an awkward ...*
- *speakerphone##i like the speakphone function a lot.*

5.2. Deneysel Parametreler

Önerilen model, 7 veri setinden her birini hedef alan olarak ele alırken, geri kalan 6 tanesi kaynak alan olarak kullanılmaktadır. Fikir hedefi çıkarımı için alanlar arası çalışmalarda, bilinmeyen hedef alan problemiyle ilgili çalışmalar oldukça sınırlıdır. Deneysel sonuçlar, alanlar arası çıkarım modeli olan CD-ALPHN (Marcacini et al., 2018) ve iyi bilinen 5 sınıflandırıcı ile karşılaştırılmıştır. Bu sınıflandırıcılar, karar ağacı (DT- ing. decision tree), k-en yakın komşu (kNN- ing. k-nearest neighbor), çok katmanlı algılayıcı (MLP- ing. multi-layer perceptron), NB ve SVM sınıflandırıcılarıdır. Kural seçiminde kullanılan minimum destek ve güven değerleri deneysel olarak belirlenmektedir. Deneyler destek değerleri için {0.0001-0.02} aralığında, güven değerleri için {0.10, 0.20, ..., 0.90} değerleri için gerçekleştirilmiştir. Bir diğer parametre tolerans oranıdır. Bu oran verilen bir yorumdan birden fazla özellik veya fikir hedefi çıkarılmasına izin vermektedir. Tolerans oranı için {0.10, 0.20, ..., 0.90} değerleri kullanılmıştır. Fikir hedefleri bir veya birden fazla kelimedenden oluşabilir ve Tablo 4.4’te gösterildiği gibi farklı etiket desenleriyle ifade edilebilir. Önerilen modelin uygun bir

şekilde karşılaştırılabilmesi için deneysel sonuçlar hem tüm veri kümeleri kullanılarak hem de farklı alt kümeler kullanılarak elde edilmiştir. Veri kümeleri için 3 alt küme oluşturulmuş ve bunlar: sadece NN grubundaki (NN, NN NN, JJ NN, NN NN NN) etiket desenlerini içeren yorumlar kümesi, tanımlanmayan yorumların haricindeki yorumlar kümesi ve bu iki kümenin kesişiminden elde edilen yorumlar kümesidir.

Tablo 5.1. Veri kümeleri ve özellikleri

ID	Veri kümesi	#Yorum	#Cümle	#Özellik	#Fikir hedefi	#Fikir hedefi: #Cümle	#Etiket Deseni: #Fikir hedefi
D1	Dijital Kamera 1	45	211	98	252	1:179, 2:24 3:7 4:1	NN:163; NN NN:25; JJ NN:16; NN NN NN:1; VB:10; VBZ:3; JJ:1; OTHER:10; UI:23
D2	Dijital Kamera 2	34	143	72	178	1:114, 2:23 3:6	NN:104; NN NN:31; JJ NN:14; NN NN NN:1; VB:4; VBZ:0; JJ:0; OTHER:3; UI:21
D3	Telefon	41	237	107	295	1:190, 2:39 3:6, 4:1 5:1	NN:212; NN NN:40; JJ NN:10; NN NN NN:0; VB:2; VBZ:1; JJ:2; OTHER:4; UI:24
D4	MP3 Player	95	615	178	715	1:551, 2:45, 3:7, 4:10 6:1, 7:1	NN:470; NN NN:77; JJ NN:32; NN NN NN:1; VB:27; VBZ:7; JJ:11; OTHER:21; UI:69
D5	DVD Player	99	263	108	337	1:202, 2:49 3:11, 4:1	NN:176; NN NN:36; JJ NN:12; NN NN NN:3; VB:24; VBZ:12; JJ:13; OTHER:14; UI:47
D6	Bilgisayar	3048	1492	956	2373	1:931, 2:355 3:141, 4:43, 5:10, 6:6, 7+:6	NN:1304; NN NN:444; JJ NN:122; NN NN NN:41; VB:81; VBZ:38; JJ:22; OTHER:297; UI:24
D7	Restoran	3044	2023	1215	3699	1:1023; 2:571; 3:270; 4:105; 5:30, 6:15; 7+:9	NN:2682; NN NN:375; JJ NN:201; NN NN NN:40; VB:11; VBZ:7; JJ:20; OTHER: 360 UI:3

5.3. Değerlendirme Ölçütleri

Önerilen modelin performansını değerlendirmek için F1-skor başarı metriği kullanılmıştır. Değerlendirme aşamasında, çıkarılan fikir hedefi/özellik ile gerçek fikir hedefi/özellik karşılaştırılırken belirli kurallar belirlenmiştir. Özellik/fikir hedefi çıkarımı işlemi 2 durum için doğru sayılmıştır. İlk durum, gerçek fikir hedefinin/özelliğın çıkarılan fikir hedefinin/özelliğın tekil formu olmasıdır. İkinci durum ise, gerçek fikir hedefinin/özelliğın çıkarılan fikir hedefinin/özelliğın alt kümesi olmasıdır. Aşağıdaki yorumlar doğru çıkarım için verilen örneklerdir.

- *picture ## the highest optical zoom pictures are perfect.*
Extracted: *pictures*
- *camera ## this is a great camera for you!*
Extracted: *camera*
- *lcd ## the moveable lcd screen is great.*
Extracted: *lcd screen*
- *depth## i can live with is the extreme depth of field obtainable ...*
Extracted: *extreme depth*
- *customer service ##... their customer service line is always busy.*
Extracted: *customer service line*

Diğer durumlarda ise çıkarılan özellikler/fikir hedefleri yanlış sayılmıştır. Aşağıdaki yorumlar, yanlış çıkarım için verilen örneklerdir.

- *four megapixel ## four megapixels is great.*
Extracted: *megapixels*
- *zoomed image ##... the zoomed images are quite impressive.*
Extracted: *images*
- *pickels and slaw ## lunch came with pickels and slaw, no extra charge.*
Extracted: *pickels, slaw*
- *glass of wine ## ... let me finish my glass of wine before offering another.*
Extracted: *wine*
- *picture quality ## great quality picture and features.*
Extracted: *quality picture*

6. DENEY SONUÇLARI VE TARTIŞMA

Özellik çıkarımı ve fikir hedefi çıkarımı için elde edilen en iyi sonuçlar Tablo 6.1’de sunulmuştur. Sadece kural transferi uygulanması durumunda, D1 ve D5 alanları için özellik çıkarımı sonuçlarının fikir hedefi çıkarımına göre daha iyi olduğu görülmektedir. Diğer alanlarda ise kural transferi için fikir hedefi çıkarımının daha başarılı olduğu görülmektedir. Kural transferi ve özellik transferinin birlikte uygulanması durumunda, D7 alanı hariç diğer alanlarda özellik çıkarımı fikir hedefi çıkarımından daha iyi sonuçlar üretmektedir. Ayrıca, deneylerde kural transferi ve özellik transferinin sonuçlara olan etkisinin görülmesi amaçlanmıştır. Özellik çıkarımı, sadece kural transferi ile gerçekleştirildiğinde, en iyi sonuç D7 alanı için elde edilmektedir. Hem kural hem özellik transferi uygulandığında ise özellik çıkarımında en iyi sonucun D2 alanında elde edildiği görülmektedir. Sonuçlar göz önüne alındığında, özellik transferinin hem özellik çıkarımının hem de fikir hedefi çıkarımının başarısını arttırdığı görülmektedir. Ancak, özellik transferinin özellik çıkarımına etkisi daha fazladır. D1-D6 arasındaki alanlar teknolojik ürünlere ait alanlar olduğundan ortak özelliklere sahip olduğu söylenebilir. Ortak özellik içeren bu alanlarda özellik transferi, özellik çıkarma başarısını daha çok etkilemektedir. Ancak, restoran yorumlarını içeren D7 alanı diğer alanlardan daha farklı özellikler içerdiğinden, özellik çıkarımının başarısı özellik transferinden ziyade kural transferine bağlıdır. Özellik transferinin fikir hedefi çıkarımı üzerine etkisine bakıldığında, D1 ve D4 alanları için fikir hedefi çıkarımının başarısının kayda değer bir biçimde arttığı, diğer alanlardaki etkisinin ise daha az olduğu görülmektedir. D7 alanı için kural transferi, başarılı bir fikir hedefi çıkarımı için yeterlidir. Fikir hedefi çıkarımında, hem kural transferi ile hem de kural ve özellik transferi birlikte uygulandığında en iyi sonuç, D7 alanı için elde edilmektedir. Tablo 6.2’de belirtilen parametreler, bütün etiket desenleri için kural ve özellik transferi yapılması durumunda yapılan deneylerde elde edilen en iyi sonuçları sağlayan parametrelerdir. Her bir alan için en iyi sonuçlar, farklı parametrelerle elde edilmiştir. Özellik çıkarımında D1 alanı için min. güven değeri 0.9000 olarak belirlenirken, diğer durumlarda güven değerlerinin 0.3000-0.6000 aralığında olduğu görülmektedir. Güven değerleri yüksek seçildiğinde, yeterince kural üretilemediğinden en iyi sonuçların düşük güven değerleri seçildiğinde elde edildiği gözlemlenmektedir. Özellik çıkarımı ve fikir hedefi çıkarımı sonuçlarının tamamında destek değerleri incelendiğinde, en düşük değerin 0.0005, en yüksek değerin 0.0200 olduğu görülmektedir. Destek değerleri, en iyi sonuçlar için her bir alanda farklı değerler

almaktadır. Dolayısıyla, destek değerlerinin hangi durumda düşük veya yüksek çıkacağı ile ilgili belirli bir çıkarım yapılamamaktadır. Ayrıca cümle başına düşen fikir hedefi sayılarının fazla olduğu D4-D7 alanlarında tolerans oranının daha küçük olduğu, fikir hedefi sayılarının az olduğu D1-D3 alanlarında ise tolerans oranının büyük olduğu görülmektedir.

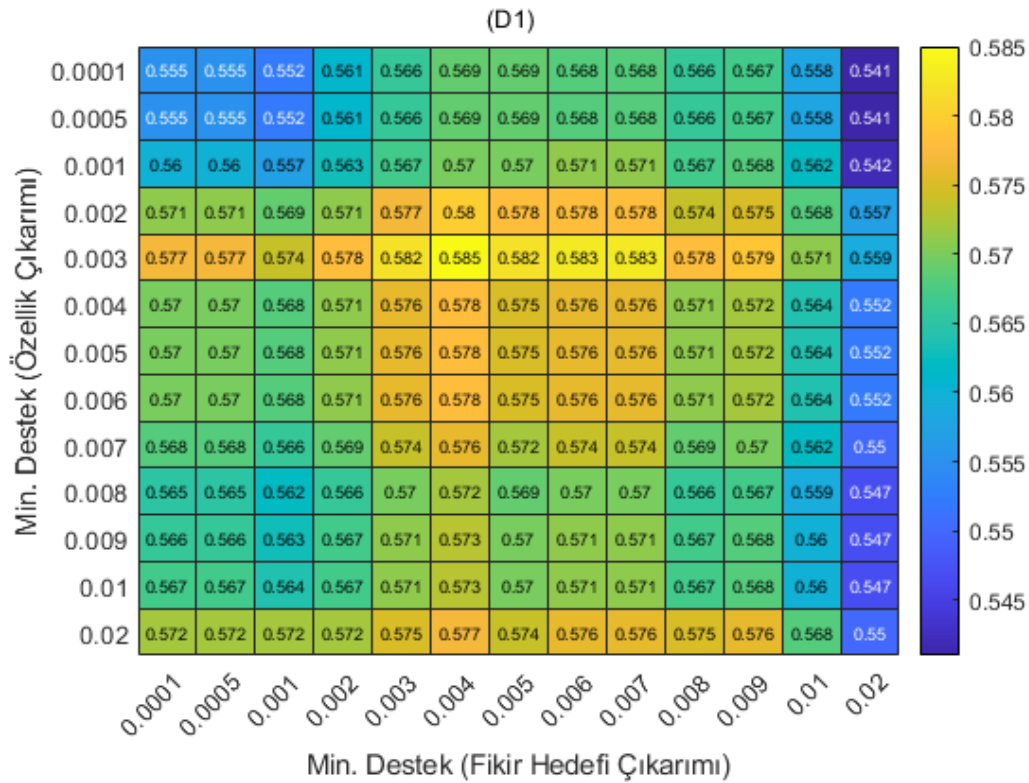
Tablo 6.1. Önerilen model ile elde edilen en iyi sonuçlar (F1-skor)

Özellik Çıkarımı			Fikir Hedefi Çıkarımı	
Hedef Alan	Kural Transferi	Kural Transferi + Özellik Transferi	Kural Transferi	Kural Transferi + Özellik Transferi
D1	0.538	0.726	0.515	0.585
D2	0.547	0.731	0.605	0.612
D3	0.632	0.692	0.634	0.642
D4	0.451	0.662	0.518	0.545
D5	0.508	0.700	0.468	0.476
D6	0.518	0.585	0.531	0.542
D7	0.635	0.643	0.682	0.681

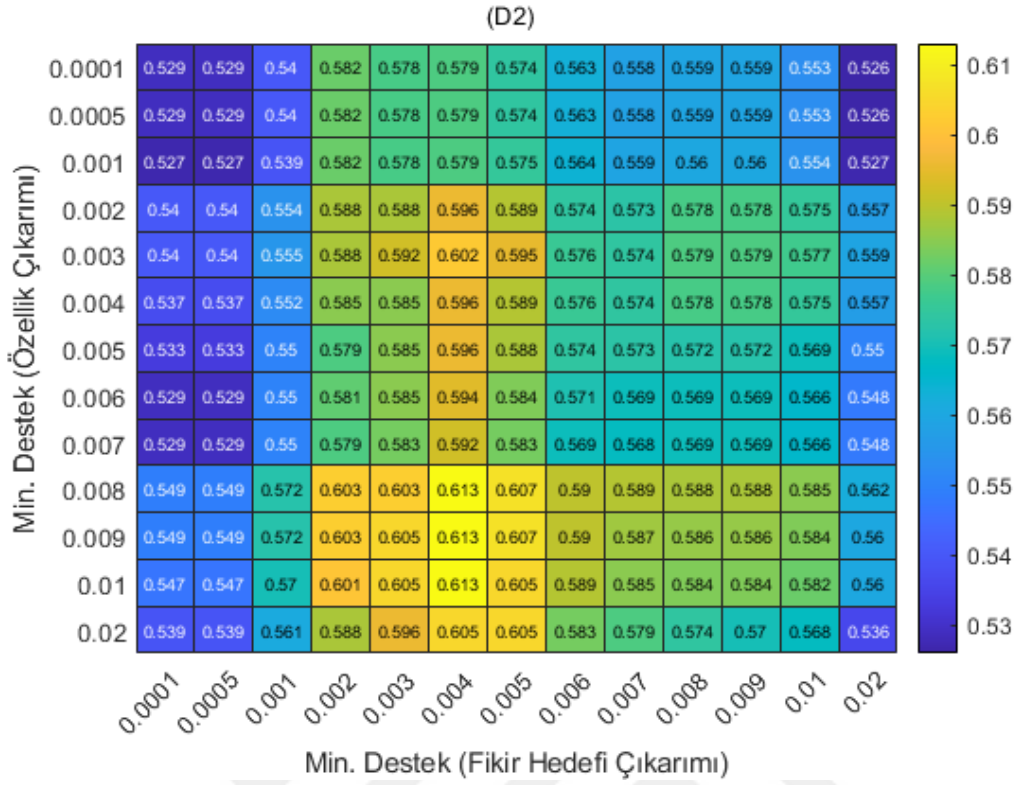
Tablo 6.2. Özellik ve fikir hedefi çıkarımı için en iyi sonuçları sağlayan parametreler

Özellik Çıkarımı							
Parametreler	D1	D2	D3	D4	D5	D6	D7
Min. Güven	0.9000	0.5000	0.4000	0.3000	0.6000	0.4000	0.4000
Min. Destek	0.0030	0.0080	0.0050	0.0090	0.0020	0.0200	0.0070
Tolerans Oranı	0.9000	0.9000	0.7000	0.9000	0.9000	0.5000	0.4000
Fikir Hedefi Çıkarımı							
Parametreler	D1	D2	D3	D4	D5	D6	D7
Min. Güven	0.6000	0.6000	0.4000	0.6000	0.4000	0.5000	0.5000
Min. Destek	0.0040	0.0040	0.0040	0.0090	0.0005	0.0010	0.0020
Tolerans Oranı	0.8000	0.9000	0.8000	0.5000	0.5000	0.5000	0.4000

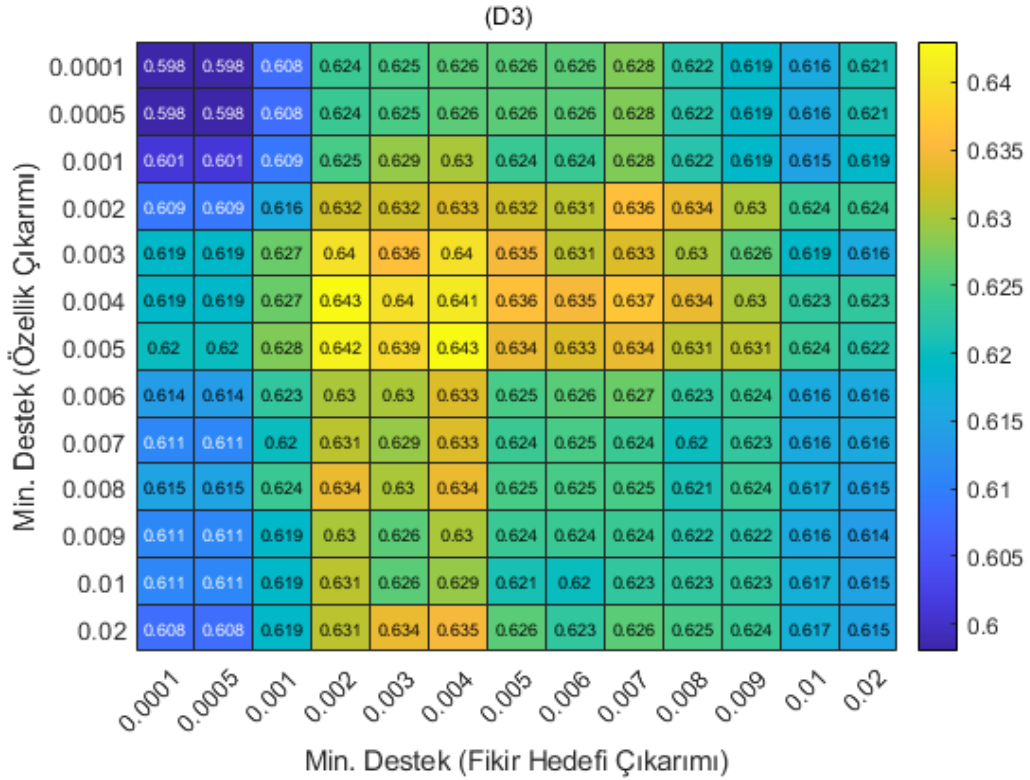
Tablo 6.2'deki parametreler dikkate alındığında, diğer parametreler sabitken farklı destek değerleri için deneyler tekrarlanmıştır. Kural transferi ve özellik transferi birlikte uygulandığında, fikir hedefi çıkarımı için elde edilen sonuçlar, her bir alan için Şekil 6.1-7'de sunulmuştur. Koyu mavi ile gösterilen kısımlar fikir hedefi çıkarımının başarısız olduğu sonuçları, sarı ile gösterilen kısımlar başarılı olan sonuçları göstermektedir. Her bir alan için en iyi sonuçların farklı destek değerleri etrafında kümелendiği görülmektedir. Örneğin; D1 alanı için sonuçları dikkate aldığımızda, özellik çıkarımı için destek değeri 0.003, fikir hedefi çıkarımı için destek değeri {0.003-0.007} aralığında seçilirse en iyi sonuçların elde edildiği görülmektedir. D6 alanı için elde edilen sonuçlara bakıldığında, en iyi sonuçların sol alt köşede yoğunlaştığı gözlemlenmektedir. Bu sonuçlara göre, özellik çıkarımı için destek değeri {0.002-0.002} aralığında, fikir hedefi çıkarımı için destek değeri {0.0001-0.001} aralığında olduğu görülmektedir. Sonuçlar ise {0.541-0.543} aralığında değişmektedir. Dolayısıyla, verilen aralıktaki destek değerlerinin değişiminin, sonucu kayda değer bir biçimde etkilemediği görülmüştür.



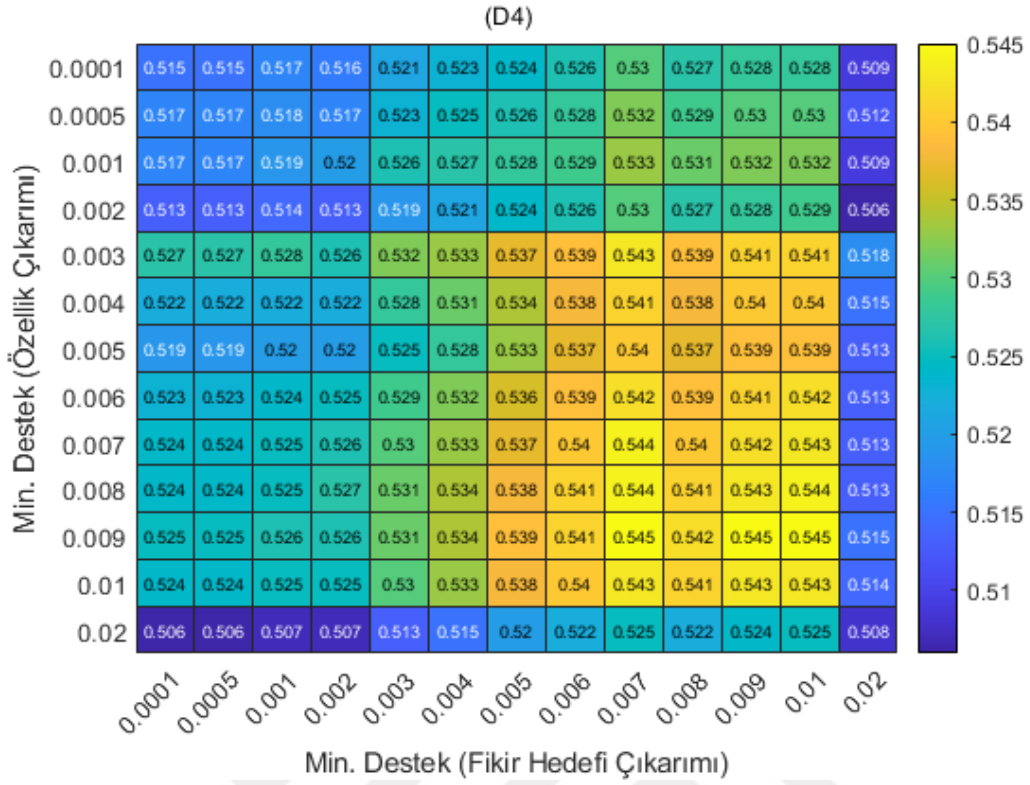
Şekil 6.1. D1 hedef alanı için en iyi sonuçları sağlayan parametrelerle özellik çıkarımı ve fikir hedefi çıkarımında minimum destek değerleri için F1-skorları



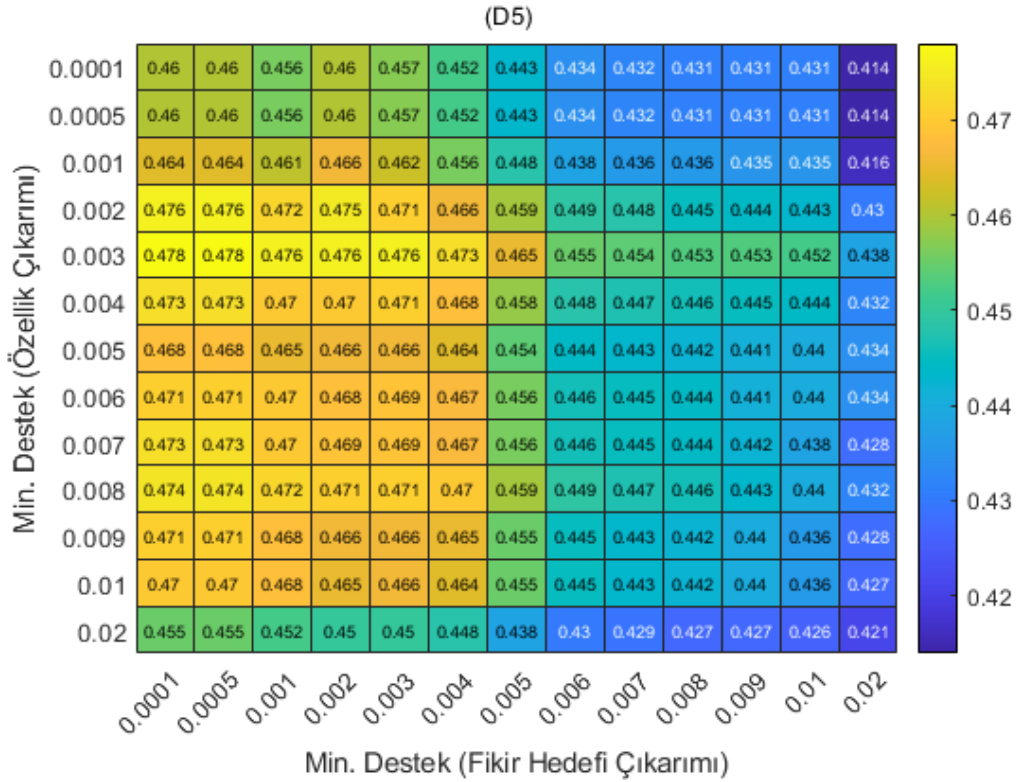
Şekil 6.2. D2 hedef alanı için en iyi sonuçları sağlayan parametrelerle özellik çıkarımı ve fikir hedefi çıkarımında minimum destek değerleri için F1-skorları



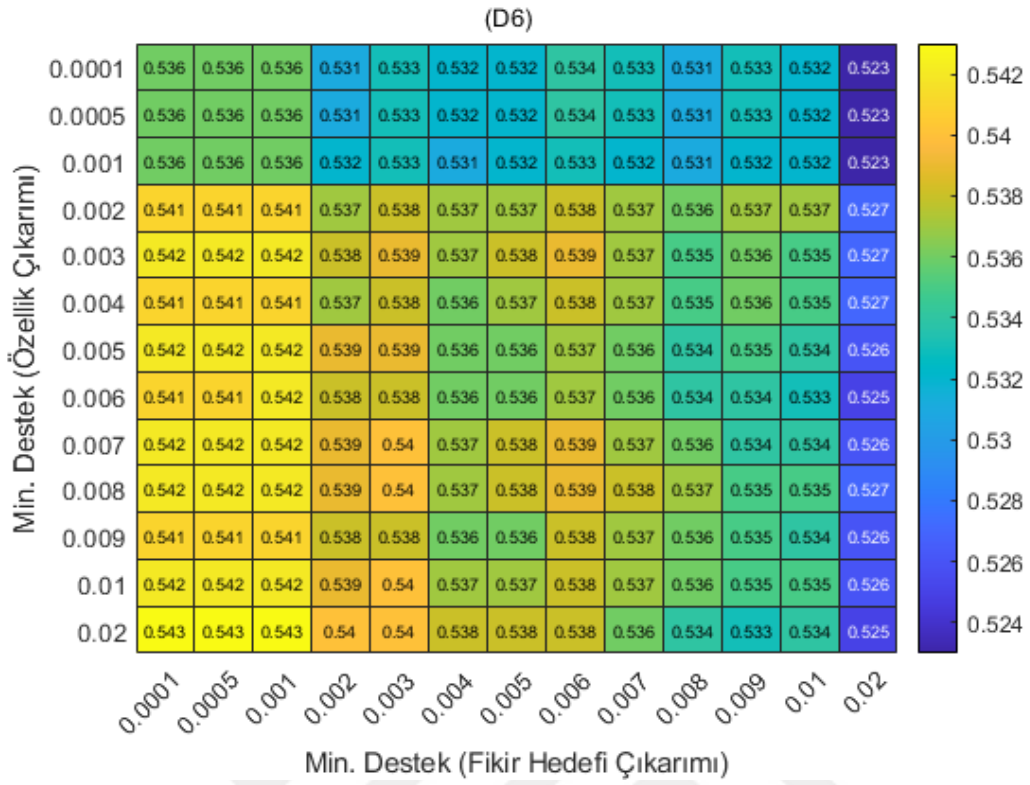
Şekil 6.3. D3 hedef alanı için en iyi sonuçları sağlayan parametrelerle özellik çıkarımı ve fikir hedefi çıkarımında minimum destek değerleri için F1-skorları



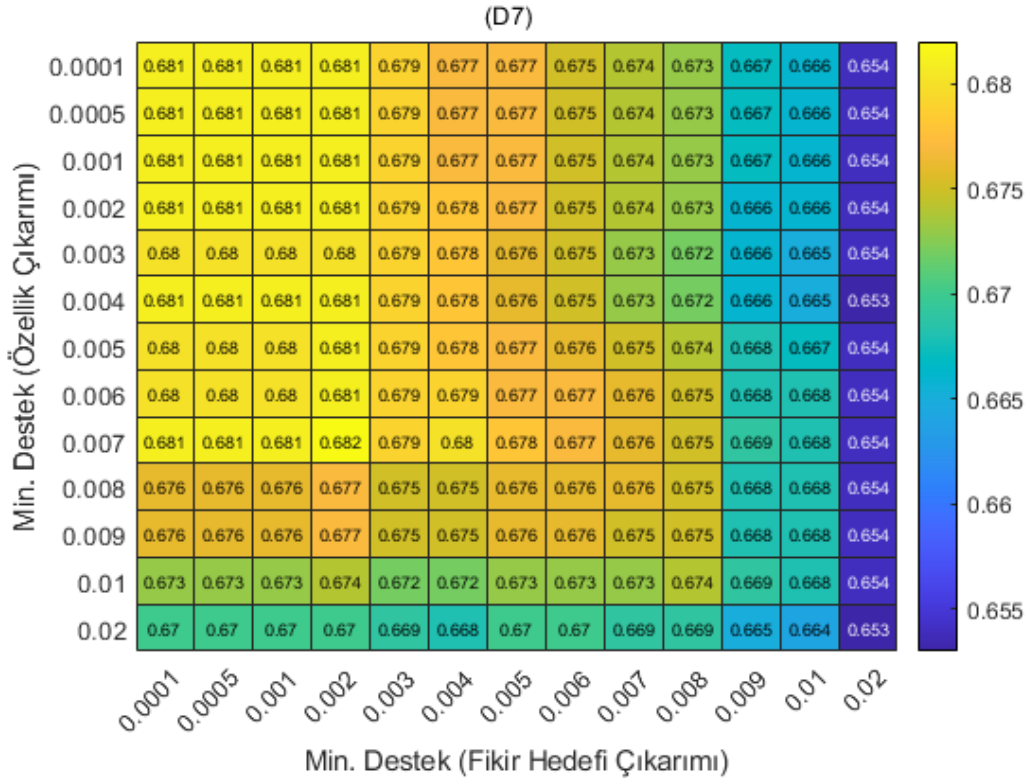
Şekil 6.4. D4 hedef alanı için en iyi sonuçları sağlayan parametrelerle özellik çıkarımı ve fikir hedefi çıkarımında minimum destek değerleri için F1-skorsları



Şekil 6.5. D5 hedef alanı için en iyi sonuçları sağlayan parametrelerle özellik çıkarımı ve fikir hedefi çıkarımında minimum destek değerleri için F1-skorsları



Şekil 6.6. D6 hedef alanı için en iyi sonuçları sağlayan parametrelerle özellik çıkarımı ve fikir hedefi çıkarımında minimum destek değerleri için F1-skorları



Şekil 6.7. D7 hedef alanı için en iyi sonuçları sağlayan parametrelerle özellik çıkarımı ve fikir hedefi çıkarımında minimum destek değerleri için F1-skorları

Fikir hedefi çıkarımı sonuçları Tablo 6.3'te diğer çalışmalarla karşılaştırılarak özetlenmiştir. En iyi sonuçlar kalın olarak gösterilmiştir. Etiket desenlerinin tamamı kullanıldığında elde edilen sonuçlara bakıldığında, D1, D2 ve D3 alanlarında diğer yaklaşımlara göre daha iyi performans sağlandığı görülmektedir. Diğer alanlar için düşük sonuçların elde edilmesinin sebebi, bu alanlarda “OTHER” etiket desenli veya “VB” ve “JJ” gruplu fikir hedefi içeren cümlelerin fazla olmasıdır. Çıkarma işlemi sadece Tablo 4.4'teki genel etiket desenleri için gerçekleştirilebildiğinden, “OTHER” etiket desenli fikir hedefleri hedef alandaki yorum içerisinden çıkarılamaz. Ayrıca, “NN” grubundaki etiket desenlerinden farklı olarak, “VB” ve “JJ” grupları veri kümelerinde nadiren gözlemlenir. Bu yüzden kaynak alanlardaki eğitim verisi yeterli olmadığından, nadiren gözlemlenen bu gruplar için az sayıda kural üretilebilmektedir. Bu kurallar hedef alan için başarılı bir fikir hedefi çıkarımında yeterli olamamaktadır. Etiket desenlerinin tamamı kullanılarak elde edilen sonuçlara ek olarak, tanımlanamayan (“UI”) fikir hedeflerinin kaldırılması ile elde edilen sonuçlar veya sadece “NN” grubundaki etiket desenleri ile elde edilen sonuçlar da sunulmuştur. Tanımlanamayan (“UI”) fikir hedeflerinin kaldırılması fikir hedefi çıkarımı başarısını D7 alanı hariç arttırmıştır. Tablo 6.1'de görülebildiği gibi D7 alanı hemen hemen hiç “UI” fikir hedefi içermemektedir, bu yüzden kaldırma işlemi sonucu pozitif anlamda etkilememektedir. “UI” fikir hedeflerinin kaldırılması, kaynak alanlardaki eğitim verisinde azalmaya yol açtığından, D7 alanı için fikir hedefi çıkarımı başarısı azalmaktadır. Sadece “NN” grubunun kullanımı fikir hedefi çıkarımının başarısını D2 alanı hariç tüm alanlarda arttırmaktadır. D6 ve D7 alanları için elde edilen performans artışının diğer yaklaşımlara göre daha iyi olduğu görülmektedir. Bu artışın nedeni D6 ve D7 alanlarında “OTHER” olarak gösterilen genel etiket desenleri dışındaki desenlerin diğer alanlara göre fazla olmasıdır. Sadece “NN” grubu kullanıldığında “OTHER” olan etiket desenleri dikkate alınmadığından başarı artmaktadır. D2 alanında “NN” grubu dışındaki desenlerden çok az bulunmaktadır. Dolayısıyla “NN” grubu dışındaki desenlerin kaldırılmasının, D2 alanına pozitif bir etki yapması beklenmez. Eğitim verilerinin azalması ise sonucun az miktarda düşmesine neden olduğu görülmektedir. “UI” fikir hedefleri kaldırıldıktan sonra sadece “NN” grup kullanılarak deneyler gerçekleştirilmiş, diğer yaklaşımlarla karşılaştırıldığında D5 alanı haricindeki diğer alanlar için en iyi sonuçlar elde edilmiştir.

Tablo 6.3. Her bir hedef alan için fikir hedefi çıkarımının karşılaştırmalı sonuçları (F1-skor)

Hedef Alan	CD-ALPHN	DT	kNN	MLP	NB	SVM	Önerilen Model			
							Bütün etiket desenleri	UI haricindeki bütün etiket desenleri	NN grubundaki etiket desenleri	UI haricindeki NN grubundaki etiket desenleri
D1	0.560	0.554	0.550	0.570	0.559	0.566	0.585	0.628	0.620	0.660
D2	0.579	0.579	0.572	0.581	0.572	0.574	0.612	0.669	0.599	0.664
D3	0.600	0.585	0.589	0.598	0.591	0.598	0.642	0.689	0.644	0.686
D4	0.589	0.583	0.581	0.587	0.585	0.595	0.545	0.589	0.575	0.628
D5	0.570	0.573	0.578	0.581	0.565	0.583	0.476	0.519	0.492	0.561
D6	0.610	0.480	0.547	0.581	0.570	0.504	0.542	0.545	0.611	0.620
D7	0.683	0.543	0.583	0.668	0.649	0.569	0.681	0.677	0.733	0.730

Önerilen model, diğer alanlar arası modellerle çeşitli ölçütlere göre karşılaştırılmış ve Tablo 6.4'te sunulmuştur. Bu ölçütler sırasıyla şunlardır: çıkarım modeli, veri kümesi özellikleri, fikir hedefi çıkarımı performansı. CRF (Jakob & Gurevych, 2010), heterojen ağ temelli modeller (Marcacini, Rossi, Matsuno, & Rezende, 2018), tekrarlayan (recursive) sinir ağı (Wang & Pan, 2018) (Wang & Pan, 2020) (Ding, Yu, & Jiang, 2017), BERT modeli ile öz dikkat mekanizması (self-attention mechanism) (Pereg, Korat, & Wasserblat, 2020) fikir hedefi çıkarımında kullanılan modellerdir. Bizim çalışmamızda ise sıralı desen temelli kural madenciliği modeli önerilmiştir. Önceki çalışmaların birçoğunda (Jakob & Gurevych, 2010) (Ding, Yu, & Jiang, 2017) (Wang & Pan, 2018) (Wang & Pan, 2020) (Pereg, Korat, & Wasserblat, 2020) göreceli olarak daha az alana sahip veri kümelerinin kullanıldığı gözlemlenmektedir. Bizim çalışmamızda ise toplam 7 alan için 6406 yorum kullanılmıştır. Çalışmaların birçoğunda (Jakob & Gurevych, 2010) (Ding, Yu, & Jiang, 2017) (Wang & Pan, 2018) (Wang & Pan, 2020) (Pereg, Korat, & Wasserblat, 2020), sadece bir kaynak alana karşılık bir hedef alan ele alınmıştır. Bizim çalışmamızda ise bir hedef alanın bütün kombinasyonlarına karşılık geri kalan bütün kaynak alanlar test edilmiştir. Sonuç olarak, diğer alanlar arası çalışmalarının performansları dikkate alındığında, önerilen model ile fikir hedefi çıkarımı için daha yüksek F-skor değerleri elde edildiği görülmektedir.

Tablo 6.4. Fikir hedefi çıkarımında önerilen modelle ilgili çalışmaların karşılaştırılması

Referans, Yıl	Çıkarım Modeli	Veri kümesi Boyutu	Veri kümesi Alanları	F-skor olarak çıkarım performansı Kaynak Alan → Hedef Alan	
(Jakob & Gurevych, 2010)	CRF	24555 6091 10969 5261 Toplam (46876)	Web-servisleri (W) Filmler (M) Arabalar (CR) Kameralar (CM)	M→W: 31.6 CR→W: 34.0 CW→W: 34.5 W→M: 42.2 CR→M: 47.9 CM→M: 49.9	W→CR: 32.4 M→CR: 38.4 CM→CR: 46.5 W→CM: 39.9 CR→CM: 47.5 M→CM: 39.1
(Ding, Yu, & Jiang, 2017)	RNN	5841 3845 3836 8545 Toplam (22067)	Restoran (R) Bilgisayar (L) Cihaz (D) Servis (S)	L→R: 46.7 D→R: 50.4 S→R: 52.0 R→L: 31.7 D→L: 36.2 S→L: 30.0	R→D: 32.0 L→D: 31.6 S→D: 33.4 R→S: 19.8 L→S: 23.4 D→S: 23.5
(Marcacini, Rossi, Matsuno, & Rezende, 2018)	Heterojen ağ temelli	45 34 41 95 99 3045 3045 Toplam (6404)	Dijital kamera 1 (D1) Dijital kamera 2 (D2) DVD player (D3) MP3 player (D4) Telefon (D5) Bilgisayar (D6) Restoran (D7)	D2,D3,D4,D5,D6,D7→D1:56.0 D1,D3,D4,D5,D6,D7→D2:57.9 D1,D2,D4,D5,D6,D7→D3:60.0 D1,D2,D3,D5,D6,D7→D4:58.9 D1,D2,D3,D4,D6,D7→D5:57.0 D1,D2,D3,D4,D5,D7→D6:61.0 D1,D2,D3,D4,D5,D6→D7:68.3	
(Wang & Pan, 2018)	RNN	5841 3845 3836 Toplam (13522)	Restoran (R) Bilgisayar (L) Cihaz (D)	R→L: 40.4 R→D: 35.1 L→R: 52.9	L→D: 40.4 D→R: 48.4 D→L: 51.1
(Wang & Pan, 2020)	RNN	5841 3845 3836 Toplam (13522)	Restoran (R) Bilgisayar (L) Cihaz (D)	R→L: 40.1 R→D: 37.3 L→R: 53.8	L→D: 41.2 D→R: 51.2 D→L: 51.7
(Pereg, Korat, & Wasserblat, 2020)	BERT modeli ile öz dikkat mekanizması	5841 3845 3836 Toplam (13522)	Restoran (R) Bilgisayar (L) Cihaz (D)	R→L: 47.6 R→D: 40.5 L→R: 54.7	L→D: 42.2 D→R: 54.5 D→L: 47.7
Önerilen Model	Sıralı desen temelli	45 34 41 95 99 3048 3044 Toplam (6406)	Dijital kamera 1 (D1) Dijital kamera 2 (D2) DVD player (D3) MP3 player (D4) Telefon (D5) Bilgisayar (D6) Restoran (D7)	D2,D3,D4,D5,D6,D7→D1:66.0 D1,D3,D4,D5,D6,D7→D2:66.4 D1,D2,D4,D5,D6,D7→D3:68.6 D1,D2,D3,D5,D6,D7→D4:62.8 D1,D2,D3,D4,D6,D7→D5:56.1 D1,D2,D3,D4,D5,D7→D6:62.0 D1,D2,D3,D4,D5,D6→D7:73.0	

7. SONUÇ

Bu çalışmada, duygu analizinde alanlar arası fikir hedefi çıkarımı için sıralı desen-temelli kural madenciliği modeli önerilmiştir. Daha önce yapılan çalışmaların birçoğunda özellik ve fikir hedefi kavramlarının birbirleri yerine kullanıldığı görülmektedir. Bu çalışmada, bu kavramların farkları ortaya konulmuş ve önerilen model bu temel üzerine kurulmuştur. İlk adım olan özellik çıkarım aşamasında özellikler çıkarılırken ikinci aşamada ise fikir hedeflerinin çıkarılması sağlanmıştır. Literatürdeki diğer çalışmalar, yorumlarda bulunan fikir hedefi olmayan özellikleri dikkate almamaktadır. Bu çalışmada, bahsedilen özelliklerin de etiketlenerek eğitime katılması sağlanmıştır. Özellik çıkarım aşamasında, kaynak alanlara ait özelliklerin hedef alana transfer edilerek fikir hedefi çıkarım başarısının arttığı gözlemlenmiştir. Dahası POS etiketleme ve etiketleme aşamalarında yeni yaklaşımlar uygulanmıştır. Bazı POS etiketlerinin etiketleme işlemine dahil edildiğinde bilgi kaybından dolayı fikir hedefi çıkarımının başarısını düşürdüğü görülmüş ve bu POS etiketleri işleme katılmamıştır. Fikir kelimelerinin, özelliklerin ve fikir hedeflerinin etiketlenmesi aşamasında, diğer çalışmalardan farklı olarak etiket grupları da kullanılmıştır. Bu sayede, kural üretiminde örneğin, isim ve isim ifadeleri için ayrı, fiiller için ayrı kuralların üretilmesi sağlanmıştır. Sonuç olarak etiketlerin sözcük türlerine göre özelleştirilmesi sağlanarak fikir hedefi çıkarımı başarısı artırılmıştır.

Bu çalışmada, ayrıca, fikir hedefi çıkarımı için alanlar arası bir model tasarlanması amaçlanmıştır. Hedef alanın bilinmeyen bir alan olduğu varsayılmış ve sadece kaynak alanlar kullanılarak fikir hedefi çıkarımı gerçekleştirilmiştir. Deneysel sonuçlar, önerilen modelin yorumların alanına bakılmaksızın fikir hedefi çıkarımı için ümit verici bir performans sunduğunu ortaya koymuştur. Kaynak alanlardan özellik transferinin hem özellik çıkarımında hem de fikir hedefi çıkarımında başarıyı arttırdığı görülmüştür. Farklı senaryolardaki bütün sonuçlar dikkate alındığında, önerilen modelin daha önceki çalışmalarla ve iyi bilinen sınıflandırıcılarla çoğu durumda daha iyi olduğu görülmektedir. Ayrıca, önerilen model, alanlar arası diğer modellerle karşılaştırıldığında, başarısının bu modellerle rekabet edecek düzeyde olduğu görülmüştür. Bu çalışmada, ürün yorumlarının farklı alanlarda olsa bile ortak desenlere sahip olduğu gözlemlenmektedir. Bu sonuç, sıralı desen temelli kural madenciliği yaklaşımlarının, alan bağımsız fikir hedefi çıkarım sistemleri için kullanılabileceğini göstermektedir.

Önerilen modelde, özellik çıkarımı için kullanılan kurallar ve fikir hedefi çıkarımı için kullanılan kurallar kaynak alana ait etiketli veriler kullanılarak üretilmektedir. Özellik çıkarımı için kullanılan kuralların üretiminde etiketlenen özellikler, etiketli verilerden elde edilen fikir hedeflerinin toplamından oluşmaktadır. Ancak, özelliklerin etiketsiz veriler kullanılarak da çıkarılması mümkündür. Böylece, büyük miktarda etiketsiz verinin özellik çıkarma aşamasında kullanılması, sonraki çalışmalar arasında yer alabilir. Önerilen model, özellikle fikir hedefi çıkarımına odaklanırken, duygu sınıflandırma adımı ile birlikte özellik tabanlı duygu özetleme için kurulması mümkün olabilecektir. Önerilen modelde ön işleme adımları İngilizce diline özel yapılmaktadır. Ancak, makine çeviri sistemleri kullanılarak diğer dillerde yazılmış yorumların İngilizce'ye çevrilmesiyle fikir hedefleri çıkarılabilecektir. Önerilen modelin, cümleyi analiz ederken bağılılık ayrıştırıcısı kullanmamasından dolayı çeviri hatalarından daha az etkileneceği düşünülmektedir.

KAYNAKÇA

- Qiu, G., Liu, B., Bu, J., & Chen, C. (2011). Opinion word expansion and target extraction through double propagation. *Computational Linguistics*, 37(1), 9-27.
- Alharbi, A., Taileb, M., & Kalkatawi, M. (2021). Deep learning in Arabic sentiment analysis: an overview. *Journal of Information Science*, 47(1), 129-140.
- Al-Smadi, M., Al-Ayyoub, M., Jararweh, Y., & Qawasmeh, O. (2019). Enhancing aspect-based sentiment analysis of Arabic hotels' reviews using morphological, syntactic and semantic features. *Information Processing & Management*, 56(2), 308-319.
- Al-Smadi, M., Qawasmeh, O., Al-Ayyoub, M., Jararweh, Y., & Gupta, B. (2018). Deep recurrent neural network vs. support vector machine for aspect-based sentiment analysis of Arabic hotels' reviews. *Journal of Computational Science*, 27, 386-393.
- Asghar, M. Z., Khan, A., Zahra, S. R., Ahmad, S., & Kundi, F. M. (2019). Aspect-based opinion mining framework using heuristic patterns. *Cluster Computing*, 22(3), 7181-7199.
- Blei, D., Ng, A., & Jordan, M. (2003). Latent dirichlet allocation. *The Journal of Machine Learning Research*, 3, 993-1022.
- Boiy, E., & Moens, M.-F. (2009). A machine learning approach to sentiment analysis in multilingual web texts. *Information Retrieval*, 12(5), 526-558.
- Brody, S., & Elhadad, N. (2010). An unsupervised aspect-sentiment model for online reviews. *Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics*, (s. 804-812).
- Chen, Z., Mukherjee, A., & Liu, B. (2014). Aspect extraction with automated prior knowledge learning. *Proc. of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, (s. 347-358).
- Ding, X., Liu, B., & Yu, P. (2008). A holistic lexicon-based approach to opinion mining. *Proceedings of the 2008 International Conference on Web Search and Data Mining*, (s. 231-240). Palo Alto, California, USA.

- Ding, Y., Yu, J., & Jiang, J. (2017). Recurrent neural networks with auxiliary labels for cross-domain opinion target extraction. *Proc. of the AAAI Conference on Artificial Intelligence*, (s. 3436-3442).
- Do, H., Prasad, P., Maag, A., & Alsadoon, A. (2019). Deep learning for aspect-based sentiment analysis: a comparative review. *Expert Systems with Applications*, 118, 272-299.
- Dragoni, M., Federici, M., & Rexha, A. (2019). An unsupervised aspect extraction strategy for monitoring real-time reviews stream. *Information Processing & Management*, 56(3), 1103-1118.
- Ekinci, E., & İlhan Omurca, S. (2020). Concept-LDA: incorporating Babelify into LDA for aspect extraction. *Journal of Information Science*, 46(3), 406-418.
- García-Pablos, A., Cuadros, M., & Rigau, G. (2018). W2VLDA: almost unsupervised system for aspect based sentiment analysis. *Expert Systems with Applications*, 91, 127-137.
- Hai, Z., Chang, K., & Kim, J.-j. (2011). Implicit feature identification via co-occurrence association rule mining. *International Conference on Intelligent Text Processing and Computational Linguistics*, (s. 393-404). Tokyo, Japan.
- Hai, Z., Chang, K., Kim, J.-J., & Yang, C. (2013). Identifying features in opinion mining via intrinsic and extrinsic domain relevance. *IEEE Transactions on Knowledge and Data Engineering*, 26(3), 623-634.
- Hofmann, T. (1999). Probabilistic latent semantic indexing. *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, (s. 50-57). Berkeley, California, USA.
- Htay, S., & Lynn, K. T. (2013). Extracting product features and opinion words using pattern knowledge in customer reviews. *The Scientific World Journal*, 2013, 1-5.
- Hu, M., & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, (s. 168-177). Seattle, Washington, USA.

- Hu, M., & Liu, B. (2004). Mining opinion features in customer reviews. *Proc. of the AAAI*, (s. 755-760). San Jose, California.
- Hu, M., & Liu, B. (2006). Opinion feature extraction using class sequential rules. *AAAI Spring Symposium: Computational Approaches to Analyzing Weblogs*, (s. 61-66). Stanford, USA.
- Jakob, N., & Gurevych, I. (2010). Extracting opinion targets in a single-and cross-domain setting with conditional random fields. *Proceedings of the 2010 Conference on Empirical Methods in Natural Language Processing*, (s. 1035-1045).
- Jiang, L., Yu, M., Zhou, M., Liu, X., & Zhao, T. (2011). Target-dependent twitter sentiment classification. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, (s. 151-160). Portland, Oregon, USA.
- Jin, W., Ho, H., & Srihari, R. (2009). A novel lexicalized HMM-based learning framework for web opinion mining. *Proceedings of the 26th Annual International Conference on Machine Learning*, (s. 465-472). Montreal, Canada.
- Jindal, N., & Liu, B. (2006). Mining comparative sentences and relations. *Proceedings of National Conference on Artificial Intelligence*, (s. 9).
- Kang, Y., & Zhou, L. (2017). RubE: Rule-based methods for extracting product features from online consumer reviews. *Information & Management*, 54(2), 166-176.
- Lafferty, J., McCallum, A., & Pereira, F. (2001). Conditional random fields: probabilistic models for segmenting and labeling sequence data. *Proceedings of the Eighteenth International Conference on Machine Learning*, (s. 282-289). San Francisco, USA.
- Li, F., Han, C., Huang, M., Zhu, X., Xia, Y.-J., Zhang, S., & Yu, H. (2010). Structure-aware review mining and summarization. *Proceedings of the Coling 2010: Posters*, (s. 653-661). Beijing, China.
- Li, S., Zhou, L., & Li, Y. (2015). Improving aspect extraction by augmenting a frequency-based method with web-based similarity measures. *Information Processing & Management*, 51(1), 58-67.

- Li, X., & Lei, L. (2021). A bibliometric analysis of topic modelling studies (2000–2017). *Journal of Information Science*, 47(2), 161-175.
- Li, Z., Li, X., Wei, Y., Bing, L., Zhang, Y., & Yang, Q. (2019). Transferable end-to-end aspect-based sentiment analysis with selective adversarial learning. *arXiv:1910.14192*.
- Liu, B. (2007). *Web data mining: exploring hyperlinks, contents, and usage data*. Springer Science & Business Media.
- Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.
- Liu, B. (2017). Many facets of sentiment analysis. *A Practical Guide to Sentiment Analysis*, 11-39.
- Liu, B., Hu, M., & Cheng, J. (2005). Opinion observer: analyzing and comparing opinions on the web. *Proceedings of the 14th International Conference on World Wide Web*, (s. 342-351). Chiba, Japan.
- Liu, K., Xu, L., & Zhao, J. (2012). Opinion target extraction using word-based translation model. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, (s. 1346-1356). Jeju Island, Korea.
- Liu, R., Shi, Y., Ji, C., & Jia, M. (2019). A survey of sentiment analysis based on transfer learning. *IEEE Access*, 85401-85412.
- Luo, Z., Huang, S., & Zhu., K. (2019). Knowledge empowered prominent aspect extraction from product reviews. *Information Processing & Management*, 56(3), 408-423.
- Maharani, W., Widyanoro, D. H., & Khodra, M. L. (2015). Aspect extraction in customer reviews using syntactic pattern. *Procedia Computer Science*, 59, 244-253.
- Marcacini, R., Rossi, R., Matsuno, I., & Rezende, S. (2018). Cross-domain aspect extraction for sentiment analysis: a transductive learning approach. *Decision Support Systems*, 114, 70-80.

- Pereg, O., Korat, D., & Wasserblat, M. (2020). Syntactically aware cross-domain aspect and opinion terms extraction. *Proc. of the 28th International Conference on Computational Linguistics*, (s. 1772-1777).
- Popescu, A.-M., & Etzioni, O. (2007). Extracting product features and opinions from reviews. *Natural Language Processing and Text Mining* (s. 9-28). içinde London: Springer.
- Poria, S., Cambria, E., Ku, L.-W., Gui, C., & Gelbukh, A. (2014). A rule-based approach to aspect extraction from product reviews. *Proceedings of the Second Workshop on Natural Language Processing for Social Media*, (s. 28-37). Dublin, Ireland.
- Quan, C., & Ren, F. (2014). Unsupervised product feature extraction for feature-oriented opinion determination. *Information Sciences*, 272, 16-28.
- Rabiner, L. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257–286.
- Rana, T., & Cheah, Y.-N. (2016). Aspect extraction in sentiment analysis: comparative analysis and survey. *Artificial Intelligence Review*, 46(4), 459-483.
- Rana, T., & Cheah, Y.-N. (2018). Sequential patterns-based rules for aspect-based sentiment analysis. *Advanced Science Letters*, 24(2), 1370-1374.
- Rana, T., & Cheah, Y.-N. (2019). Sequential patterns rule-based approach for opinion target extraction from customer reviews. *Journal of Information Science*, 45(5), 643-655.
- Ray, P., & Chakrabarti, A. (2020). A mixed approach of deep learning method and rule-based method to improve aspect level sentiment analysis. *Applied Computing and Informatics*, (in-press).
- Samha, A., Li, Y., & Zhang, J. (2014). Aspect-based opinion extraction from customer reviews. *arXiv preprint arXiv:1404.1982*.
- Shu, L., Liu, B., Xu, H., & Kim, A. (2016). Supervised opinion aspect extraction by exploiting past extraction results. *arXiv preprint arXiv:1612.07940*.
- Shu, L., Xu, H., & Liu, B. (2017). Lifelong learning crf for supervised aspect extraction. *arXiv preprint arXiv:1705.00251*.

- Tubishat, M., Norisma, I., & Abushariah, M. (2018). Implicit aspect extraction in sentiment analysis: review, taxonomy, opportunities, and open challenges. *Information Processing & Management*, 54(4), 545-563.
- Turney, P. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. *arXiv preprint cs/0212032*.
- Wang, W., & Pan, S. (2018). Recursive neural structural correspondence network for cross-domain aspect and opinion co-extraction. *Proc. of the 56th Annual Meeting of the Association for Computational Linguistics*, (s. 2171-2181).
- Wang, W., & Pan, S. (2020). Syntactically meaningful and transferable recursive neural networks for aspect and opinion extraction. *Computational Linguistics*, 45(4), 705-736.
- Wang, W., & Pan, S. J. (2019). Transferable interactive memory network for domain adaptation in fine-grained opinion extraction. *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*, (s. 7192-7199). Hawaii, USA.
- Wang, Y., He, W., Jiang, M., Huang, Y., & Qiu, P. (2020). CHOpinionMiner: an unsupervised system for Chinese opinion target extraction. *Concurrency and Computation: Practice and Experience*, 32(7), e5582.
- Wei, W., & Gulla, J. A. (2010). Sentiment learning on product reviews via sentiment ontology tree. *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, (s. 404-413). Uppsala, Sweden.
- Yu, J., Zha, Z.-J., Wang, M., & Chua, T.-S. (2011). Aspect ranking: identifying important product aspects from online consumer reviews. *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies*, (s. 1496-1505). Portland, Oregon, USA.
- Zhang, L., & Liu, B. (2014). Aspect and entity extraction for opinion mining. *Data Mining and Knowledge Discovery for Big Data* (s. 1-40). içinde Berlin: Springer.
- Zhang, L., Liu, B., Lim, S. H., & O'Brien-Strain, E. (2010). Extracting and ranking product features in opinion documents. *Proceedings of the 23rd International*

Conference on Computational Linguistics (Coling 2010), (s. 1462-1470). Beijing, China.

Zhuang, L., Jing, F., & Zhu, X. (2006). Movie review mining and summarization. *Proceedings of 15th ACM International Conference on Information and Knowledge Management*, (s. 43-50). Arlington, Virginia, USA.

ÖZGEÇMİŞ

Adı Soyadı: Muhammet Yasin PAK

Yabancı Dil: İngilizce

İletişim Yolu: Eskişehir Yolu: Konya Caddesi: 5.09

E-posta: y.pak@eskisehir.edu.tr

Öğrenim Geçmişi:

Doktora (2018-2021): Eskişehir Teknik Üniversitesi, Fen Bil. Ens., Bilgisayar Mühendisliği A.D.

Doktora (2015-2018): Anadolu Üniversitesi, Fen Bil. Ens., Bilgisayar Mühendisliği A.D.

Yüksek Lisans (2013-2015): Anadolu Üniversitesi, Fen Bil. Ens., Bilgisayar Mühendisliği A.D.

Lisans (2006-2011): Ege Üniversitesi, Bilgisayar Mühendisliği Bölümü

Meslek Geçmişi:

Araştırma Görevlisi (2018-2021): Eskişehir Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü

Araştırma Görevlisi (2015-2018): Anadolu Üniversitesi, Bilgisayar Mühendisliği Bölümü

SCI ve SCI-Expanded kapsamındaki dergilerde yayınlanan makaleler:

Pak M. Y., Gunal S., “A Model for Cross-Domain Opinion Target Extraction in Sentiment Analysis”, Computer Systems Science and Engineering (accepted), 2021

Pak M. Y., Gunal S., "Sentiment classification based on domain prediction", Elektronika ir Elektrotechnika, 22(2), 96-99, Apr 2016.

Demirkiran, E. T., Pak, M. Y., Cekik, R., "Multi-criteria Collaborative Filtering Using Rough Sets Theory", Journal of Intelligent & Fuzzy Systems, 40(1), 907–917, 2021.

Diğer Dergi Makaleleri:

Pak M.Y., Gunal S., "The impact of text representation and preprocessing on author identification", Anadolu University Journal of Science and Technology A - Applied Sciences and Engineering, 18 (1), 218–224, 2017.