

PAPER DETAILS

TITLE: Review of Sentiment Analysis and Opinion Mining Algorithms

AUTHORS: Ümit Can,Bilal Alatas

PAGES: 75-111

ORIGINAL PDF URL: <https://dergipark.org.tr/tr/download/article-file/317689>

Duygu Analizi ve Fikir Madenciliği Algoritmalarının İncelenmesi

Umit Can^{1*}, Bilal Alatas²

¹Munzur Üniversitesi, Bilgisayar Mühendisliği Bölümü, Tunceli, Türkiye

²Fırat Üniversitesi, Yazılım Mühendisliği Bölümü, Elazığ, Türkiye

*Yazışmalardan sorumlu yazar: E-mail: ucan@munzur.edu.tr

Makale gönderme tarihi: 06.04.2017, Makale kabul tarihi: 13.06.2017

Özet

Duygu Analizi veya Fikir Madenciliği, metin madenciliğinin önemli bir alanı ve son yılların önemli araştırma konularından biridir. Günümüzde şirketlerin veya kişilerin ürettikleri ürünler çok hızlı bir şekilde tüketiciye ulaşmakta ve bu ürünlerle ilgili yapılan yorumlarda gelişen teknoloji ile beraber internet dünyasına yansımaktadır. Bu yorumların ne anlama geldiği üreticiler için çok önemlidir. Bunun dışında Duygu Analizi veya Fikir Madenciliği finanstan tutun da tıp alanına kadar birçok alanda kullanılabilir. Duygu Analizi; bir metni ele alarak bu metnin olumlu, olumsuz veya tarafsız bir içeriğe sahip olup olmadığını inceler. Genel olarak fikirlerin, duyguların ve metinlerin nesnelliğinin hesaplanma işlemi de denilebilir.

Bu çalışma da, Duygu Analizi hakkında araştırma yapılmış olup Duygu Analizi sınıflandırma teknikleri incelenerek tüm alt bileşenleri ile beraber anlatılmıştır. Duygu Analizi ile ilgili birçok güncel makale incelenmiş ve kısaca anlatılmıştır. En son olarak genel bir değerlendirme ve sonuç yazılarak çalışma bitirilmiştir.

Anahtar Kelimeler: Duygu analizi, duygu analizi algoritmaları, duygu analizi uygulamaları, fikir madenciliği

Review of Sentiment Analysis and Opinion Mining Algorithms

Abstract

Sentiment Analysis or Opinion Mining is an important field in text mining. Nowadays the products which are produced by companies or persons are reached to consumers mercerially and reviews about these products issued on web pages. As understood easily these reviews are very significant for producers. In addition to that, Sentiment Analysis can be used from financial field to medicine field. Sentiment Analysis investigates a text that has a positive, negative or a neutral meaning. In general, we can imagine Sentiment Analysis as the computational treatment of opinions, sentiments, and subjectivity of text.

In this study, a research about Sentiment Analysis has been performed and Sentiment Analysis classification techniques have been explained with its all parts. Many articles related with Sentiment Analysis have been studied and briefly explained. Then, one application about Sentiment Analysis has been shown for understanding more about Sentiment Analysis. Consequently, a general assessment of this issue has been done and the study has been finished with the result section.

Keywords: Opinion mining, sentiment analysis, sentiment analysis algorithms, sentiment analysis applications

GİRİŞ

Duygu Analizi (DA) veya Fikir Madenciliği (FM) insanların görüşlerinin, tutumlarının ve duygularının bir varlığı ifade ettiği bilişimsel çalışmalardır. Bu varlık, bireyleri, olayları ya da konuları temsil edebilir. Konular ise çoğunlukla yorumları kapsamaktadır. DA ve FM genellikle birbirlerinin yerine kullanılırlar. Fakat bazı araştırmacılar DA ve FM'nin birbirinden kısmen farklı olduğunu söylemeye başlamışlardır. FM insanların bir varlık hakkındaki düşüncelerini analiz edip onu açığa çıkarır. Öbür yandan DA ise bir metni analiz ederek oradaki duygusal ifadeleri ortaya koyar. Bu yüzden DA düşünceleri

bulur, bunların ifade ettiği duyguları açığa çıkartır ve onların polaritelerini Şekil 1'de olduğu gibi sınıflandırır (Medhat ve ark., 2014).

Ürün yorumları üzerinde gerçekleştirilen bir DA süreci Şekil 1'de gösterildiği gibidir. DA'da üç tane temel sınıflandırma seviyesi vardır: doküman seviyesi, cümle seviyesi ve görüş seviyesi. Doküman seviyesi bir fikir dokümanını, pozitif veya negatif bir yorum ya da duygu ifade edip etmemesine göre sınıflandırır. Tüm dokümanı temel bir bilgi birimi olarak görür. Cümle seviyesi ise her bir cümledeki duyguyu sınıflandırmayı amaçlar. İlk adımda cümlelerin subjektif veya objektif olup olmadığı belirlenir.

Eğer cümle subjektif ise, cümle seviyesi DA, cümlelerin negatif veya pozitif fikir ifade edip etmediğine karar verir (Medhat ve ark., 2014).



Şekil 1. Ürün yorumlarındaki DA süreci

Wilson ve ark (2005) duygusal ifadelerin doğada subjektif olma zorunluluklarının olmadığına işaret etmişlerdir. Ayrıca doküman ve cümle seviye sınıflandırmaları arasında büyük bir fark yoktur, çünkü cümleler sadece küçük dokümanlardır (Liu, 2012). Doküman ve cümle seviyelerindeki sınıflandırmalarda yorumların detaylandırılması zorunluluğu yoktur. Bu görüş seviyesinde detaylandırılır. Görüş seviyesi DA, duyguyu varlıkların belirli yönlerine göre sınıflandırmayı amaçlar. İlk adım, varlıkların ve onların özelliklerini belirlemektir. Yorumcular aynı varlığın farklı özellikleri için farklı yorumlar yapabilirler. "Telefonun ses kalitesi hiç iyi değil fakat şarj ömrü uzun" gibi bir cümle bu duruma örnek gösterilebilir. Bu çalışma ilk iki çeşit DA'yı ele almıştır.

DA'da kullanılan veri setleri bu alandaki önemli konulardan biridir. Verinin ana kaynağı genelde ürün yorumlarıdır. Bu yorumlar yaptıkları iş konusunda yatırım yapacak iş sahipleri için önemlidir, çünkü bu yorumları analiz ederek kendi ürünleri hakkında bir fikir elde ederler. Yorumların elde edildikleri kaynaklar büyük oranda yorum siteleridir. DA sadece ürünlerin yorumlarında uygulanmaz; bunların dışında stok marketlere (Yu ve ark., 2013; Hagenau ve ark., 2013), yeni başlıklara (Xu ve ark., 2012) veya politik tartışmalara da (Maks ve Vossen, 2012) uygulanabilir. Politik tartışmalar konusunda örnek vermek gerekirse, bundan insanların adaylar ve siyasi partiler

hakkındaki görüşleri anlaşılabilir. Politik gönderilerden seçimlerin sonuçları tahmin edilebilir. Sosyal ağ ve küçük blog sayfaları çok iyi birer bilgi kaynağı olarak gösterilebilir, çünkü insanlar buralarda birçok konu hakkında yorum yaparlar. Bu yorumlar DA'da veri kaynakları olarak kullanılır.

Son zamanlarda DA algoritmaları konusunda birçok uygulama ve iyileştirme yapılmıştır. Bu çalışmada bu iyileştirmelere daha yakından bakılarak, DA teknikleri incelenmiş ve sınıflandırılmıştır. Son zamanlarda yapılan birçok yayın incelenmiştir. Şekil 1'deki Öznitelik Seçimi (ÖS) teknikleri tartışılarak ilgili makaleler gösterilmiştir. Duygu Sınıflandırma (DS) teknikleri ise Şekil 2'de gösterilmiştir.

Bu çalışma popüler birçok DA tekniğini ve uygulamalarını incelemiş ve tek bir makalede araştırmacılara sunmuştur. Bu açıdan araştırmacılar için son derece faydalı bir çalışmadır. Bu çalışmada diğer çalışmalarda var olmayan bir şekilde birçok DA tekniği kategorize edilmiştir. Bu makale DA ile ilgili yeni alanları da tartışmıştır. Bu alanlar Duygu Tespiti (DT), Kaynak İnşası (Kİ) ve Transfer Öğrenmesidir (TÖ). DT cümle içinde açık veya örtülü biçimde olan duyguları ortaya çıkarıp analiz etmeyi amaçlar. TÖ ve Çapraz Alan sınıflandırması ise bir alandan analiz edilerek elde edilen veriyle ve sonraki hedef alanla ilgili sonuçlarla alakalıdır. Kİ, fikirlerin polaritelerine göre ifade edildiği durumlarda sözlük ve korpus oluşturmayı amaçlar. Bu çalışmada bu konulara yakından bakılmıştır.

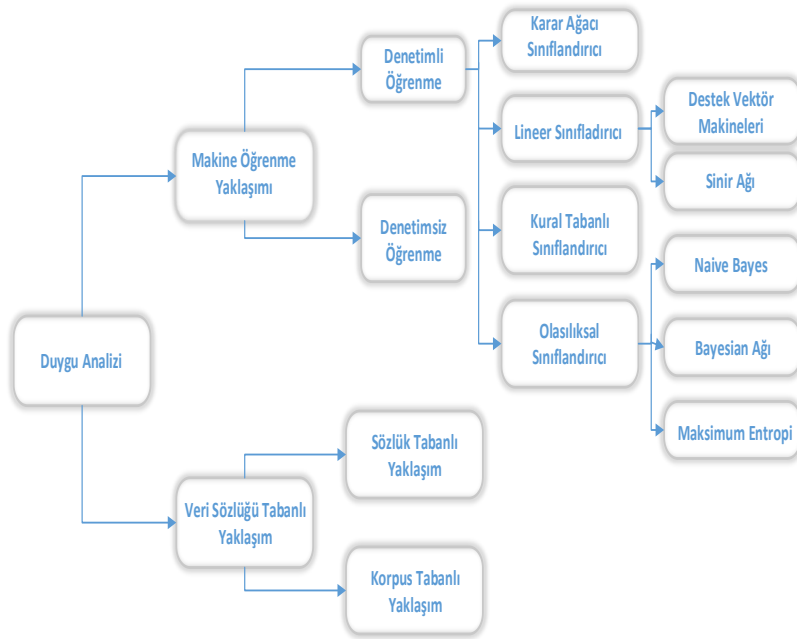
DA alanında her yıl onlarca yayın yapılmaktadır ve yıllar boyunca da bu oran artmaktadır. Bu artış DA ile ilgili araştırmaları ve trendleri sınıflandırmayı ve sistematize etmeyi gerektirir. Okuyucular (Wilson ve ark., 2005; Pang ve Lee, 2008; Liu, 2012; Montoyo ve ark., 2012; Piryani ve ark., 2017) referanslı makalelerde sofistike ve detaylı araştırmalar bulabilirler. Bu araştırmalar DA uygulamalarındaki problemleri tartışmış ancak DA tekniklerini tartışmamıştır.

İki uzun ve detaylı araştırma Pang ve Lee (2008) ve Liu (2012) tarafından gösterilmiştir. DA uygulamaları ve DA'da karşılaşılan zorluklar üzerine odaklanmışlardır. DA'daki her bir problemi çözmek için kullanılan tekniklerden bahsetmişlerdir. Cambria ve ark (2013), Feldman (2013) ve Montoyo ve ark (2012) DA'daki yeni

trendleri göstermek için küçük araştırmalar sunmuşlardır. Tsytsarau ve Palpanas (2012) DA'daki ana konuları tartışan bir araştırma sunmuşlardır. Her bir konu için konuların tanımlarını, problemleri ve gelişmelerini anlatmışlardır. Bu araştırma yazısında anlatılanlar bir önceki cümlede anlatılan çalışma Tsytsarau ve Palpanas (2012) ile benzerlik göstermektedir, fakat bu araştırma perspektif ve yaklaşım farklılıkları sunmaktadır.

Bu çalışmanın birçok nedenden dolayı önemli katkıları vardır. İlk olarak bu araştırma son zamanlarda çıkan birçok tekniği kategorize

etmiştir. Bu araştırmacılara uygulamaları konusunda yardımcı olacaktır. İkinci olarak, çeşitli DA teknikleri algoritmaları kısa detayları ve onların orijinal referansları dikkate alınarak kategorize edilmiştir. Bu, DA konusunda yeni olanlar için panoramik bir bakış açısı sağlayabilir. Üçüncü olarak, uygun kalite testi veri setleri tartışılmış ve belirli uygulamalarına göre kategorize edilmiştir. Sonuç olarak bu araştırma, DA ilgili olarak duygu tespiti, kaynak inşası ve öğrenme transferi gibi konuları içerecek şekilde tartışılarak geliştirilmiştir.



Şekil 2. Duygu sınıflandırma teknikleri

LİTERATÜR TARAMASI

Bu çalışmada özetlenmiş 99 makale Çizelge 1'de gösterilmiştir. Üçüncü kolonda makalelerin amaçları gösterilmiş ve DA, DT, ÖS, TÖ ve Kİ olmak üzere altı kategoriye bölünmüşlerdir. Kİ kategorisi; sözcük, korpora ve sözlükler olarak sınıflandırılabilir. Duygu sınıflandırma problemini çözen makale DS olarak kategorize edilmiştir. Duygu Analiz problemlerini çözen makaleler ise DA olarak kategorize edilmiştir. Özellik seçimine katkı sunan makaleler ise ÖS olarak kategorize edilmişlerdir. Daha sonra DT, Kİ ve TÖ olarak gösterilmiştir.

Dördüncü kolonda, makale Evet/Hayır (E veya H) soruları ile yönlendirilmiş ise alan odaklı, alana-özgü veriler DA'da kullanılıyor demektir.

Beşinci kolonda ise kullanılan algoritmalar gösterilmiş ve Şekil 2' deki gibi kategorileri belirtilmiştir. Bölüm 4'te gösterildiği gibi bazı makalelerde farklı algoritmalar kullanmıştır. Bu uygulamalar örneğin Steinberger ve ark (2012) tarafından gösterilmiştir. Bu durumda algoritmanın sadece ismi yazılmıştır. Altıncı kolonda ise makalenin DA tekniklerini genel metin analizinde (G) mi yoksa ikili sınıflandırma (Pozitif/Negatif) probleminde mi kullandığı gösterilmektedir. Yedinci kolon ise makalelerin kullandığı algoritmaların değerlendirilmesinde kullanılan verilerin kapsamını göstermektedir. Bu veriler; yorumlar, haberler, web sayfaları, mikro bloglar ve diğerlerinden oluşabilir. Sekizinci kolon ise benchmark veri setlerini veya kullanılmış iyi bilinen veri kaynaklarını

göstermiştir. Son kolon makalede İngilizce'den başka bir dil analiz edilmişse onu belirtmiştir.

Bu araştırmanın metodolojisi popüler ÖS ve DS algoritmalarının kısa özetleri ve ilgili alanlarda tartışılması şeklinde olup; bu

makalelerin bu algoritmalara katkıları ve DA'daki özel problemleri nasıl çözdükleri gösterilmiştir.

Bu araştırmanın ana amacı DA için tek bir kategori oluşturmaktır.

Çizelge 1. Makale özetleri

Referanslar	Yıl	Görev	Alan- Odaklı	Kullanılan Algoritmalar	Polarite	Veri Kapsamı	Veri Seti/kaynağı	Diğer Diller
Montoyo ve ark.	2010	DA	E	Kural- tabanlı	G	Web forumları	automotviefo rums.com	
Qiu ve ark.	2010	DT	H	Web-tabanlı, Semantik etiketleme ve kural tabanlı	Poz/Ne g	Web sayfaları	N/A	
Lu ve ark.	2010	DT	H	Sözcük tabanlı, Semantik	G	Kişisel hikayeler	experiencepr oject.com	
Neviarouska ya	2010	DS	H	Markov Blanket, DVM, NB, ME	Poz/Ne g	Film yorumları, Haber makaleleri	IMDB	
Yan-Yan ve ark.	2010	DS	E	Grafik tabanlı yaklaşım	Poz/Ne g	Film, Ürün yorumları	N/A	Çince
Bai (2011)	2011	DS	H	Grafik tabanlı yaklaşım, NB, DVM	Poz/Ne g	Kamera yorumları	Çince fikir analizi alanı	Çince
Hu ve Wenjie	2011	DA	H	Semantik, GSA tabanlı	G	Yazılım programı kullanıcılar ının fikirleri	CNETD	
Cao ve ark.	2011	DS	E	Zayıf ve yarı denetimli sınıflandırm a	Poz/Ne g	Film yorumları, Çoklu alan duygu veri seti	IMDB, amazon.com	
He ve Zhou	2011	Kİ-İNG	E	Rastgele yürüme algoritması	G	Elektronik aletler, Stok, Otel yorumları	Alana özgü Çince korpus	Çince
Tan ve Wu	2011	TÖ- İNG	E	Entropi tabanlı algoritma	G	Eğitim, Stok, Bilgisayar yorumları	Alana-özgü çince veri seti	Çince
Tan ve Wang	2011	TÖ- İNG	E	Sıralama agoritması	G	Kitap, Otel, Dizüstü bilgisayar, Yorumları	Alana-özgü çince veri seti	Çince
Wu ve Songbo	2011	DS	H	KRA	Poz/Ne g	Araba, Otel,	N/A	

						Bilgisayar yorumları		
Jiao ve Zhou	2011	TÖ-İNG	E	SAR	G	Film yorumları, QA	MPQA, RIMDB, CHES	
Lambov ve ark.	2011	DA	H	2-seviyeli KRA	G	Mobil müşteri Yorumları	amazon.com, epinions.com, blogs, SNS ve CRM eposta	
Xu ve ark.	2011	DA	H	Çoklu-sınıf DVM	G	Dijital kameralar, MP3 yorumları	N/A	
Chien ve You-De	2011	DA	E	DVM-ki kare	G	Alicıların web sayfalarında ki gönderileri	ebay.com, wikipedia.com, epinions.com	
Fan ve Chang	2011	DA	H	Semantik	G	Çince eğitim verisi	NTCIR7 MOAT	Çince
Zhou ve ark.	2011	DS	H	Sözlük tabanlı, semantik	Poz/Ne g	Film yorumları	MDB	
Heerschop ve ark.	2011	DS	H	İstatiksel, semantik	Poz/Ne g	Ürün yorumları	amazon.com	
Zirn ve ark.	2011	DA	H	İstatiksel	G	Kitap yorumları	amazon.com	
Gupt ve ark.	2012	ÖS	H	İstatiksel (HMM - GDT), ME	Poz/Ne g	Film yorumları	N/A	
Duric ve Song	2012	Kİ-İNG	E	Semantik	G	Restorant yorumları	N/A	İspanyolca
Robaldo ve Luigi	2012	DA	E	İçerik tabanlı metot, DDİ	G	Restorant yorumları	N/A	
Kang ve ark (2012)	2012	DA	H	Sözlük tabanlı, DDİ	G	Haberler	N/A	
Moreo ve ark.	2012	DA	H	NKB, semantik	G	Ürün yorumları	N/A	Çince
Zhang ve ark.	2012	DA	H	DDİ	G	Ürün yorumları	amazon.com	
Min ve Park	2012	DS	H	Yarı denetimli, BN	G	Yapay veri setleri	N/A	
Ortigosa ve ark.	2012	Kİ-İNG	E	DDİ	G	Bloglar	ISEAR	İspanyolca, italyanca
Boldrini ve ark.	2012	DT	E	Korpus tabanlı	G	Blog verileri	Canlı gazete blogları, Perimalleri, Açıklamalı bloglar	
Keshtkar ve	2012	Kİ-İNG	H	Üçgenleştir	G	Haber	Duygu	Diğer

Inkpen				me		sayfaları	sözlükleri	latin, Arapça
Steinberger ve ark.	2012	DS	E	NB, DVM, kural tabanlı	G	2 taraflı tartışmalar	convinceme.net	
Walker ve ark.	2012	DT	H	Sözlük tabanlı, DVM	G	Duygu korpusu	ISEAR, Emotinet	
Xu ve ark.	2012	DA	H	Semantik	Poz/Ne g	Sözcükler(Veri sözlüğü)	Felemenkçe wordnet	Felemenkçe
Balahur ve ark.	2012	DA	H	DVM, K-nearest neighbor, NB, BN, DT, bir kural öğrenici	Poz/Ne g	Medya	Media-analysis şirketi	
Lane ve ark.	2012	DS	H	DVM, 1-NN	Poz/Ne g	İlişkiler, biyografi	BWSA	Felemenkçe
Camp ve Bosch	2012	ÖS	H	Semantik, NB, DVM, DT	G	Haberler, Yergili yazılar, Müşteri yorumları	amazon.com	
Reyes ve Rosso	2012	DT	H	Sözlük tabanlı	G	Emailler, Kitaplar, Romanlar, Peri masalları	Enron eposta korpus	
Mohammad	2012	DS	H	Denetimsiz, GDT	G	Sosyal yorumlar	2000-SINA blog veri seti, 300-SINA Hownet veri sözlüğü	Çince
Kaufmann	2012	DS	H	DVM, YSA	Poz/Ne g	Film, GPS, Kamera, Kitap yorumları	amazon.com	
Liu	2012	ÖS	E	NKB tabanlı	G	Stok Haberleri	N/A	
Moraes ve ark.	2013	DS	H	DVM, NB, C4.5	Poz/Ne g	Filim yorumları	MC, MCE korpus	İspanyolca
Martin-Valdivia ve ark.	2013	DA	E	BKA	G	Akıllı telefonlar, Tweetler	Twitter	
Kontopoulou ve ark.	2013	DS	H	NB, DVM	Poz/Ne g	Filim yorumları, Tweetler	Twitter	
Rui ve ark.	2013	DS	H	DVM	G	Tweetler	Twitter	
Yu ve ark.	2013	ÖS	E	KI-KARE, BNS, DVM	G	Stok market	DGAP, EuroAhoc	
Li ve Li	2013	DT	E	DDI	G	Hikâyeler	Aozora Bunko	Japonca
Ptaszynski ve ark.	2013	DA	H	Semantik	Poz/Ne g	Fast food yorumları	N/A	Tayvanca
Pai ve ark.	2013	DS	E	Taksonomi	Poz/Ne	Telefon,	epinions.com	

				tabanlı, korpus tabanlı	g	Araba, otel yorumları		
Fermin ve ark.	2013	DS	H	Semantik	Poz/Ne g	Blog Gönderiler i	TREC 2006, TREC 2007, and TREC 2008	
Li ve Tsai	2013	Kİ	H	ME	G	Korpus	N/A	Felem enkçe, Çince
Hagenau ve ark.	2013	DA	H	S-HAL, SO- NKB	G	Haber sayfaları	Sogou CS korpus	Çince
Di Caro ve Grella	2013	DS	H	NB, DVM	Poz/Ne g	Restorant yorumları	N/A	
Alvaro	2013	DA	H	Sözlüksel tabanlı yaklaşım, MÖ	G	Facebook mesajları	www.facebo ok.com	
Ayoub ve ark.	2013	DA	H	İteratif önyüklemeli algoritma	G	Çevrimiçi ürün yorumları	N/A	
Efstratios ark.	2013	DA	E	Ontoloji tabanlı teknik	G	Tweeter gönderileri	Twitter	
Haeng-Jin	2013			Kişisel değer tutum (KDT)	G	Youtube'de ki bilgi teknoloji alanındaki yorumlar	Youtube	
Emma ve ark.	2013	ÖS	E	DVM, önişlemeli metotlar, ki- kare		Çevrimiçi film yorumları		
Nizam ve Akin	2013	DA	H	NB, Random Forest, SMO, Karar Ağacı, 1-Nearest Neighbors (IB1)	G	Veri setleri	Weka	
Zhao ve ark.	2013	DA	H	SAMC	G	Mikro- blog içeriği	Micro-blog	Çince
Shi ve ark.	2013	DA	E	Bulanık duygu ontolojisi	G	Mikro- blog gönderileri	SinaWeibo	Çince
Jebaseeli ve Kirubakar	2013	DS	H	DDİ, Duyarlılık değeri, Anma ve F- ölçümü				
Isa	2013	DS	H	Ters Porter algoritması, Geri-İleri Algoritması Bağışıklık Ağı Algoritması	G	Malaysian newspaper news	Berita Harian	Malay ca

Smailovic	2014	DA		DVM	G		Twitter
Balahur ve Turchi	2014	DT, DS		DDİ	G		
Hogenboom ve ark.	2014	DA		Makine Transferi (MT) temeli	Poz/Ne g	Web siteleri, forumlar ve bloglar	SentiWordNet, DutchWordNet, İngilizce, Hollandaca
Lau ve ark.	2014	DA	E	Sosyal analitik metodolojisi, yarı denetimli bulanık ontoloji madencilik algoritması	G	Sosyal medya yorumları	N/A
Kranjc ve ark.	2014	DS		DVM, ClowdFlows		Micro-blog verileri	
Habernal ve ark.	2014	DA		DDİ, ME, DVM, NB	Poz/Ne g G	Facebook marka sayfalarındaki Çek gönderileri, film yorumları	Facebook, Çek film veri seti, Çekçe
Bravo-Marquez ve ark.	2014	DS			Poz/Ne g	Twitter gönderileri	Twitter
Agarwal ve Mittala.	2014	ÖS		IG özellik seçim metotları, MKTNB, DVM	G	Film yorumları, kitap, elektronik, mutfak eşyaları gibi ürün yorumları	N/A
Meral ve Diri.	2014	DS		NB, RF (rastgele orman), DVM		Tweetter verileri	Twitter
Yu ve Wang	2015	DA		Veri-sözlüğü yaklaşımı	Poz/Ne g	Tweetter gönderileri	Twitter
Jurado ve Rodriguez	2015	DT		Veri-sözlüğü yaklaşımı, Porter algoritması, NLTK	G	9 büyük projedeki 10,829 konu	N/A
Katz ve ark.	2015	DA		Kontekst tabanlı yaklaşım	G	Otel ve film yorumları, Twitter gönderileri	TripAdvisor, IMDB, Twitter
Van de Kauter ve ark.	2015	DA	E	Yeni bir detaylandırılmış (fine-grained)	Poz/Ne g	Belli bir şirket hakkındaki	N/A

				grained) duygu açıklama şeması, veri sözlüğü yaklaşımı		haberler		
Fattah	2015	DS		Yeni bir ağırlıklandır ma şeması, Birçok sınıflandırıcı nı kombinasyo nu	Poz/Ne g	Film yorumları ve çoklu alanlı duygu veri seti	N/A	
Al-Rowaily ve ark.	2015	DA	E		Poz/Ne g	Sosyal form gönderileri	Dark Web forums, Aloka web forum	İspanya olca
Williams ve ark.	2015	DA	H	Naive Bayes, DVM	Poz/Ne g	Dil siteleri	Learn English Today, British National Corpus (BNC)	
Khan ve ark.	2016	DA	H	DVM,eSap	Poz/Ne g	Film yorumları	Large Movie Review, Cornell Movie Review, Multi- Domain Sentiment veri setleri	
Hur ve ark.	2016	DT	E	CART, YSA, DVR	Poz/Ne g	Film yorumları	Korea film market, NAVER Movie site.	Korece
Appel ve ark (2016)	2016	DA	H	Naive Bayes, Maksimum Entropi	Poz/Ne g	Sosyal medya ve film yorumları	Twitter veri seti, Film yorumları veri seti	
Pröllochs ve ark.	2016		E	Hidden Markov modeli, Baum- Welch, Algoritması	G	Finansal haberler	Thomson Reuters Datastream, Loughran and McDonald Finansal Duygu Sözlüğü	Almanca
Schumaker ve ark.	2016	DA	E	Central Support Model	Poz/Ne g	Futbol tarafı twitleri	Twitter veri seti	
Meire	2016	DA	E	DVM, Random	Poz/Ne g	Sosyal medya	Facebook	Flemenççe

			Forest Sınıflandırma		yorumları			
Chan ve Chong	2016		H	KA(Karar Ağaçları), DVM	Poz/Neg	Film yorumları, Finansal metinler	Rotten Tomatoes web sitesi, Finet Hong Kong finans web sitesi	Çince
Poria ve ark.	2016	ÖS	E	ELM, SVM(DVM), ANN(YSA)	G	Sosyal medya	Youtube veri seti	
Sun ve ark.	2016	ÖS	H	DNN	Poz/Neg	Sosyal medya	Sina Mikroblog	Çince
Khan ve ark.	2016		H	DVM	Poz/Neg		Benchmark datasets, Cornell Movie Datasets, Large Movie Review	
Deng ve ark.	2017	DA	E	General Inquirer (GI), MPQA,	Poz/Neg	Sosyal medya, Finans siteleri	Twitter veri seti	
Giatsoglou ve ark.	2017	DT	H	DVM	G	Film yorumları, ürün yorumları	Greek e-shopping site, MOBILE-PAR veri seti, Rotten Tomatoes, The Large Movie Review veri seti	Yunanca
Guo ve ark.	2017	DS	E	Logistic model, Thermal Optimal Path (TOP)	G	Sosyal medya	Xueqiu borsa sitesi	Çince
Wang ve ark.	2017	DA	H	DVM	G	Sosyal medya	Sina Weibo veri seti	Çince
Vilares ve ark.	2017	DA	H	SentiStrength, SO-CAL	Poz/Neg	Film yorumları, sosyal medya	Amazon	İspanyolca, Almanca
Chen ve ark.	2017	DA		CNN, RAE, RNTN, DCNN, NBSVM, tree-CRF		Ürün yorumları	Stanford sentiment treebank, customer reviews dataset	

Wu ve ark.	2017	DA	E	DSSC,ADM M,FISTA	Poz/Ne g	Ürün yorumları,	Amazon product review ve Twitter dataset veri setleri	
Liu ve ark.	2017	DA		PROMETH EE II	Poz/Ne g	Otomobille rle ilgili ürün yorumları	Bazhuayu, Kuan- gren web siteleri	Çince
Piryanı ve ark.	2017	DA		Naive Bayes ve DVM	Poz/Ne g	Sosyal medya	Twitter	

DUYGU SINIFLANDIRMASINDA ÖZİNİTELİK SEÇİMİ

DA'nın görevi duygu sınıflandırma problemi olarak düşünülebilir. DS'de ilk adım metnin özniteliklerini seçip çıkarmaktır. Bazı mevcut öznitelikler şunlardır (Aggarwal ve ark., 2012).

Terimlerin varlığı ve sıklığı: Bu öznitelikler bireysel kelimeler veya kelime n-gramları ve bunların sıklık sayılarıdır. Terimlerin varlığı ve sıklığı kelimelerin ikili ağırlıklarını (kelime varsa sıfır çıkar yoksa bir) veya kullanılan terimlerin sıklık ağırlıkları göreceli olarak özniteliklerin önemini gösterir (Yelena ve Padmini, 2011).

Konuşma parçaları: Görüşlerin önemli bir belirteci olan sıfatları bulma.

Görüş bildiren kelimeler ve deyimler: İyi veya kötü, hoşlanma ya da nefret belirten görüşleri ifade etmek için genelde kullanılan kelimelerdir. Öbür yandan bazı deyimler görüş belirten kelimeler kullanmadan da görüşleri ifade eder.

Olumsuz ifadeler: Negatif kelimeler görüşlerin yönelimini değiştirebilir.

Öznitelik seçme metotları

Öznitelik seçme metotları, insanların ek açıklamalarına ihtiyaç duyan veri sözlüğü tabanlı ve daha sıklıkla kullanılan statik metotlar olarak ikiye ayrılırlar. Veri sözlüğü tabanlı yaklaşımlar küçük bir “tohum” kelime seti ile başlarlar. Sonra benzerlik tespiti veya geniş bir sözlük elde etmek için çevrimiçi kaynaklar kullanarak bu sete önyüklemeye yaparlar. Bu Whitelaw ve ark (2005) tarafından da söylendiği gibi birçok zorluk içermektedir. Öbür yandan statik yaklaşımlar bütün olarak otomatiktirler.

Öznitelik seçme teknikleri dokümanları, bir Kelime Grubu (KG) olarak veya dokümanda kelime dizilerini tutan bir dizi olarak işlerler. KG

sınıflandırma işleminde daha basit olduğundan daha fazla kullanılır. En çok kullanılan öznitelik seçimi adımı durdurma kelimelerinin kaldırılması ve köklere ayırmadır. Bir sonraki alt bölümde, ÖS’de sık kullanılan üç tane istatistiksel metot gösterilmiştir. Bunların dışında ÖS’de kullanılan bilgi kazanımı ve Gini indeksi gibi metotlar vardır (Aggarwal ve ark., 2012).

Noktasal karşılıklı bilgi (NKB)

Karşılıklı bilgi ölçümü, öznitelik ve sınıflar arasındaki karşılıklı bilgiyi modellemeyi sağlar. Bu ölçüm bilgi teorisinden türetilmiştir (Cover ve Tomas, 1991). NKB $M_i(w)$, kelime w ile sınıf i , sınıf i ile kelime w arasındaki eşdizimlilik seviyesinin temelini üzerinde tanımlanır. Beklenen kelime w ile sınıf i arasındaki eşdizimlilik, karşılıklı bağımsızlık temelinde, $P_i * F(w)$ olarak verilir ve gerçek eşdizimlilik $F(w) * p_i(w)$ dir. Karşılıklı bilgi bu iki değer arasındaki oran olarak tanımlanır ve şöyle gösterilmiştir:

$$M_i(w) = \log \left(\frac{F(w) \cdot p_i(w)}{F(w) \cdot P_i} \right) = \log \left(\frac{p_i(w)}{P_i} \right) \quad (1)$$

$M_i(w)$ ‘nun 0’ dan büyük olduğu durumda w kelimesi sınıf i ile pozitif olarak ilişkilidir. w kelimesi sınıf i ile $M_i(w)$ ‘nun 0’ dan küçük olduğu durumda negatif ilişkilidir.

NKB birçok uygulamada kullanılmıştır ve NKB’de birçok iyileştirme yapılmıştır. NKB eşdizimlilik kuvveti olarak düşünülür. Yu ve ark (2013) basit NKB modelini bağlamsal entropi modeli ile geliştirmişlerdir. Geliştirdikleri bağlamsal entropi modeli, bir entropi ölçümünü kullanarak kelimelerin bağlamsal dağılımlarını karşılaştırıp iki kelime arasındaki benzerliği ölçer; bu kelimelerin kök kelimeleri ile olan benzerliğinin keşfine olanak sağlar. Kök

kelimeleri bir kez genişletilince hem kök kelimeleri hem de genişletilmiş kelimeler yeni makalelerin duygularını sınıflandırmak için kullanılır. Kullandıkları metotların daha yararlı duygu kelimelerini keşfettiği gösterilmiştir. Yapılan bu çalışmalar, NKB bazlı geliştirilmiş metotları üstün hale getirmiştir.

Ki-kare (χ^2)

n , koleksiyondaki toplam doküman sayısıdır, $p_i(w)$, w içeren dokümanlar için sınıf i 'nin şartlı olasılığıdır. P_i , sınıf i 'yi oluşturan dokümanların global kesridir ve $F(w)$, kelime w 'yu içeren dokümanların global kesridir. Bu yüzden kelime w ile sınıf i arasındaki kelime (χ^2)-istatistik şöyle gösterilmiştir (Aggarwal ve ark., 2012):

$$\chi^2_i = \frac{n.F(w)^2.(p_i(w)-P_i)^2}{F(w).(1-F(w)).P_i.(1-P_i)} \quad (2)$$

χ^2 ve NKB, terimler ve kategoriler arasındaki korelasyonu ölçen iki farklı yoldur. χ^2 normalleştirilmiş bir değerde NKB'den daha iyidir. Bu yüzden bu değerler aynı kategoride daha karşılaştırılabilir terimlerdir (Aggarwal ve ark., 2012).

χ^2 birçok uygulamada kullanılmaktadır. Bunlardan biri Fan ve Chang (2011) 'in gösterdiği bağlamsal reklamcılıktır. Fan ve Chang (2011) çevrimiçi bağlamsal reklamcılığı geliştirmek için bloggerların acil kişisel çıkarlarını-İlgilerini keşfetmişlerdir. Gerçek reklamlarla ve ebay.com, wikipedia.com ve epinions.com gibi aktif blog sayfaları üzerinde çalışmışlardır. Sınıflandırma için Destek Vektör Makinesi (DVM) ve ÖS için χ^2 kullanmışlardır. Buldukları sonuçlar, metotlarının reklamları bloggerların kişisel ilgileri ile pozitif korelasyon yaparak etkili bir biçimde tanımlandığını göstermektedir.

Hagenau ve ark (2013) stok market verilerini dikkate alarak kendi özellik seçimlerinin bir parçası olarak market geri besleme özelliklerini kullanmışlardır. Sonra bunları χ^2 ve İki Normal Ayırma (İNA) ile kullanmışlardır. Gürbüz özellik seçiminin kompleks özellik tipleri ile birleştirildiğinde sınıflandırma doğruluğunu önemli bir ölçüde arttırdığı görülmüştür. Yaklaşımları semantik olarak ilgili yapıları seçmeye izin verir ve bir Makine Öğrenmesi (MÖ) yaklaşımı uygulandığında aşırı uyma problemini azaltır.

Sınıflandırıcı olarak DVM kullanmışlardır. Elde ettikleri sonuçlarda ileri özellik çıkarım metotlarının ve kendi geri besleme tabanlı özellik seçimlerinin kombinasyonunun sınıflandırma doğruluğunu arttırdığı ve duygu analitiğini geliştirmeye izin verdiği görülmüştür. Bunun sebebi yaklaşımlarının daha az-açıklayıcı özelliklerin sayısını azaltmasıdır. Örneğin gürültü ve metin mesajlarına MÖ yaklaşımlarını uygularken aşırı uymanın negatif etkilerini sınırlandırmasıdır.

Gizli semantik indeksleme (GSİ)

Özellik seçme metotları, orijinal öznitelik setlerinden toplama yaparak verilerin boyutsallığını düşürmeye çalışırlar. Özellik dönüşüm metotları orijinal özellik setlerinin bir fonksiyonu olarak daha küçük özellik setleri oluştururlar. GSİ ünlü dönüşüm metotlarından biridir (Deerwester ve ark., 1990). GSİ metin boşluğunu orijinal kelime özelliğinin lineer bir kombinasyonu olan yeni bir eksen sistemine dönüştürür. Temel Bileşen Analizi (TBA) teknikleri bu hedefe ulaşmak için kullanılır (Jolliffe, 2002). Altta yatan öznitelik değerlerin çeşitlilik bilgisinin en büyük seviyesini sürdürecektir eksen-sistemine karar verir. GSİ'nin en önemli dezavantajı sınıf-dağıtımı temeline kör olan denetimsiz bir teknik olmasıdır.

ÖS'de kullanılabilecek Saklı Markov Modeli (SMM) ve Gizli Dirichlet Tahsisi (GDT) gibi birçok istatistiksel yaklaşım vardır. Bu yaklaşımlar, Duric ve Song (2012) tarafından inceleme dokümanındaki varlıkları, polarite bakımından tanımlayan subjektif ifadelerden ayırmak için kullanılmıştır. GDT dokümanları gözlemlenemeyen konular tarafından tanımlanmasına izin veren üretken modellerdir. SMM-GDT eşzamanlı model konuları ve doküman koleksiyonunun sözdizimsel yapılarıdır (Griffits ve ark., 2005). Özellik seçimi şemaları Duric ve Song (2012) tarafından önerilmiştir. Çalışma doküman polaritesi sınıflandırması konusunda önemli bir başarı elde etmiştir. Film yorumları üzerine çalışıp Maksimum Entropi (ME) sınıflandırıcısını kullanmışlardır.

ÖS'deki zorluklar

Özellikleri çıkarmadaki en büyük zorluklardan biri ironi algılamadır. Bu görevin objektifliği ironi yorumlarının tanımlanmasıdır. Reyes ve Rosso (2012) bu konuda bir çalışma

önermişlerdir. Bu tür subjektif yorumların altını çizmek ve göstermek için bir özellik modeli tanımlamayı ve göze çarpan ironi karakteristiklerini göstermeyi hedeflemişlerdir. Altı özellik kategorisinde sözel ironiyi göstermek için bir model kurmuşlardır: n-grams, POS-grams, komik profillemeye, olumlu/olumsuz profillemeye, duygusal profillemeye ve memnuniyet profillemeye. Amazon.com sitesinden haber ve satiric makaleler ile müşteri yazılarındaki ironik yorumlardan serbestçe kullanılabilecek bir veri seti oluşturmuşlardır. Çevrimiçi viral efektler üzerine; örneğin insanlar arasında zincirleme bir reaksiyonu tetikleyen içerikler gönderilmişlerdir. Sınıflandırma önerileri için NB, DVM ve DT kullanmışlardır. Bu üç sınıflandırıcı ile doğruluk, hassaslık, anma ve F-ölçütü bakımından tatmin edici sonuçlar elde edilmiştir.

DUYGU SINIFLANDIRMA TEKNİKLERİ

Duygu Sınıflandırma (DS) teknikleri; MÖ yaklaşımı, veri sözlüğü tabanlı yaklaşım ve hibrid yaklaşım olarak üçe bölünebilir (Diana ve Adam, 2011). MÖ yaklaşımı, popüler MÖ algoritmalarını ve dilbilimsel özellikleri uygulamıştır. Veri sözlüğü tabanlı yaklaşım daha önceden bilinen ve derlenmiş olan terimlerden oluşan bir duygu sözlüğüne dayanmaktadır. Bu yaklaşım, sözlük tabanlı yaklaşım ve duygu kutuplarını bulmak için istatistiksel ve semantik metotları kullanan korpus tabanlı yaklaşım olmak üzere bölünür. Hibrid yaklaşım bu iki yaklaşımın birleşmesidir ve duygu veri sözlükleriyle birlikte önemli bir rol oynarlar. Çeşitli yaklaşımlar ve popüler DS algoritmaları Şekil 2'de gösterilmiştir.

MÖ yaklaşımı kullanan metin sınıflandırma metotları denetimli ve denetimsiz öğrenme metotları olarak ikiye ayrılabilirler. Denetimli metotlar büyük sayıda etiketli eğitim dokümanlarını kullanılabilir hale getirirler. Denetimsiz metotlar ise bu etiketli eğitim dokümanlarını bulunmasının zor olduğu durumlarda kullanılırlar

Veri sözlüğü tabanlı yaklaşım, metni analiz etmek için kullanılan fikir veri sözlüğünü bulmaya bağlıdır. Bu yaklaşımda iki farklı metot vardır. Sözlük tabanlı yaklaşım eşanlamlılık ve zıt anlamlılıklardan oluşan sözlükleri araştırır ve fikir tohum kelimeleri bulmaya dayanan bir yaklaşımdır. Korpus tabanlı yaklaşım fikir kelimelerinin tohum listesi ile başlar ve sonra geniş bir korpusta diğer fikirleri bulur. Bu

istatistiksel ve semantik metotlar kullanılarak yapılabilir. Bu iki yaklaşımın kısa tanımları ve ilgili araştırmalar alt bölümde gösterilecektir.

Makine öğrenme yaklaşımı

MÖ yaklaşımı, sözdizimsel ve/veya dilbilimsel özelliklerin kullanılmasını yapan normal bir metin sınıflandırma problemi olan DA'yı çözmek için kullanılan MÖ algoritmalarına dayanır.

Metin Sınıflandırma Problemi Tanımı: Eğitim kayıtları seti $D = \{X_1, X_2, \dots, X_n\}$ her bir kayıt bir sınıfa etiketlenmiştir. Daha sonra bilinmeyen bir sınıf örneği için bu model bu sınıfa bir etiket tahmini yapmak için kullanılır. Zor bir sınıflandırma problemi bir örneğe sadece bir etiket atanmasıdır. Daha kolay bir sınıflandırma problemi ise bir örneğe etiketlerin olasılıksal bir değerinin atanmasıdır.

Denetimli öğrenme

Denetimli öğrenme metodu etiketlenmiş eğitim dokümanlarının varlığına bağlıdır. Literatürde birçok çeşit denetimli sınıflandırıcı vardır. Nizam ve Akın (2014) yaptıkları çalışmada denetimli öğrenme yaklaşımını kullanarak sosyal medya DA ile ilgili sonuçlar elde etmişlerdir. Eğiticiyiz yöntem kullanılarak Tweeter'daki tüm kelimeler ele alınmış yani analiz edilmiştir. Tweetler pozitif, negatif ve nötr olarak işaretlenerek 3 sınıfta toplanmıştır. Bazı sınıflandırma algoritmaları kullanılarak çeşitli sonuçlar elde edilmiştir. Sonraki alt bölümlerde DA'da sık kullanılan sınıflandırıcılar anlatılmıştır.

Habernal ve ark (2014), Çek sosyal medyasının DA için MÖ metotları üzerine derinlemesine bir çalışma yapmışlardır. Çalışma sırasında birçok metot, ön işlemeli teknikler keşfedilmiş ve çeşitli özellik ve sınıflandırıcılar çalıştırılmıştır. Ayrıca beş farklı özellik seçimi algoritması kullanılmış ve adlandırılmış varlık tanınması ile DA'da ön işlemenin performansı araştırılmıştır. Çalışma sonucunda yararlı sonuçlar elde edilmiştir.

Olasılıksal sınıflandırıcılar

Olasılıklı sınıflandırıcılar sınıflandırmanın karışım modellerini kullanırlar. Karışım modeli her bir sınıfın karışımın bir bileşeni olduğunu farz eder. Her bir karışım bileşeni bu bileşen için belirli örnekleme olasılığını sağlayan üretici

modellerdir. Bu tür sınıflandırıcılara üretken sınıflandırıcılar da denir. Bir sonraki alt bölümde üç tane popüler olasılıklı sınıflandırıcı tartışılmıştır.

Naive bayes sınıflandırıcısı (NB)

Naive Bayes sınıflandırıcısı oldukça basit ve çok kullanılan bir sınıflandırıcıdır. NB sınıflandırıcısı modeli bir sınıfın sonraki olasılığını hesaplar ve dokümandaki kelimelerin dağılımını temel alır. Model dokümandaki kelimenin pozisyonunu yok sayan bir KG özellik çıkarımı ile çalışır. Belirli bir etikete ait verilen bir özellik setine bir olasılık tahmin etmek için Bayes Teoremi kullanılır:

$$\frac{P(etiket \setminus \text{özellikler})}{P(etiket) * P(\text{özellikler} \setminus etiket)} = \frac{P(\text{özellikler})}{P(etiket)} \quad (3)$$

$P(etiket)$ etiketin önceki olasılığıdır ya da etiketi rastgele özellik kümeleme ihtimalidir. $P(\text{özellikler} \setminus etiket)$, verilen bir özellik setini bir etiket olarak sınıflandırma ön olasılığıdır. $P(\text{özellikler})$ ise verilen bir özellik setinin oluşma ön olasılığıdır. Her bir özelliği bağımsız olarak belirten bir Naive varsayımı denklemi yeniden şöyle yazılabilir:

$$\frac{P(etiket \setminus \text{özellikler})}{P(etiket) * P(f_1 \setminus etiket) * \dots * P(f_n \setminus etiket)} = \frac{P(\text{özellikler})}{P(etiket)} \quad (4)$$

Kang ve ark (2012) pozitif sınıflandırma doğruluğu sağlama yönelimi problemini çözmek için bir NB sınıflandırıcı önermişlerdir. Bu, iki sınıfın doğruluklarının ortalama bir değer olarak belirttikleri durumlarda ortalama doğruluğun azalması problemini doğurur. Kang ve ark (2012) bu algoritmayı restoran yorumlarında kullanarak pozitif doğruluk ile negatif doğruluk arasındaki boşluğu daraltmışlardır. Bu sonuçları NB ve DVM ile karşılaştırmışlardır. Hatırlama hassaslıktaki doğruluk NB ve DVM ile karşılaştırılmıştır.

Bayesian ağ (BA)

NB sınıflandırıcısının en temel varsayımı özelliklerin birbirinden bağımsız olmasıdır. Tersine varsayım ise özelliklerin bütünen bağımlı olmasıdır. BA, değişkenler ve onların ilişkileri için eksiksiz bir model olarak düşünülür. Bu yüzden tüm değişkenler için eksiksiz ortak

olasılık dağılımı bir model için belirtilmiştir. Metin madenciliğinde BA'nın hesaplama karmaşıklığı çok pahalıdır ve bu yüzden çok kullanılmaz (Aggarwal ve ark., 2012).

Ortigosa ve ark (2012) BA'yı üç farklı hedef değişken tarafından karakterize edilmiş yazarın tutumunu içinde barındıran gerçek dünya probleminde kullanmıştır. Çok-boyutlu BA sınıflandırıcısını önermişlerdir. Bu farklı hedef değişkenlerini, aralarındaki potansiyel ilişkilerden faydalanmak için aynı sınıflandırma görevi için birleştirmiştir. Bu konteksteki büyük miktardaki uygun etiketsiz bilginin avantajını almak için yarı denetimsiz alana çok boyutlu bir sınıflandırma çerçevesi çıkarmışlardır. Kendi yarı denetimli çok boyutlu yaklaşımlarının birçok DA yaklaşımından daha iyi performans verdiğini göstermişlerdir.

Maksimum entropi sınıflandırıcısı (MES)

Maxent Sınıflandırıcı (koşullu üssel sınıflandırıcı olarak bilinir) etiketlenmiş özellik kümelerini vektör kullanan kodlara çevirir. Bu kodlanmış vektör sonradan her bir özellik için ağırlıklandırmayı hesaplamada kullanılır ve sonra özellik seti için etikete karar verme ile birleştirilebilir. Bu sınıflandırıcı $X\{kodlayıcı\}$ tarafından oluşturulmuş özellik seti ile birleştirilmiş $X\{ağırlıklar\}$ seti olarak ifade edilmiştir. Özellikle her bir $C\{(\text{özellik seti}, etiket)\}$ kodlama haritası bir vektörle çiftleştirilir. Her bir etiketin olasılığı şu denklemle bulunur:

$$P(f_s \setminus etiket) = \frac{içcarpım(ağırlıklar, kodlama(f_s, etiket))}{toplaml(içcarpım(ağırlıklar, kodlama(f_s, l)))} \quad (5)$$

MES sınıflandırıcı Kaufmann (2012) tarafından herhangi bir dil çifti ile küçük eğitim verileri arasındaki paralel cümleleri bulmak için kullanılmıştır. Paralel olmayan korpora kullanan dil tekniklerini kullanan veya belli miktar eğitilmiş verilere ihtiyaç duyan otomatik olarak paralel veri çıkarmak için geliştirilmiş başka teknikler vardır. Sonuçlar MES sınıflandırıcıların hemen hemen tüm dil çiftleri için yararlı sonuçlar ürettiğini göstermiştir. Bu birçok yeni dil için paralel korpora oluşumuna izin verebilir.

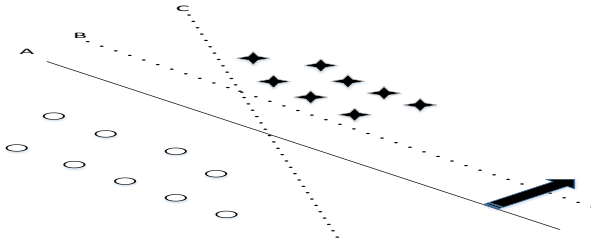
Lineer sınıflandırıcılar

$\bar{X} = \{x_1, \dots, x_n\}$ normalize edilmiş kelime frekansı, $\bar{A} = \{a_1, \dots, a_n\}$ özellik uzayındaki gibi

aynı boyutsallıktaki lineer katsayı vektörüdür ve b skalerdir; lineer öngörücü çıktısı $p = \bar{A}\bar{X} + b$ şeklinde tanımlanır ve bu lineer sınıflandırıcının çıktısıdır. Tahmin edici p iki farklı sınıf arasındaki ayrıştırıcı hiperdüzlemdir. Birçok lineer sınıflandırıcı çeşidi vardır. Bunlardan biri iki farklı sınıf arasında iyi lineer ayrıştırıcılara karar vermeye çalışan DVM'dir (Cortes ve Vapnik, 1995; Vapnik, 1995). İki popüler lineer sınıflandırıcı devam eden bölümlerde gösterilmiştir.

Destek vektör makineleri (DVM) sınıflandırıcısı

DVM'nin temel prensibi farklı sınıfları en iyi biçimde ayrıştırarak arama uzayındaki en iyi lineer ayrıştırıcıya karar vermektir. Şekil 3'te 2 tane x sınıfı vardır, o ve üç tane hiperdüzlem A, B ve C vardır. Hiperdüzlem A sınıflar arasındaki en iyi ayrışmayı sağlar, çünkü herhangi veri noktalarının normal mesafesi en büyüktür ve bu maksimum ayrıştırma marjını gösterir.



Şekil 3. Bir sınıflandırma probleminde destek vektör makinesinin kullanımı

Metin verisi metinlerin aralıklı doğasından kaynaklı DVM için sınıflandırmaya uygundur. Fakat bazı özellikleri düzensizdir. Genel olarak biri diğeri ile ilişki kurma eğilimindedir ve lineer olan ayrıştırılabilir kategoriler şeklinde organize olmuşlardır (Joachims, 2002). DVM veri örneklerini lineer olmayan şekilde eşleştirme yaparak orijinal arama uzayında lineer olmayan bir karar yüzeyi inşa eder (Aizerman, 1964).

Chen ve Tsehng (2011) iki çoklu-sınıflı DVM temelli yaklaşım kullanmışlardır. Çalışmalarında bire-karşı-hepsi DVM'si ve tekli makine çoklu-sınıf DVM'si yorumları kategorize eder. Chen ve Tsehng ürün yorumlarındaki bilgi kalitesini hesaplamak için bir metot önermişlerdir. Aynı zamanda bilgi odaklı özellik setlerini bulmak için bilgi kalite çerçevesi benimsemişlerdir. Dijital kamera ve MP3 yorumları üzerine çalışmışlardır. Buldukları

sonuçlar kullandıkları metodun yorumları kalitelerine göre doğru bir şekilde sınıflandırdığını göstermiştir

DVM'ler Li ve Li (2013) tarafından duygu polarite sınıflandırıcı olarak kullanılmıştır. İkili sınıflandırma probleminden farklı olarak fikir subjektifliğini tartışmışlardır. Mikro blog platformlardaki kompakt nümerik fikir özetleri üzerine bir çerçeve önermişlerdir. Kullanıcıların sorgulamalardan çıkan fikirlerle ilişkili konuları çıkarmış ve kullanmışlardır sonra bu fikirleri DVM kullanarak sınıflandırmışlardır. Twitter gönderileri üzerinde çalışarak tecrübe edinmiş ve kullanıcı güvenilirliğinin ve fikir öznelliğinin mikro-blog fikirleri kümelemek için gerekli olduğunu bulmuşlardır. Gerçek zamanda işin farklı yönlerindeki harici fikirleri izlemek için bir izleme sistemi kurarak, destekleyici-karar yapıcılar için kullandıkları mekanizmanın etkili bir biçimde market zekâsı keşfettiğini ispatlamışlardır.

Emma ve ark (2013) yaptıkları çalışmada DA'da metin ön işlemenin önemini keşfetmişlerdir. Elde ettikleri deney sonuçlarında uygun özellik seçimi ile DVM'yi kullanarak DA doğruluğunu ciddi bir oranda geliştirmişlerdir. DA'nın zor bir problem olması düşünüldüğünde konu kategorizasyonu konusunda literatürde ki çalışmalarla karşılaştırıldığında önemli sonuçlar elde etmişlerdir.

Smailovic ve ark (2014) yaptıkları çalışmada ilgili şirketleri ve onların ürünlerini tartışan, bunlarla ilgili olan tweetlerin stok fiyatlarının değişimini gösterebileceğini analiz etmişlerdir. Çalışmalarında aktif bir öğrenme yaklaşımı geliştirilmiş ve stok market alanı ile ilgili tweetleri DA'da kullanılmıştır. Statik Twitter verileri analiz problemi ortaya konulmuş ve DVM sınıflandırıcısı kullanılmıştır. Twitter gönderilerini DVM sınıflandırıcı kullanarak sadece pozitif ve negatif değil bunlara ek olarak nötr şekilde kategorize etmişlerdir. Sonuçlar pozitif duygulardaki değişim ihtimalinin stok kapanış fiyatlarının bir göstergesi olarak kullanılabileceğini göstermiştir.

Kranjc ve ark (2014) yaptıkları çalışmada bir bulut tabanlı bilişimsel iş akışı platformu olan ClowdFlows'su anlatmış ve bu platformun uzantılarının veri akışlarını analiz edebileceğini ve aktif öğrenme yapabileceğini göstermişlerdir. Dahası ClowdFlows'da veri ve iş akışından faydalanarak örneklerin etiketlenmesi kitle

kaynak kullanımı boyunca dağıtılabilir. Burada aktif öğrenme lineer bir DVM ile kullanılarak duygu sınıflandırmada kullanılmıştır. Ayrıca microblogging veri akılarında uygulanmıştır.

Meire ve ark (2016) ilk defa yaptıkları çalışmada kullanıcı gönderilerine ek olarak kullanıcının ön tanımlayıcı (kullanıcı profili ve önceki gönderiler) ve sonraki (beğenmeleri ve retweetleri arası etkileşim) bilgileri de dikkate alan bir yol izlemişlerdir. Bunun birçok nedeni olduğunu düşünmüşlerdir: İlk olarak sosyal medyada argo kelimelerin fazla olması geleneksel yöntemlerin başarıları oranlarını etkilemektedir. İkinci olarak ön tanımlayıcı kullanıcı bilgisi sayılabilecek kişinin refahı, ruh hali ve demografik bilgileri gelecek duygusu için önemli bir bilgi olabilir. Üçüncü olarak gönderiden sonraki yorumlar, beğeniler ile gönderi duygusu arasında bir bağlantı vardır. Yapılan çalışmada DVM ve Random Forest Sınıflandırma (RFS) algoritması kullanılmış ve kullanıcının ön tanımlayıcı bilgilerinin ve gönderi sonrası bilgilerinin duygu analizi modeline doğru tahmin yönünden katkı sunduğu ortaya çıkmıştır.

Sinir ağı sistemi (SAS)

Sinir Ağı, nöronların temel birim olduğu birçok nörondan oluşmaktadır. Nöronlara girişler i 'ninci dokümandaki kelime frekansı olan vektör X_i tarafından ifade edilir. Bir fonksiyonun $f(\bullet)$ girdilerini hesaplamak için her bir nöronla ilişkili A ağırlıkları seti vardır. Sinir ağının lineer fonksiyonu, $p_i = A \cdot \bar{X}_i$ dur. Bir ikili sınıflandırma probleminde, $\bar{X}_i$ sınıf etiketinin y_i tarafından ifade edildiği farz edilir ve p_i tahmin edilen fonksiyonunun işaretli sınıf etiketini üretir.

Çok-katmalı sinir ağları, lineer olmayan sınırlar için kullanılır. Bu çoklu-katmanlar belirli bir sınıfa ait kapalı bölgelere yaklaşmak için kullanılan çoklu parçalı lineer sınırlar oluşturmak için kullanılır. Önceki katmanlardaki nöronların çıktıları sonraki katmanlardaki nöronların içini besler. Eğitim süreci çok karmaşıktır, çünkü hataların birçok katman üzerinden geri yayılması gerekmektedir. Metin verileri için birçok SAS uygulaması vardır (Ruiz ve Srinivasan, 1999; Ng ve ark., 1997).

DVM ve Yapay Sinir Ağları (YSA) arasında deneysel bir karşılaştırma Moraes ve ark (2013) tarafından doküman-seviye DA dikkate alınarak yapılmıştır. Bu karşılaştırmayı, YSA'ların duygu öğrenmesinde daha az dikkat çeken bir yaklaşım

olmalarından ve DVM'nin SAS'da geniş ve başarılı bir şekilde kullanılmasından dolayı yapmışlardır. İki yaklaşımın daha iyi seviyede sınıflandırma doğruluğu elde edebildiği durumlardaki; ihtiyaçları, sonuç modellerini ve konteksleri tartışmışlardır. Ek olarak geleneksel KG modelinde özellik seçimi ve ağırlıklandırma için popüler denetimli metotlar ile standart bir hesaplama konteksi ben imsemişlerdir. Deneyler bazı dengesiz veri kontekslerinin dışında YSA'nın DVM'den daha üstün sonuçlar verdiğini göstermiştir. Üç tane kıyas veri seti ve amazon.com sitesinden alınan film, gps, kamera ve kitap yorumları üzerinde denemişlerdir. Film yorumlarında YSA'nın DVM'den istatistiksel olarak önemli bir fark oluşturduğu görülmüştür. Her iki metotta, DA literatürün de nadiren tartışılan bazı potansiyel sınırlamalara onay vermişlerdir. Örneğin yürütme zamanında DVM ve eğitim durumunda YSA'nın hesaplama tutarı gibi. Bilgi kazanımının (hesaplaması ucuz özellik seçimi metodu) kullanılmasının sınıflandırma doğruluğuna önemli bir etkisi olmadan hesaplama performansını düşürdüğünü ispatlamışlardır.

DVM ve SAS, Camp ve Bosch (2012) tarafından gösterildiği gibi biyografiksel metinlerdeki kişisel ilişkileri sınıflandırmada kullanılmışlardır. İki insan arasındaki ilişkileri pozitif, sinirsel veya bilinmeyen olarak belirmişlerdir. Çalışmaları insanları belirli bir etki alanında, alanda ve zaman çerçevesinde tanımlayan tarihsel biyografiksel bilgileri temel almıştır. Kullandıkları sınıflandırıcıların bu ilişkileri iyi bir oranda etiketleyebildiği gösterilmiştir. Bir eğitim setinin ilişkileri içerdiği, birden fazla kişiyi çevrelediği ve bir tane belirli varlığa odaklanan kümeden daha iyi sonuçlar verdiğini bulmuşlardır. DVM'nin ve tek katman SAS (1-SAS) algoritmasının en yüksek skoru aldığını ispatlamışlardır.

Jebaseeli ve Kirubakaran (2013) yaptıkları çalışmada fikirleri sınıflandırmak için yapay sinir tabanlı algoritmalarını Genetik Algoritma (GA) kullanarak geliştirmişlerdir. Bu teknikler, genetik olarak optimize edildiklerinde kesinlik üzerinde etkili olmuşlardır. Rastgele üretilen momentumun seçiminde kullanılan uygunluk değerine yani optimize edilmiş momentuma ve öğrenme oranlarına karar verilmiştir. Sonuçlar diğer algoritmalar ile karşılaştırılmış ve sınıflandırma kalitesi, hassasiyet, anma ve F-

ölçütü ile ölçülmüştür. Sonuçlar genetik tabanlı YSA'nın diğer sınıflandırıcılara kıyasla daha iyi olduğunu göstermiştir.

Kısa uzunluktaki mikroblog metin mesajları mikrobloglardan öznitelik çıkarılma işlemini zorlaştırmaktadır. Çince yazılan mikrobloglardaki bu zorluğu aşmak için Sun ve Ren (2016) öznitelik çıkarma işleminde bir mikroblog diyalogundaki gönderileri ve ilgili yorumları birleştirerek bir içerik genişletme çerçevesi önermişlerdir. Mikroblog diyaloglarından gönderilerin öznitelikleri olarak sayılabilecek bir bağlamsal bilgi çıkarmak için yeni bir erişimli otomatik kodlayıcı oluşturulmuş ve sinir ağının yapısını başlatmak için birkaç katmanlı Kısıtlı Boltzmann Makinesi (KBM) ile özelleştirilmiş Derin Sinir Ağı (DSA) modeli uygulanmıştır.

Karar ağacı sınıflandırıcıları

Quin (1986), Karar Ağacı (KA) sınıflandırıcıların öznitelik değerinin bölündüğü şartlarda, eğitim veri uzayının hiyerarşik olarak ayrışmasını sağlamışlardır. Koşul ya da yüklem bir ya da daha fazla kelimenin varlığı ya da yokluğudur. Yaprak düğümler sınıflandırmanın amacı için kayıtların kesin minimum sayılarını içerene kadar veri uzayının bölünmesi özinelemeli olarak devam eder.

Dokümanların parçalanması için kullanılabilen terimler seti ile ilişkilendirilen dokümanların benzerliğine dayanan farklı türde yüklemeler vardır. Üç çeşit tek özellik bölme vardır (Lewis ve Ringuette, 1994). Bu bölme uygulamak için ağaçtaki belli bir düğümdeki belirli kelimelerin ya da ifadelerin var olmasında veya yok olmasında kullanılır. Benzerlik tabanlı çoklu-nitelikli bölme dokümanları kullanır veya sık kelime kümelerini ve belgelerin benzerliğini bu kelime kümelerini bölmek için kullanır (Chakrabarti ve ark., 2003).

Metin sınıflandırmadaki KA uygulamaları ID3 VE C4.5. gibi standart paketler üzerinde küçük varyasyonlar olmaya eğilimlidir. Li ve Jain (1998) C5 algoritmasını kullanmışlardır. Ağaç konseptine bağlı olarak; Hu ve Wenjie (2010) tarafından kelime-seviyesi kontekstindeki konu ile ilgili terimlerin içerik yapılarını araştırıp bulmak için bir yaklaşım önerilmiştir. Bunu da, konu ile ilgili terim "t" ve onun kontekst kelimeleri arasındaki linkleri keşfetmek için Maksimum Açılım Ağacı (MAA) yapısını

kullanarak yapmışlardır. Buna göre duygu sınıflandırması için Konusal Terim Tanımlama Modeli (KTTM) geliştirmişlerdir. Kendi tanımlamalarında "konu ile ilgili terimler" belirlenmiş varlıklardır ya da belli bir alanda varlıkların belirli yönleridir. Konu ile ilgili terimlerin otomatik çıkarımını tanıtmışlardır. Daha sonra bu çıkarılan terimleri doküman konularını ayırt etmek için kullanmışlardır. Bu yapı duygu bilgisini aktarır. Yaklaşımları diğer MÖ algoritmalarından farklıdır ayrıca pozitif veya negatif bağlamsal bilgiyi etkili bir biçimde öğrenme becerisi vardır.

Grafik tabanlı bir yaklaşım Yan-Yan ve ark (2010) tarafından önerilmiştir. Kelime yapılarının içini ve dışını birleştirmek için yayılım yaklaşımı sunmuşlardır. Bu iki cümle özelliği, doküman içi ve dokümanlar arası ilişkinin kanıtıdır. Bir yorum cümlesindeki duygu oryantasyonuna karar vermek için cümlenin içindeki özelliklerden daha fazlasına ihtiyaç duyulur. Yan ve Bing kamera alanı üzerinde çalışmışlardır ve kendi metodlarını denetimsiz ve denetimli olan (NB, DVM) yaklaşımları ile karşılaştırmışlardır. Elde ettikleri sonuçlar önerdikleri yaklaşımın dış cümle özelliğini kullanmadan her iki yaklaşımdan daha iyi sonuç elde ettiğini göstermiştir.

Kural-tabanlı sınıflandırıcılar

Kural tabanlı sınıflandırıcılarda veri uzayı bir kural seti ile modellenmiştir. Sol el tarafı (sol taraf), sağ el tarafının (sağ taraf) sınıf etiketi olduğu bir durumda, ayırıcı normal formda özellik setindeki bir şartı gösterir. Şartlar terimlerin varlığındadır. Terimlerin yokluğu durumu nadiren kullanılır çünkü seyrek veride bilgilendirici değildir.

Kural üretmek için birçok sayıda kriter vardır. Bu kriterlere bağlı olarak eğitim safhası tüm kuralları inşa eder. En genel iki kriter destek ve güvendir (Liu ve ark., 1998). Destek kriteri kurallara uygun olan eğitim setindeki mutlak örnek sayıdır. Güven, sol el kuralı tatmin olduğunda sağ el kuralının da tatmin olma yapısal şartı ile ilgilidir. Birleştirilmiş kural algoritmaları önerilmiştir (Medhat ve ark., 2008).

Hem karar ağaçları ve hem de karar kuralları özellik uzayında kuralları kodlamaya yönelimlidir, fakat karar ağacı bu amaca hiyerarşik yaklaşım ile ulaşmaya eğilimlidir. Quinlan (1986) karar ağacı ve karar kuralı problemlerini tek bir çerçevede çalışmıştır. Bir

metin örneğinin sınıflandırılması için karar ağacındaki belirli bir yol bir kural olarak düşünülebilir. Karar ağaçları ve karar kuralları arasındaki en önemli fark DT'nin veri uzayının katı hiyerarşik bölünmesidir.

Zayıf, yarı ve denetimsiz öğrenme

Metin sınıflandırmanın ana amacı dokümanları belirli sayıda ön tanımlı kategori ile sınıflandırmaktır. Bunu başarmak için denetimsiz öğrenmede geniş sayıda etiketlenmiş eğitim dokümanları kullanılır. Metin sınıflandırmasında, bu etiketli eğitim dokümanlarını oluşturmak bazen zordur fakat etiketsiz dokümanları toplamak kolaydır. Denetimsiz öğrenme metotları bu zorlukların üstesinden gelir. Ko ve Seo (2000) gibi bu alanda birçok çalışma önerilmiştir. Ko ve Seo (2000) dokümanları kelimelere bölen ve her bir kelimeyi, her bir kategori listesini ve cümle benzerlik ölçümünü kullanarak bölen bir metot önermişlerdir.

Zayıf ve yarı-denetimli kavram birçok uygulamada kullanılmıştır. He ve Zhou (2011) örneklerden ziyade özellik aşamasında zayıf denetim sağlayan bir strateji önermişlerdir. He ve Zhou (2011) var olan duygu veri sözlüğünden çıkarılmış ön bilgiyi duygu sınıflandırıcı model öğrenmesine dâhil ederek bir başlangıç sınıflandırıcı elde etmişlerdir. Etiketlenmiş özellik olan ön bilgilere işaret etmişlerdir ve bunları, modellerin tahminini etiketsiz özelliklerde sınırlandırmak için kullanmışlardır. Çalışmalarında düşünceleri berraklaştırmak için alana özgü polarite kelimelerini belirleyebilmişlerdir. Kelime polariteleri alandan alana farklı olabilmektedir. IMDB ve amazon.com sitelerinden aldıkları film yorumları ve çoklu-alan duygu veri setleri üzerinde çalışmışlardır. Elde ettikleri sonuçlar yaklaşımlarının diğer zayıf denetimli duygu sınıflandırma metotlarından daha iyi olduğunu göstermiştir ve metotları uygun ön bilgiler olduğu zaman herhangi bir sınıflandırma görevini yapabilmektedir.

Xianghua ve ark (2013) Çin'deki sosyal yorumlardaki farklı açılar ve aynı zamanda farklı boyutlarda ifade edilen duyguların otomatik olarak keşfedilmesi için denetimsiz bir yaklaşım kullanmışlardır. Çok-yönlü global sosyal yorum konularını keşfetmek için GDP modelini kullanmışlardır. Sonra yerel konuları ve yorum metni ile ilgili window kontekstini bölmeyi temel alan ilişkilendirilmiş duyguyu

çıkarmışlardır. Blog veri seti (2000-SINA) ve veri sözlüğünden (300-SINA Hownet) çıkardıkları sosyal yorumlar üzerine çalışmışlardır. Yaklaşımlarının konuyu ayırttırma ile ilgili iyi neticeler aldığı ve DA doğruluğunu geliştirdiği görülmüştür.

Ayoub ve arkadaşları çevrimiçi yorumlarda DA ve fikir madenciliği üzerine çalışmışlardır. Çalışmalarında denetimsiz alan ve dil bağımsız bir modeli yorumlardan açık ve kapalı yönleri tespit etmek için önermişlerdir. Elde ettikleri sonuçlar kullandıkları tekniklerin görevi başarılı bir şekilde yaptığını ve diğer tekniklerden iyi sonuçlar verdiğini göstermiştir (Ayoub ve ark., 2013). Bunların dışında semantik oryantasyon kullanan NKB 'ye bağlı olan farklı denetimsiz yaklaşımlarda vardır (Turney, 2002).

Duygu sınıflandırması, MÖ teknikleri ve sözcük tabanlı metodlarla veya bunların birleştirilmesi ile yapılabilir. Khan ve ark (2016) alandan bağımsız sözlüklerin etki alanını arttırmak için SWIMS adlı sözcük tabanlı yaklaşım ile MÖ yaklaşımını içeren bir çerçeve önermiştir. SentiWordNet duygu sözlüğünü öznelik seçmede kullanılmıştır. DVM öznelik ağırlıklarının öğrenilmesinde kullanılmış ve de akıllı model seçimi yaklaşımı da sınıflandırma performansını arttırmada uygulanmıştır.

Vilares ve ark (2017) kompozisyonel söz dimi kuralları tarafından yönlendirilen çok dilli duygu analizi için denetimsiz bir metot önermişlerdir. Önerdikleri bu metodu en yeni denetimsiz ve denetimli yaklaşımlarla karşılaştırmışlardır. Algoritmaları en çok kullanılan iki denetimsiz metottan daha iyi sonuçlar vermiştir.

Birçok konuda duygu analizi çalışmaları olmasına rağmen felaket olayları ile ilgili kullanıcılar tarafından ifade edilen duygunun polaritesinin ne olduğuna odaklanan çalışma sayısı azdır. Neppalli ve ark (2017) Sandy kasırgası sırasında atılan twitler için duygu analizi yapmış ve çevirim içi kullanıcıların duygularını gösteren kasırga merkezli bir coğrafi harita oluşturmuşlardır. Böylece kullanıcıların sadece bulundukları yere göre değil ayrıca kasırganın uzaklığına göre duygularının nasıl değiştiğini göstermişlerdir. Yapılan çalışmalar bir felâket sırasında ortaya çıkan duyguların, acil müdahale uzmanlarının felaket bölgesinin kendisi için daha güçlü bir farkındalık geliştirmesine yardımcı olabileceğini görülmüştür.

Meta sınıflandırıcı

Birçok durumda, araştırmacılar kendi çalışmalarında bir ya da daha fazla sınıflandırıcıyı çalışmalarını test etmek için kullanmışlardır. Bu çalışmalardan biri Lane ve ark (2012) tarafından önerilmiştir. Medya analizinde pozitif veya negatif uygunluk taşıyan dokümanların bulunma problemini çözmek için bir MÖ yaklaşımı önermişlerdir. Negatif ve pozitif örneklerin dağılımındaki dengesizlik, zamanla dokümanlardaki değişiklikler, etkili eğitim ve modeller için değerlendirme prosedürleri hedeflerine ulaşırken karşılaştıkları zorluklardır. Medya analiz şirketi tarafından üretilmiş üç veri seti üzerinde çalışmışlardır. Dokümanları iki yolla sınıflandırmışlardır. Bunlardan biri uygunluğun varlığını belirleme ve negatif veya pozitif uygunluğu değerlendirmedir. Ham metinden veri setleri oluşturmak için beş tane farklı tip özellik kullanmışlardır. DVM, K-en yakın komşu, NB, BN, DT, bir kural öğrenici ve diğerleri arasından en iyi olanını bulmak için birçok sınıflandırıcıyı test etmişlerdir. Eğitim verilerinde sınıf dağılımını dengelemenin performansı geliştirmede yararlı olabileceğini göstermişlerdir fakat NB bundan olumsuz etkilenebilir.

Twitter'dan akan verilere MÖ algoritmalarının uygulanması, Rui ve ark (2013) tarafından incelenmiştir. Çalışmalarında dinamik panel veri modeli kurarak Twitter Ağzı Kelimeleri'nin (TAK) film satışlarına etkisini araştırmışlardır. Sınıflandırma için NB ve DVM kullanmışlardır. En büyük katlıları tweetlerin tek olma karakteristiklerini düşünerek tweetleri sınıflandırmalarıdır. Ön-tüketim fikri (henüz ürünü almayanlar) ile tüketim sonrası fikri (ürünü almış olanlar) arasında ayırma yapmışlardır. Benchmark film yorumları ve twitter verileri üzerine çalışmışlardır. Twitter API kullanarak, Twitter TAK verileri ve Box-OfficeMojo.com sitesinden de film satış verilerini toplamışlardır. Buldukları sonuçlar, ürün satışları üzerine TAK'ın etkisinin ciddi oranda takipçisi olan twitter kullanıcılarına, az takipçisi olanlara göre daha fazla olduğunu göstermiştir. Ön tüketim TAK'ın film satışlarına tüketim sonrası TAK'tan daha etkili olduğunu bulmuşlardır.

Bir diğer sınıflandırıcı karşılaştırma çalışması, istatistiksel olarak Markov Model tabanlı sınıflandırıcıların uygulanmasıdır. Bu,

kelimeler arasındaki bağımlılıkları yakalamak ve birçok popüler sınıflandırıcının öngörü performansını geliştiren bir kelime hazinesi sağlamaktır. Bu iki-aşamalı öngörü algoritması Bai (2011) tarafından gösterilmiştir. İlk aşamada sınıflandırıcısı kelimeler arasındaki koşullu bağımlılıkları öğrenir ve bunları duygusal bir değişken için Markov Blanket Directed Acyclic Graph ile kodlar. İkinci aşamada ise yüksek çapraz-geçerlilik doğruluğu üretmek için algoritmasında ince ayar yaparak bir meta sezgisel strateji kullanır. IMDB'den alınan iki çevrimiçi film yorumu ve üç çevrimiçi haber koleksiyonu üzerine çalışmış ve sonra kendi algoritmasını DVM, NB, ME ve diğerleri ile karşılaştırmıştır. Metodunun öngörü özellikleri cimri seti tanımlayabildiğini ve duygu oryantasyonu ile ilgili daha iyi öngörü sonuçları bulduğunu göstermiştir. Bulduğu sonuçlar duyguların kelimeler arasındaki koşullu bağımlılıklar, anahtar kelimeler veya yüksek frekanslı kelimeler tarafından yakalandığını göstermiştir. Birçok örnekte modelinin karmaşıklığının nedeni lineer olmasıdır.

Denetimli ve denetimsiz yaklaşımlar birlikte kullanılabilirler. Bu yaklaşım Martin-Valdivia ve ark (2013) tarafından uygulanmıştır. Polarite sınıflandırma sistemi geliştirmek için bir tane meta-sınıflandırıcı önermişlerdir. Paralel olarak İngilizceye çevrilen korpus ile İspanyolca film yorumlarının korpusu üzerine çalışmışlardır. İlk olarak bu iki korporayı kullanarak iki ayrı model üretmişlerdir. Sonra DVM, NB, C4.5 ve diğer MÖ algoritmalarını uygulamışlardır. Sonra İngilizce korpusa SentiWordNet duygu korpusunu entegre ederek semantik oryantasyon yaklaşımını kullanan bir yeni bir denetimsiz model üretmişlerdir. Bundan sonra ise bir meta-sınıflandırıcıyı kullanarak üç sistemi birleştirmişlerdir. Tek korpus kullananlara göre buldukları sonuçlar daha iyi olmuştur ve paralel korporaların olduğu durumlarda polarite sınıflandırmasın için yaklaşımlarının iyi bir strateji olarak ortaya çıkmıştır.

MÖ sınıflandırıcıları Walker ve ark (2012) tarafından sınıflandırma tutumu için kullanılmışlardır. Tutum bir nesne, fikir ya da konuma yönelmiş olan kişi tarafından tutulan pozisyon olarak tanımlanmıştır (Samusanduran ve Wiebei, 2009). Tutum bir perspektife benzer ve örneğin bir politik karara karşı olan bir konuşmacının tarafını tanımlarken görülebilir.

Walker ve arkadaşları insanların takındığı ve politik tartışmalara uyguladığı tutumları sınıflandırmışlardır. Convinceme.com sitesindeki 14 farklı tartışma konusu için 103 karşılıklı tartışmadan faydalanmışlardır ve konuşmacıların durumlarını ve tutumunu açıklamaya çalışmışlardır. Temel hedefleri tartışma konulu sınıflandırmaya bağlamsal diyalog özelliklerinin potansiyel katkısının ne olduğudur. Sınıflandırma için DVM, NB ve kural tabanlı sınıflandırıcı kullanmışlardır. Tartışma-araflı sınıflandırma doğruluklarına ulaşmışlardır.

Bravo-Marquez ve ark (2014) yaptıkları çalışmada tweetlerdeki öznel bilgiyi belirleyip çıkarmak için Twitter DA'nın web madenciliğini ele almışlardır. Bu çalışma ilk olarak farklı duygu kapsamları olan sözcüksel kaynakların birbirlerini nasıl tamamladığı üzerine odaklanmıştır. Bu amaçla hangi sözcüksel kaynağın diğerinden daha yararlı olduğu tanımlanmıştır. Dahası meta-seviyesi özelliklerine dayanan yeni bir duygu sınıflandırma metodu önerilmiştir. Bu denetimli yaklaşım ortaya çıkan subjektiflik sınıflandırmasını ve Twitter'daki polarite tespitini güçlendirmiştir. Çalışma sonucunda iyi sonuçlar elde edilmiştir.

Veri-Sözlüğü tabanlı yaklaşım

Görüş belirten kelimeler birçok duygu sınıflandırmasında kullanılabilir. Pozitif görüş kelimeleri arzu edilen durumları, negatif görüş kelimeleri ise arzu edilmeyen durumları ifade etmede kullanılır. Aynı zamanda görüş ifadelerinin ve deyimlerinin olduğu bir görüş veri sözlüğü vardır. Görüş kelimeleri listesini toplamak ve derlemek için üç temel yaklaşım vardır. Manuel yaklaşım çok zaman tüketir ve yalnız kullanılmaz. Final kontrolü olarak genelde diğer iki otomatik yaklaşımla birleştirilir ve bu metotlardan kaynaklı hataları önlemek için kullanılır. Yu ve Wang (2015) yaptıkları çalışmada Twitter araştırma API'sını kullanarak 2014 FİFA Dünya Kupası oyunları boyunca ABD futbol fanatiklerinin gerçek zamanlı yorumlarını toplayarak, bunların duygusal tepkilerini, gollerden sonraki duygu değişimlerini bulmak için DA yapmışlardır. ABD takımında oyunu sırasında karşı takım gol attığında korku ve sinirin genel negatif duygular olduğu ve kendi takımları gol attığında ise bu duyguların azaldığını tespit etmişlerdir. Daha sonra diğer

takımlar arasında oynanan maçlarda ABD taraftarlarının negatif duygulardan daha fazla olumlu duygu hissettikleri gözlemlenmiştir. Otomatik yaklaşımlar ilerleyen alt bölümlerde gösterilmiştir.

Sözlük tabanlı yaklaşım

Sözlük tabanlı yaklaşımın ana stratejisi [Hu ve Liu, 2004; Kim ve Hovy, 2004] tarafından sunulmuştur. Bilinen oryantasyonlarla küçük bir görüş kelime seti manuel olarak toplanır ve sonra bu küme iyi bilinen WordNet (Miller ve ark., 1990) korporası veya eş anlamlı ve zıt anlamları için araştırma (Mohammad ve ark., 1990) yapılarak geliştirilir. Yeni bulunan kelimeler tohum listesine eklenir ve bir sonraki iterasyon başlar. Yeni kelimelerin bulunması sona erene kadar iterasyon devam eder. İşlem bittiğinde manuel kontrol, hataları düzelterek ve ayıklayarak başılır.

Sözlük tabanlı yaklaşımın bir tane ciddi dezavantajı vardır. Bu da görüş kelimelerini alanları ve kontekst özellikli oryantasyonu ile bulamamasıdır. Qiu ve He (2010) sözlük tabanlı yaklaşımı bağlamsal reklamdaki duygu cümlelerini belirlemek için kullanmışlardır. Reklam ilgisini ve kullanıcı deneyimini geliştirmek için bir reklam stratejisi önermişlerdir. Sözdizimsel ayrıştırma için duygu sözlüğünü kullanmışlardır. Konu kelimelerinin yakalanması ve reklam anahtar kelimelerini çıkarmadaki müşterilerin tutumlarının belirlenmesi için bir kural tabanlı yaklaşım önermişlerdir. Automotvieforums.com sitesindeki web forumları üzerine çalışmışlardır. Önerdikleri yaklaşım reklam kelime çıkarımı ve reklam seçiminde etkili sonuçlar vermiştir.

Alvaro ve ark (2013) yaptıkları çalışmada öğrencilerin Facebook'a yazdıkları mesajların duygularını yüksek doğruluk oranı ile çıkarmanın olabilirliğini göstermişlerdir. Facebook üzerinde DA yapmak için yeni bir metot önermişlerdir. Yaklaşımları için bir Facebook uygulaması olan SentBuk'u geliştirmişlerdir. Önerdikleri metot sözlük tabanlı yaklaşım ile MÖ tekniklerinden oluşan hibrid bir metottur.

Polarite veya öznellik sınıflandırma belli bir metnin pozitif veya negatif sınıflara ayrıştırılması işlemidir. Bu konuda birçok denetimli ve denetimsiz metotlar önerilmiştir. SentiWordNet (SWN) çoğunlukla duygu analizi için bir sözcük kaynağı olarak kullanılmıştır. Khan ve ark (2016)

yaptıkları bu çalışmada SWN'i etiketlenmiş eğitim korpusu olarak bütünleştirmiş ve duygu puanları konuşma bilgisi bölümlerine dayalı olarak çıkarılmıştır. Gözden geçirilmiş duygu puanları ile bir SWN-V sözlüğü oluşturulmuş ve bu daha sonra DVM öğrenme modelinde ve sınıflandırma işleminde kullanılmıştır. Bu sözlüğe dayanarak "Enhanced Sentiment Analysis and Polarity Classification (eSAP)" çalışması önerilmiştir.

Korpus tabanlı yaklaşım

Korpus temelli yaklaşım görüş belirten kelimelerin kontekt özellikli oryantasyon ile bulunması problemini çözmeye yardım eder. Korpus tabanlı metotlar sözdizimsel örüntülere veya tohum görüş kelimeleri listeleri ile geniş bir korpusta diğer görüş kelimelerini bulmak için ortaya çıkan örüntülere dayanır. Bu metotlardan biri Hatzivassiloglou ve McKeown (1997) tarafından gösterilmiştir. Öncelikle bir görüş sıfatları tohum listesi ile başlamışlardır ve bunları bir dilbilimsel kısıtlamalar seti ile ek sıfat görüş kelimelerini ve onların oryantasyonunu belirlemek için kullanmışlardır. VE, VEYA, FAKAT, YA, YA DA gibi bağlaçlar için kısıtlamalar: bağlaç VE örneği şöyle der; sıralı sıfatlar genelde aynı oryantasyona sahiptir. Bu fikir pratikte her zaman tutarlı olmayan duygu tutarlılığı olarak adlandırılır. Fakat, bununla birlikte karşıt fikir belirten ifadeler de vardır. İki sıralanmış sıfatın aynı ya da farklı oryantasyonlara sahip olup olmadığına karar vermek için öğrenme geniş bir korpuse uygulanır. Sonra sıfatlar arasındaki bağlantılar bir grafik oluşturur ve bu grafik üzerinde kümeleme işlemi, pozitif ve negatif kelime kümeleri oluşturmak için uygulanır.

Lafferty ve ark (2001), Koşullu Rastgele Alanlar (KRA) metodunu bir dizi teknik ile birleştirerek görüş ifadelerini çıkarmak için kullanmıştır. Bu aynı zamanda çoklu-karakter dizisi örüntü eşleştirme algoritması ile duygu polaritesini ayırmak için kullanılmıştır. Algoritmaları Çin çevrimiçi yorumlarına uygulanmıştır. Birçok duygu sözlüğü oluşturmuşlardır. Araba, otel ve bilgisayar yorumları üzerine çalışmışlardır. Buldukları sonuçlar metotlarının yüksek performansla sahip olduğunu göstermiştir. Xu ve ark (2011) iki seviyeli KRA modelini düzeltilmemiş bağımlılıklar ile karşılaştırmalı ilişkileri çıkarmak

için kullanmışlardır. Bu; ilişkiler, varlıklar, kelimeler ve ilişkiler arasındaki düzeltilmemiş bağımlılıklardan faydalanılarak yapılmıştır. Amaçları müşteri yorumlarından elde ettikleri sonuçlar arasındaki karşılaştırmalı ilişkileri gösteren bir grafiksel model yapmaktır. Şirket risk yönetiminde karar desteğini göstermek için bir karşılaştırmalı ilişki haritası göstermişlerdir. Amazon.com, epinions.com, blogs, SNS ve maillerdeki mobil müşteri yorumları üzerine çalışmışlardır. Metotlarının diğerlerine göre karşılaştırmalı ilişkileri çıkarmada daha başarılı olduğu görülmüştür ve karşılaştırmalı ilişki haritalarının şirket risk yönetimi ve karar vermede çok etkili bir araç olduğu görülmüştür.

Fermin ve ark (2013) tarafından özellik taksonomisi için bir taksonomi tabanlı yaklaşım önerilmiştir. Bu taksonomi, görüş kısımlarının ve bir nesnenin öznelitliklerinin anlamsal bir gösterimidir. Verilen bir alan hakkında insanların ifade ettikleri görüşlerden değerli bir bilgiyi yakalayan alana özgü bir kaynaklar seti tanımlamışlardır. Doküman setinden otomatik olarak sonuç çıkarılmış kaynakları kullanmışlardır. Epinions.com sitesinden aldıkları kulaklıklar, oteller ve arabalar başlıkları üzerine çalışmışlardır. Yaklaşımlarını alandan bağımsız diğer tekniklerle karşılaştırmışlardır. Doğru görüş çıkarımı sistemi inşa etmek için alanın önemini ispatlamışlardır.

Hogenboom ve ark (2014) göre birçok DA metodu duygu veri sözlüklerine dayanır ve bunların içerdikleri kelimeler ve onlarla ilişkili duygular belli bir dil için uygun hale getirilmiştir. Fakat hızlı bir şekilde farklı dillerde büyüyen veri miktarı çok-dilli desteğin önemini arttırmıştır. Bu çalışmada birçok metot ilave bir dilin veri sözlüğünün DA'da desteklenmesi için incelenmiştir. Çalışmalar duygunun sadece kelime anlamlarına bağlı olmadığını ayrıca dilin belirli boyutlarına yönelimli olduğunu göstermiştir.

Sosyal medyada DA'nın büyük bir potansiyeli vardır fakat alan heterojenliğinden kaynaklı büyük zorluklarla karşılaşmaktadır. Kelimelerin duygu oriyantasyonu içerik alanı tarafından çeşitlendirilir fakat sosyal medya alanlarındaki kontekse özgü duyguyu öğrenmenin zorlukları devam etmektedir. Dilin sosyal medyada geleneksel medyaya göre oldukça farklı kullanılması da başka bir zorluk alanıdır. Bu zorluklar dikkate alınarak Deng ve

ark (2017) mevcut duygu sözlüklerini alana özgü duygu sınıflandırması için ayrıntılandırılmamış bir korpus ya da sözlük kullanarak adapte eden yeni bir metot önermişlerdir. Sonuçlar duygu sınıflandırma performansında büyük bir gelişme olduğunu göstermiştir.

Korpus tabanlı yaklaşımı tek başına kullanmak sözlük tabanlı yaklaşımı kullanmak kadar etkili değildir çünkü tüm İngilizce kelimeler için büyük bir korpus oluşturmak çok zordur fakat bu yaklaşımın önemli bir avantajı vardır bu da alan bulmaya ve kontekst belirli görüş kelimelerini bulmaya yardım eder. Korpus tabanlı yaklaşımı, istatistiksel yaklaşım veya anlamsal yaklaşımın uygulanmasında kullanılır ve bu ilerleyen bölümlerde gösterilmiştir.

İstatistiksel yaklaşım

Eşdizimlilik örüntüleri veya tohum görüş kelimelerini bulmak istatistiksel teknikler kullanılarak yapılabilir. Bu, Fahrni ve Klenner (2008) önerdiği gibi korpusdaki sıfatların eşdizimliliklerini kullanarak sonraki polariteleri üretmekle yapılabilir. Sözlük inşa etmek için web’de korpus olarak indekslenmiş tüm dokümanları kullanmak olanaklıdır. Bu kullanılan korpus yeterince geniş değilse bazı kelimelerin uygun olmama probleminin üstesinden gelir.

Geniş bir şekilde çıkarılmış bir metin korpusundaki kelimenin var olma frekansını çalışarak bir kelimenin polaritesi belirlenebilir (Read ve Carrol, 2009). Eğer kelime pozitif metinlerde daha sık var oluyorsa onun polaritesi pozitifdir. Eğer negatif metinlerde daha sık var oluyorsa o zaman polaritesi negatiftir. Eğer eşit frekansta ise o zaman nötrdür.

Benzer görüş kelimeleri bir korpusta sık sık beraber görünürler. Bu yüzden aynı kontekste iki kelime sıklıkla yan yana var oluyorsa, bunlar muhtemelen aynı polaritelere sahiptirler. Bu yüzden bilinmeyen bir kelimenin polaritesi başka bir kelime ile birlikte gerçekleşme göreceli sıklığı hesaplanarak bulunabilir. Bu NKB kullanılarak yapılabilir (Turney, 2002).

İstatistiksel metotlar DA ile ilgili birçok uygulamada kullanılmıştır. Bunlardan biri çalışma testi olarak adlandırılan bir istatistiksel rastlantısallık testi tarafından yönetilerek yorum manipülasyonlarını saptamaktır. Hu ve ark (2012) eğer yorumlar müşteriler tarafından aktüel olarak yazılmışsa müşterilerin çeşitli arka-planlarından

dolayı yorumların yazılma stillerinin rastgele olabileceğini tahmin etmişlerdir. Amazon.com sitesindeki kitap yorumları üzerine çalışmışlardır ve yaklaşık olarak ürünlerin %10.3’ün çevrimiçi yorumların manipülasyonuna maruz kaldıklarını keşfetmişlerdir.

Gizli Semantik Analiz (GSA) bir doküman seti ile bu dokümanın içinde söz edilen terimler arasındaki ilişkileri analiz eder ve bunları dokümanlar ve terimler arasında anlamlı örüntüler üretmek için kullanır (Deerwester ve ark., 1990). Cao ve Duan (2011) çeşitli özelliklerin etkilerini uygulamak için yorum metinlerinden semantik karakteristikler bularak GSA’yı kullanmışlardır. Çalışmalarının amacı; bazı yorumların neden hiç yardımsever oylar almadığı ya da az aldığı durumlarda diğerlerinin fazla oy aldığıdır. Bu yüzden hiç oy almayan yorumlar için bir yararlı öngörü yerine oy alan belli yorumların bu oyları almalarına etki eden faktörleri araştırmışlardır (evet ya da hayır oylarını içerir). CNET download.com sitesinde yazılım kullanan kullanıcıların yorumlarından faydalanmışlardır. Semantik karakteristiklerin yardımsever oy yorumları almada diğer karakteristiklere göre daha etkili olduğunu ortaya koymuşlardır.

Bir kelimenin semantik oryantasyonu da istatistiksel bir yaklaşımdır ve NKB metotları ile kullanılır. Aynı zamanda Dil Hiperuzay Analogu (DHA) adıyla bir semantik uzay uygulaması Lund ve Burgess, 1996 (Lund ve Burgess, 1996) tarafından önerilmiştir. Semantik uzay, kelimelerin noktalar tarafından gösterildiği bir uzaydır. Bu noktaların pozisyonları kelimelerin anlamları ile ilişkilidir. Xu ve ark (2012) Duygu Dil Hiperuzay Analogu adında bir yaklaşım geliştirmişlerdir. Modellerinde kelimelerin semantik oryantasyon bilgisi belirli bir vektör uzay tarafından karakterize edilir ve sonra bir sınıflandırıcı terimlerin (kelimeler ve deyimler) semantik oryantasyonunun belirlemek için eğitilir.

Hipotez NKB (SO-NKB)’deki semantik oryantasyon çıkarım metotları tarafından onaylanmıştır. Yaklaşımları çevreleyen kelimelere dayanak ağırlıklandırılmış özellikler üretir. Gazete sayfaları üzerine çalışmışlardır ve Çince bir koprus kullanmışlardır. Buldukları sonuçlar SO-NKB’den iyidir ve orijinal DHA modeli ile karşılaştırıldığında semantik oryantasyon karakteristiklerini modellemede avantaj elde etmişlerdir.

Semantik yaklaşım

Semantik yaklaşım direkt olarak duygu değerlerini verir ve kelimeler arasındaki benzerliğin farklı hesaplama prensiplerine dayanır. Bu prensipler duygu olarak yakın kelimeler için benzer duygu değerleri verir. WordNet duygu polaritelerini hesaplamak için kullanılan kelimeler arasındaki farklı özelliklerdeki duygu ilişkilerini sağlar (Kim ve Hovy, 2004).

Semantik yaklaşım fiillerin, isimlerin ve sıfatların tanımı için bir veri sözlüğü modeli oluşturmak için kullanılır ve bu Maks ve Vossen (2012) tarafından gösterilmiştir. Çalışmaları bir cümle içerisindeki aktörler arasında detaylı öznel ilişkileri tanımlamıştır. Bu öznel ilişkiler hem tutum kimliği saklayıcısı ve hem de tutumun oryantasyonu (pozitif veya negatif) ile etiketlenir. DA ile ilişkili olarak kullandıkları model bir duygu kategorizasyonu içerir. Bu; tutum saklayıcısının tanımlanması, tutumun polaritesi ve metindeki farklı aktörlerin his ve duyguların tanımlanması için çeşitli anlamlar sağlar.

Elektronik KG'lerin semantikleri için Pai ve ark (2013) tarafından eKG içerik analizi önerilmiştir. Hem olumlu hemde olumsuz değerlendirmeleri çıkarmış ve tüketicilerin karar almalarına yardımcı olmuştur. Kullandıkları metod şirketlere ürünlerini ya da hizmetleri değerlendirmelerinde yardımcı olabilir. Taywan fast food yorumları üzerine çalışmışlardır. Elde ettikleri sonuçlar eKG değerlendirmelerinin etkili olduğunu göstermiştir.

Zhang ve ark (2012) çalışmalarında gösterdiği gibi istatistiksel metodlar semantik metodlar ile harmanlanabilir. Çalışmalarında bu iki metodu, çevrimiçi yorumlardan ürünlerin zayıf yanlarını elde etmek için kullanmışlardır. Hownet tabanlı benzerlik ölçümünü sık olmayan ve aynı hedefi tanımlayan özellikleri bulmak için kullanmışlardır. Üstü kapalı özellikleri belirlemek için istatistiksel tabanlı seçim metodu olan NKB'yi kullanmışlardır. Müşteri yorumlarında muhtemelen en hoşnutsuz yön olan ürünün zayıflığını bulabilmişlerdir. Zayıflıkları bulmada çalışmaları iyi performans göstermiştir.

Veri sözlüğü tabanlı ve doğal dil işlemeli teknikler

Doğal Dil İşleme (DDİ) teknikleri sözdizimsel yapıyı ve semantik bağlantıları

bulmak için bazen veri sözlüğü yaklaşımı ile kullanılır (Bolshakov ve Gelbulch, 2004). Moreo ve ark (2012) kendi önerdikleri veri sözlüğü tabanlı DA algoritmasını kullanmadan önce ön bir işlem olarak DDİ tekniklerini kullanmışlardır. Önerdikleri sistem otomatik odaklanma tespit modülü ve DA modülünden oluşmaktadır. Çalışmaları günlük konuşma yani argonun ağırlıkta kullanıldığı konuşmalarda umut verici olmuştur.

Caro ve Grella'nın (2013) önerdiği DA yaklaşımı cümlelerin derin DDİ analizine ve bağımlılık ayrıştırma ön işlemine dayanır. Kullandıkları SA algoritması Duygu Propagandası kavramına dayanır. Metinde göze çarpan duyguları kapsamı için bir sözdizimsel tabanlı kural seti sunmuşlardır. Algoritmalarını kontekt tabanlı polarite skorlarını hesaplayarak geliştirmişlerdir. Algoritmaları 100 restoran yorumunun manuel korpusuna uygulandığında yüksek etkili sonuçlar elde edilmiştir.

Min ve Park (2012) DDİ'yi farklı bir perspektif ile kullanmışlardır. DDİ tekniklerini kipleri ve zaman ifadeleri belirlemek için madencilik teknikleri ve sıralama algoritmaları ile kullanmışlardır. Önerdikleri ölçünün iki tane parametresi vardır. Bunlardan biri ürünün kullanılması ile ilişkili ifade zamanlarının yakalanması ve diğerde ürünlerin farklı zaman aralıklarında satın alınma durumlarıdır. Çalışmalarında amazon.com isimli siteden faydalanmışlardır. Elde ettikleri sonuçlar kullandıkları ölçümün yararlı olduğunu göstermiştir.

Jurado ve Rodriguez (2015) yeni bir teknik geliştirerek bunu yazılım geliştirme sürecinde oluşan duyguyu ölçmek için kullanmışlardır. Bu durumda DDİ metodunu sunmuş ve yazılım geliştiriciler tarafından yazılan metnin altında yatan duygular analiz edilmiştir. İyi bilinen dokuz büyük projedeki 10,829 konu analiz edilmiştir. Sonuçlar geliştiricilerin metnin altında yatan duyguları terk ettiğini göstermiştir. Bu bilgi geliştirme işleminin analiz edilmesinde kullanılabilir.

Al-Rowaily ve ark (2015) yaptıkları çalışmada siber güvenlik alanında DA ve FM geliştirmek için İngilizce ve Arapça dilleri için sırasıyla SentiLen ve SentiLar duygu sözlükleri içeren Bilingual Sentiment Analysis Lexicon (BİSAL) sistemi önermişlerdir. Özellikle Arapça

için elde ettikleri duygu sözlükleri gelecek çalışmalar için umut vericidir.

Chan ve Chong (2017) yapısal olmayan büyük finansal metin de verileri üzerinde çalışma yapmışlardır. Gelişen teknoloji ile beraber finansal verilerin inanılmaz derece büyümesi ile beraber bunların incelenmesi ve anlamlı bilgiler çıkarılması zorlaşmıştır. Örneğin yatırım kararlarını vermeden önce, herhangi bir yatırımcı için ilgili finansal metin yazarlarının tercihlerini kavramak kolay olmamaktadır. Gramere dayalı dilbilimsel analizin avantajlarından yararlanarak bir duygu analizi motoru önermişlerdir.

Söylem bilgisi

DA'da söylem'in önemi gün geçtikçe artmaktadır. Söylem bilgisi hem cümlelerin arasında veya aynı cümledeki yan cümleler arasında bulunabilir. Söylem seviyesindeki duygu çıkarımı (Asher ve ark., 2008; Somasundaran ve ark., 2008)'da çalışılmıştır. Asher ve ark (2008) beş farklı tipteki retorik ilişki üzerinde çalışmışlardır: kontrast, düzeltme, destek, sonuç ve devam. Somasundaran ve ark (2008) görüş çerçevesi olarak adlandırılan bir konsept önermiştir. Görüş çerçevelerinin bileşenleri, görüşler ve bunların hedefleri ile olan ilişkileridir (Wilson ve ark., 2005). Bunlar çalışmalarını genişletmiş ve duygu sınıflandırmayı daha da geliştirmek için söylem şeması modellendirme seçeneklerini araştırmışlardır (Somasundaran ve ark., 2009).

Retorik Yapı Teorisi (RYT) (Mann ve Thompson, 1998) bir metnin aralıklara bölümünü tarif eder ve her bir aralık metnin anlamlı bir bölümünü temsil eder. Heerschop ve ark (2011) dokümanlardaki DA için bir çerçeve çalışması önermişlerdir. Bu cümle seviyesinde RYT'nin uygulanması ile elde edilen dokümandaki söylem yapısına dayanır. Şunu varsaymışlardır; duygu sınıflandırıcılarının performansı geliştirebilirler, bu da metinleri önemli ve daha az önemli metin aralıklarına bölerek yapılabilir. Film yorumlarının sınıflandırılmasında veri sözlüğü kullanmışlardır.

Zhou ve ark (2011) tarafından polarite belirsizliğini ortadan kaldırmak için kelimeler arasında söylem ilişkilerini keşfeden bir denetimsiz yaklaşım önermişlerdir. İlk olarak RYT'ye dayanan polarite üzerindeki söylem kısıtlamalarını tanımlayan bir söylem şeması belirtmişlerdir. Sonra anlamsal sıralı temsillere çevrilmiş büyük sayıda söylem örnekleri

toplamak için küçük bir işaret deyimi tabanlı örüntü setinden faydalanmışlardır. Çince eğitim verileri üzerinde çalışmışlardır. Elde ettikleri sonuçlar kullandıkları metodun tanımlanmış söylem ilişkilerini daha iyi bir şekilde bulduğunu göstermiştir.

Zirn ve ark (2011) alt cümle seviyesinde ince elenmiş DA için tam otomatik bir çerçeve önermişlerdir. Farklı duygu veri sözlüklerindeki polarite puanlarını entegre etmek için Markov matrisini kullanmışlardır ve bunu yaparken komşu segmentler arasındaki ilişkilerin bilgilerini de kullanmışlardır. Ürün yorumları üzerine çalışmışlardır. Çalışma sonucunda yapısal özelliklerin kullanımının polarite tahmin doğruluklarını %69 oranına kadar geliştirdiği ortaya çıkmıştır.

RYT'nin blog spotlarının geniş bir skaladaki polarite sıralamasında yararlı olduğu Chenlo ve ark (2017) tarafından gösterilmiştir. Cümle-seviye metodlarını anahtar cümleleri seçmek için kullanmışlardır. Daha sonra RYT analizini bu çekirdek cümlelere uygulamışlardır. Bunu da sınıflandırma polaritelerini yönetmek ve böylece belirli konuları dikkate alarak dokümanın polaritesini tahmin etmek için kullanmışlardır.

Farklı yapılarıdaki cümleler duyguları çok farklı türlerde ifade ederler. Geleneksel cümle seviyesi duygu sınıflandırması çalışmaları tüm çözümlere uygulanan bir tekniğe veya belli bir özelliğe sahip cümlelere odaklanmışlardır. Chen ve ark (2017) yaptıkları çalışmada ilk olarak cümleleri farklı tiplere göre sınıflandıran ve daha sonra cümlelerin hep bir tipe göre DA'yı uygulayan böl ve fethet yaklaşımına dayanan bir metod önermişlerdir. Yaklaşımları dört farklı sınıflandırması veri setinde değerlendirilmiş ve duygu tipi sınıflandırmasının cümle seviyesi duygu analizini geliştirebileceği ortaya çıkmıştır.

Diğer teknikler

MÖ yaklaşımı ya da veri sözlüğü tabanlı yaklaşım olarak katagorize edilemeyecek teknikler vardır. Biçimsel Kavram Analizi (BKA) bu tekniklerden birisidir. BKA, Wille (1982) tarafından önerilmiştir. Yapılanma, analiz ve görselleştirme verileri (yapılanma, analiz ve görselleştirme) için kullanılan bir matematiksel modeldir ve Galois bağlantısı (Priss, 2006) olarak bilinen dualite kavramına dayanır. Veri varlıklarının kümelerinden oluşur ve özellikleri formal kavram olarak adlandırılan formal özetler

olarak yapılandırılmıştır. Birlikte kısmi sıralı ilişkiler tarafından bir kafes kavramı oluşturulmuştur. Konsept örgüleri nesneler ve onların belirli alanlarına uygun özneliliklerin, kavramsal yapıları olarak adlandırılırlar. Nesnelerin belirlenmesi ile oluşturulur ve daha sonra aralarındaki ilişkiler gösterilir. Bulanık Biçimsel Konsept Analizi (BBKA) ise kesin ve temiz olamayan bilgileri ele almak için geliştirilmiştir. Çeşitli bilgi alan uygulamalarında başarıyla uygulanmıştır (Li ve Tsai, 2013).

BBKA ve BBKA birçok DA uygulamasında Li ve Tsai (2013) 'nin çalışmalarında gösterdiği gibi uygulanmıştır. Çalışmalarında dokümanları daha fazla kavram özetlerine göre kavramsallaştırmak için BBKA üzerine kurulu bir sınıflandırma çerçevesi önermişlerdir. Belirsiz terimlerden oluşan keyfi sonuçları geliştirmek için eğitim örnekleri kullanmışlardır. BBKA'yı, bir sınıflandırıcıyı dokümanlar yerine kavramlar kullanarak eğitmek için kullanmışlardır. Böylece sınıflandırıcının doğasında olan belirsizlikleri azaltmayı amaçlamışlardır. Benchmark test bed (Reuters 21578) ve film ve eBook yorumlarındaki iki görüş polarite veri setleri üzerine çalışmışlardır. Tüm veri setlerinde iyi sonuçlar elde etmişlerdir ve gürültüyü düşürme hassaslığı becerisinin iyi olduğunu ispatlamışlardır.

Kontopoulos ve ark (2013) BKA'yı bir ontoloji alan modeli oluşturmak için kullanmışlardır. Çalışmalarında, twitter gönderilerinin ontoloji tabanlı teknikleri kullanarak daha etkili bir DA'yı, her bir tweeti konuya uygun bir şekilde görüş kümelerine bölerek yapmışlardır. Akıllı telefon alanları üzerine çalışmışlardır. Kullandıkları mimari görüş gönderilerinin daha detaylı bir analizini vermektedir. Bunu da belli bir konuyu, konunun alan özelliklerini ve konulara verilen puanları ayırt ederek yapar.

Diğer kavram-seviyeli DA sistemleri son zamanlarda geliştirilmiştir. Mudinas ve ark (2012) bir konsept seviyeli DA sistemi olan pSenti'nin anatomisini göstermişlerdir. Bu sistem veri sözlüğü tabanlı FM ve öğrenme tabanlı yaklaşımlarla entegre edilmiştir. Sistemleri duygu polarite sınıflandırmasında yüksek doğruluk elde etmiştir. Aynı zamanda duygu gücü bulmada saf veri sözlüğü tabanlı sistemlerle karşılaştırılmıştır. İki gerçek dünya veri seti üzerine çalışmışlardır

(CNET yazılımS yorumları ve IMDB film yorumları). SentiStrenght gibi en güncel sistem benzeri hibrid yaklaşımlardan daha iyi performans göstermiştir.

Cambria ve ark (2012) SenticNet 2'yi tanıtmışlardır. Doğal kelime seviyeli dil verisi ile kavram seviyeli duygular arasındaki kavramsal ve duygusal boşluğu doldurmak için SenticNet 2'yi geliştirmişlerdir. Sistemlerini Yapay Zeka ve SemanticWeb'den yararlanan yeni bir paradigma olan sectic hesaplama ile inşa etmişlerdir. Sonuçlar sistemlerinin yapılandırılmış ve yapılandırılmamış bilgileri etkin bir biçimde birleştirmek ve karşılaştırmak için gerçek dünya uygulamalarına rahatlıkla gömülebildiğini göstermiştir.

Konsept seviye duygu analiz sistemleri e-sağlık uygulamalarında kullanılmıştır. Bu hastaların görülerinin analizini (Cambria ve ark., 2012) ve kalabalığın doğrulamasını içerir.

Zhao ve ark (2013) yaptıkları çalışmada Çince mikro-blog içeriğinde DA yapan bir metodu (SAMC) temel alan yeni bir optimizasyon algoritması önermiştir. Bu optimizasyon algoritması ile metin duygusunun hesaplanmasında etkili sonuçlar elde etmişlerdir. Bir metin DA platformu oluşturmuşlardır. Platform bir grup açık API'yi kullanıcılara sunmuştur.

İsa ve ark (2013) yaptıkları çalışmada Malayca metnlerinin köklerine ayrılması için yeni bir metot olan Reverse Porter (RP) algoritmasını önermiş ve bu Backward-Forward (B-F) adında varolan başka bir algoritma ile karşılaştırmışlardır. Çalışmada aynı zamanda Malayca metinlerinden elde edilmiş duygu değerlerine Bağışıklık Ağı Algoritması (BAA) uygulanmıştır. Çalışmanın devamında duygunun nasıl Malayca bir metinden elde edilebileceği daha iyi anlaşılmıştır. Bu çalışma biyolojiden esinlenmiş bir MÖ tekniği olan BAA'dan faydalanarak metinlerden duygu değerlerini elde etmenin yollarını bulmaya çalışmıştır.

Balahur ve Turchi (2014), metindeki duygu tespiti ve sınıflandırması ile ilgili yapılan birçok çalışmanın İngilizce dışındaki dillerde daha az yapıldığı ve buradaki eksikliklerin giderilmesi noktasından hareketle bir çalışma yapmışlardır. Bu nedenle İngilizce ile kıyaslandığında az ya da yok denecek kadar kaynakları olan dillerde duygu tespiti ve sınıflandırması yapabilen modeller geliştirmek için makine çeviri sistemleri

ve denetimli metotları uygulamışlardır. Elde ettikleri sonuçlar, makine çevirisi sistemlerinin olgun bir seviyeye ulaştığı ve uygun MÖ algoritmaları ile kombinasyonunda özellikler dikkatli seçilerek İngilizce ile rekabet edebilecek bir başarı elde edilebileceği gösterilmiştir.

Lau ve ark (2014) dizayn bilim araştırma metodolojisinin yönlendirmesi ile sosyal medyadan piyasa istihbaratı çıkarılmasında uygulamak için yeni bir sosyal analitik metodolojisi oluşturmuşlardır. Önerilen metot yeni bir yarı denetimli bulanık ontoloji madencilik algoritması ile desteklenmiştir. Gerçek sosyal medya verileri kullanılmıştır. Elde ettikleri sonuçlar, firmaların bu metodoloji ile kendi ürünlerini ve pazarlama stratejilerini geliştirebileceğidir.

Agarwal ve Mittala (2014), yaptıkları çalışmada unigrams, bigrams, çift-etiketli ve bağımlılık gibi özellikler çıkarmışlardır. Sonra belirgin özellikler, IG özellik seçim metotları kullanılarak çıkarılmıştır. Daha sonra ise semantik bilgiyi içerebilen ve MÖ algoritması için veri seyrekliği problemini azaltan yeni bir özellik seçim metodu önermişlerdir. Tüm deneysel çalışmalarda DVM ve Mantıksal Katlı Terimli Naive Bayes (MKTNB), MÖ algoritmaları sınıflandırmada kullanılmıştır. Deneysel sonuçlar önerilen özellik kümeleme metodun doküman seviyesi duygu sınıflandırmasında diğerlerinden iyi sonuçlar verdiği görülmüştür. Deneyler sırasında standart film yorum veri setleri ve kitap, elektronik, mutfak eşyaları gibi ürün yorumlarından oluşan veri setleri kullanılmıştır.

Meral ve Diri (2014) yaptıkları çalışmada algı analizi yapmak için Twitter'dan veri toplamışlardır. Bu analiz gerçekleştirilirken DDİ ve NB, Rastgele Orman ve DVM gibi MÖ yöntemleri kullanılarak, akıllı bir sistem oluşturulmuş ve elde edilen sonuçlar karşılaştırmalı olarak verilmiştir. Sınıflandırma işleminde bazı durumlarda özellik uzayının boyutunun artması sınıflandırıcının başarısını azaltıcı etki yapar. Bu nedenlerden dolayı bu çalışmada özellik seçimi yapılmıştır ve yöntem olarak korelasyon tabanlı özellik seçimi kullanılmıştır. Dokuz farklı alan ve birleşimden oluşturulan veri setleri ile sistem eğitildiğinde kelime tabanlı yöntem ile kullanıldığında %89.5'lik başarı elde edilmiştir. Kullanılan Tweeter verileri sınıflandırıcıların

eğitiminde kullanılarak %90'a yakın bir sınıflandırma başarısı elde eden bir sistem geliştirilmiştir.

Katz ve ark (2015) çalışmalarında DA yapmak için yeni bir kontekst tabanlı yaklaşım olan CoSent'i önermişlerdir. Yaklaşımlarını var olan duyguların gösterimindeki anahtar terimleri bulmak için gerekli teknikler üzerine inşa etmişlerdir. Bu terimleri ve denetimli öğrenme için üretilmiş özellikleri kullanan bir içerik modellemişlerdir. Önerdikleri modelin iki güçlü yanı vardır. Biri gürültüye karşı gürbüz olmasıdır ve diğeri ise özellik kümesine çoklu kaynaklardan kolayca özellik eklenebilmesidir. Çoklu gerçek dünya alanları üzerine yapılan deneylerde yaklaşımlarının güncel metotlar ile karşılaştırıldığında, hem gürültüsüz ve gürültülü metinlerde daha iyi sonuç verdiği görülmüştür.

Williams ve ark (2016) otomatik duygu analizinde yaklaşımlarında deyimlerin önemini ne olduğu ile ilgili çalışma yapmışlardır. Deyimlerin duygu analizine etkisini ortaya koymak için buldukları sonuçları NB, DVM gibi sınıflandırma metotları kullanarak elde ettikleri sonuçlar ile karşılaştırmışlardır. Elde ettikleri sonuçlarda hassasiyet, geri çağırma ve F-ölçümü gibi üç tane ölçütte daha iyi sonuçlar elde etmişlerdir.

Gişe tahmini zor fakat film dağıtımçıları için karar alma sürecini etkileyen önemli bir olaydır. Bu konuda birçok çalışma yapılmıştır fakat değişken çeşitliliklerinin eksikliği ve tahmin algoritmalarının basitliğinden dolayı tatmin edici sonuçlar alamamışlardır. Hur ve ark (2016) yaptıkları bu çalışmada gişe tahmin doğruluk oranının arttırmak için yorumcuların duygularından faydalanarak bir gişe tahmin modeli gösterilmiştir. Bu çalışmada gişe ve onun tahmincileri arasında non-lineer bir bağlantı kurmak için Sınıflandırma ve Regresyon Ağacı (CART), YSA ve Destek Vektör Regresyon (DVR) gibi üç tane makine öğrenme algoritması kullanılmıştır. Ayrıca makine öğrenmeye dayalı tahmin algoritmalarında değişkenlerin önemini sağlamak için bağımsız bir altuzay metodu (ISM) uygulanmıştır.

Appel ve ark (2016) tarafından cümle seviyesindeki DA problemini çözmek için hibrid bir yaklaşım önerilmiştir. Bu yeni model doğal dil işleme için gerekli olan teknikleri kullanmıştır. SentiWordNet kullanılarak bir duygu sözlüğü geliştirmiş ve semantik oryantasyon polaritesini

tahmin etmek için bulanık kümeler kullanmıştır. Tüm bunlar duygu hesaplamak için bir temel oluşturmuştur. Bu yeni yöntem üç farkı veri kümesine uygulanmış ve sonuçlar ise NB ve ME teknikleri ile karşılaştırılmıştır. Sonuçlar önerilen hibrid yöntemin daha kesin ve hassas sonuçlar verdiğini göstermiştir.

Finansal haberlerde karar desteği için kullanılan doğal dil işleme olumsuz olan tüm cümleleri doğru bir şekilde işlemek için sağlam metotlara ihtiyaç duymaktadır. Uygun negatif kapsamı tahmin etmek için genellikle kural tabanlı algoritmalar ve üretken olasılıksal modeller kullanılmıştır. Pröllochs ve ark (2016) yaptığı çalışmada bunun tersine herhangi bir uzunluktaki öğrenme görevini başarabilen güçlendirilmiş bir öğrenme metodunu önermişlerdir. Yapılan karşılaştırmalar sonucu negatif kapsam keşfinin duygulardan karar desteği ortaya çıkarmadan önemli bir aşama olduğu açığa çıkmıştır.

Schumaker ve ark (2016) maç öncesi atılan tweetlerin duygularından yola çıkarak sonuçlarının doğru tahmin edilip edilemeyeceği ile ilgili bir çalışma yapmışlardır. İngiltere premier liginde oynayan yirmi tane kulüp ile ilgili duyguları analiz etmek için ve bir bahis karar destek sistemi kurmak için Central Support adlı bir sistem kurmuşlardır. Çalışma sonucunda tweet duygusunun içinde gizlenmiş bilgilerin gücünün olduğu ve otomatik bahis sistemi tasarımı için öngörülen etkilerinin olduğu ortaya çıkmıştır.

Her gün büyük miktarda video facebook ve youtube gibi sosyal media platformlarında yayınlanmaktadır. Bu büyük miktardaki verilerden anlamlı bilgiler çıkarmak oldukça zorlu bir görevdir. Poria ve ark (2016) yaptıkları çalışmada sesli, görsel ve metinsel yapıları kullanarak videolarda multimodal DA yapan yeni bir metot önermişlerdir. Youtube veri kümesi ile daha önceden yapılan çalışma sonuçları incelendiğinde önerilen multimodal sistemin önceki çalışmaların sonuçlarından %20 oranında daha fazla başarı elde ederek %80 oranında doğruluk başarısına ulaştığı görülmüştür.

Sosyal ağın gelişmesiyle, borsadaki yatırımcılar arasındaki etkileşim daha hızlı ve kolay hale geldi. Bu yüzden yatırım kararlarını etkileyebilecek yatırımcıların düşünceleri sosyal ağda kolayca yayılarak borsayı etkileyebilir. Guo ve ark (2017) Çin borsasının sosyal medya

ağından aldıkları yorumları kullanarak DA çalışması yapmışlardır. Thermal Optimal Path (TOP) isimli bir metot önermişlerdir. Sonuçlar duygu verisinin her zaman borsa fiyatlarını belirlemediğini ve sadece borsaya yüksek oran da yatırımcı ilgisi olduğunda kullanılabileceğini göstermiştir.

Mesajların retweet edilmesi duyguların yayılmasını sağlar ve bu duygu yayılması olarak adlandırılabilir. Wang ve ark (2017) duygu yayılmasını çalışmak için duyguya dayalı bağımsız katmanlı bir model önermişlerdir. İlk olarak, kullanıcıların iletileri yeniden iletip iyileştirmede tahmin etmek için yayılma olasılığı getirilmiştir. Sonra bir öğrenme modeli, kullanıcıların iletileri yeniden yazması durumunda, retweetlerin duygularının orijinal tweet'lerin duygularından farklı olup olmadığını öngörmek için kullanılmıştır. Öğrenme modelinde kullanıcı özellikleri, yapısal özellikler ve tweet özellikleri uygulanmıştır. Daha sonra retweetlerin duyguları orijinal tweet'inkinden farklıysa, retweetlerin duygularının nasıl değiştiğini tahmin etmek için ağırlıkları değiştirilmiştir. Çalışma sonucunda elde edilen bulgular bu çalışmanın diğer modellere göre %15.78 başarı ve %4.9 performans üstünlüğü sağladığı görülmüştür.

Farklı alanların genellikle farklı duygu ifadeleri olmaktadır ve genel bir duygu sınıflandırıcısı tüm alanlar için yeterli gelmemektedir. Her bir hedef alan için o alana özgü bir duygu sınıflandırıcısı geliştirmek doğal bir çözümdür. Fakat hedef alanda etiketlenmiş veriler genellikle yetersiz, masraflı ve zaman alıcı olmaktadır. Bu problemleri aşmak için duygu bilgisini birçok kaynaktan alarak birleştiren bir alana özgü duygu sınıflandırıcı yaklaşımı önerilmiştir. Önerilen metot büyük oranda etiketli veri bağımlılığını azaltmıştır. Örneğin mutfak alanında hedef alanla ilgili sadece 200 örnek etiketli iken %87,22 oranında doğruluk oranına ulaşılmıştır. Diğer denetimli duygu sınıflandırıcılar ile kıyaslandığında Amazon ve Twitter veri setlerinde %8.98 ve %7.92 oranında performans artışı sağlanmıştır.

Tüketicilerin satın alma kararlarını desteklemek için çevrimiçi yorumlar ürünlerin sıralamasını yapmakta değerli bir araştırma konusudur ancak bu konuyla ilgili araştırma nispeten azdır. Liu ve ark (2017) yaptıkları çalışmada ürün sıralaması yapmak için DA

teknikğine ve ürünler sezgisel bulanık küme teorisine dayalı bir metot önermişlerdir. Tanımlanmış pozitif, negatif ve nötr duygu oryantasyonlarının sezgisel bulanık sayılara dönüştürülmesi büyük miktardaki çevirim içi yorumların duygu oryantasyonlarının birleştirilmesi ve işlenmesi için yeni bir fikirdir.

DUYGU ANALİZİ İLE İLİŞKİLİ ALANLAR

Duygu tespiti

DA bazen bir varlık hakkındaki görüşleri keşfetmek için bir DDİ görevi gibi düşünülebilir. Çünkü görüş, duygu ve his arasındaki farklılıklarda bazı belirsizlikler vardır. Görüş, geçişken bir kavram olarak bir varlığa karşı gösterilen tutum gibi düşünülmüştür. Duygu ise bir tutumu yansıttığı zaman hissi veya duyguyu yansıtır (Tsutsarau ve Palpanas, 2012).

Plutchik (1980) sekiz tane temel duygu olduğunu düşünmüştür. Bunlar: sevinç, üzüntü, öfke, korku, güven, iğrenme, sürpriz ve beklentidir. Duygu Tespiti (DT) bir DA olarak düşünülebilir. DA genel olarak pozitif ve negatif görüşlerin belirlenmesi ile ilgilidir fakat DT ise metindeki çeşitli duyguların algılanması ile ilgilidir. Bir görev olarak DA, DT, MÖ ve Veri sözlüğü tabanlı yaklaşımları kullanarak uygulanabilir fakat veri sözlüğü tabanlı yaklaşım daha sıklıkla kullanılır.

Cümle seciyesi DT, Lu ve (2010) tarafından önerilmiştir. İngilizce cümlelere gömülü olan bireysel olayların duygularını algılamak için web tabanlı metin madenciliği yaklaşımı önermişlerdir. Yaklaşımları bir olayın öznesi ve nesnesi arasındaki karşılıklı genel aksiyonların olasılıksal olarak dağıtılması üzerine kuruludur. Web tabanlı metin madenciliği ile semantik rol etiketlemesi tekniklerini entegre etmişlerdir. Yaklaşımları pozitif, negatif ve nötr duyguların bulunmasında iyi sonuçlar vermiştir. His belirleme probleminin aynı zamanda kontekst-hassas olduğunu ispatlamışlardır.

MÖ ve veri sözlüğü tabanlı yaklaşımın beraber kullanılması Balahur ve ark (2012) tarafından gösterilmiştir. Genel his bilgisinin depolandığı duygu korpusuna (EmotiNet) dayanan bir metot önermişlerdir. Duyguların her zaman duygusal anlamlar içeren örneğin “mutlu” gibi kelimeler tarafından ifade edilemeyebileceğini söylemişlerdir öbür yandan gerçek yaşamdaki durumların tanımlanması ile okuyucular belirli duyguları algılayabilir.

Çalışmalarında DVM ve DVM-SO algoritmaları kullanmışlardır. EmotiNet’ti temel alan metotları hiçbir duygusal kelimenin olmadığı kontekstlerde, duyguların algılanması görevinde iyi sonuçlar vermiştir. ISEAR korpusundaki bir metindeki duygunun genel his bilgisine dayananan yaklaşımlar kullanılarak en iyi şekilde tespit edildiği ispatlanmıştır. EmotiNet’i kullanarak daha iyi sonuçlar elde etmişlerdir.

Etki Analizi (EA) belli göstergesel yöntemlerle ile ortaya çıkarılan duyguların tanınmasını sağlar. Neviarouskaya ve ark (2009) bir Etki Analizi Modeli (EAM) önermişlerdir. EAM dört tane aşamadan oluşur. Bunlar; sembolik işaret, sözdizimsel yapı, kelime düzeyi ve cümle düzeyinde analizdir. Neviarouskaya’nın çalışmasında EAM birçok uygulamada kullanılması gösterilmiştir.

İnce elenmiş tutum tiplerini kullanarak cümlelerin sınıflandırılması Neviarouskaya ve ark (2010) tarafından başka bir çalışmada gösterilmiştir. Tümleme prensibine dayanan bir sistem geliştirmişlerdir. www.experienceproject.com sitesinden aldıkları 1000 cümle ile çalışmışlardır. Bu site insanların deneyimlerini, görüşlerini, duygularını, tutkularını ve itiraflarını kişisel bir hikâyeler ağı üzerinden paylaştığı bir yerdir. Metinsel tutum analizi konusunda güvenilir sonuçlar elde etmişlerdir.

Etki duygusu kelimeleri, Keshtkar ve Inkpen (2012) tarafından gösterildiği gibi bir korpus tabanlı teknik ile kullanılabilir. Çalışmalarında, kontekstsel ve sözcüksel özelliklere dayanan bir önyükleme algoritmasını tanıtmışlardır. Küçük sayıda tohum kelimeleri ile (WordNet etki duygu kelimeleri) ile başlamışlardır. Yaklaşımları altı duygu sınıfı örüntüsünü çıkarmayı öğrenir. Açıklamalı blogları ve farklı veri kümelerini yorumlar çıkarmak için kullanmışlardır. Çevrimiçi dergi blogları, metin etkisi, peri masalları ve açıklamalı bloglardan elde ettikleri veriler üzerine çalışmışlardır. Kullandıkları algoritma yararlandıkları veri kümeleri üzerinde başarılı sonuçlar vermiştir.

Ptaszynski ve ark (2013) Aozora Bunko’dan alınan anlatıların metin tabanlı EA üzerine çalışmışlardır. Anlatılardaki etkinin tanınması ile ilgili olarak kişi/karater problemleri üzerine yoğunlaşmışlardır. İlk olarak anaförük ifadelerle dayanarak duygu öznelere bir cümleden

çıkartılmış ve sonra EA prosedürü anlatının her bir bölümündeki her bir karakterin duygu durumu tahmin edilmiştir.

Maillerde ve kitaplarda EA'nın kullanılması Mohammad (2012) tarafından gösterilmiştir. Mohammad Enron e-posta korpusunu analiz etmiş ve iş yeri postalarında kullandıkları duygu kelimelerinin farklı cinsiyetler arasında farklı şekilde kullandıklarını ispatlamıştır. Kelime ilişkilerinin pozitif veya negatif polaritelerini manuel olarak açıklayan bir veri sözlüğü oluşturmuştur ve bu veri sözlüğünü sekiz temel duygu ile oluşturmuştur. Bunu duygu kelimelerini, bunların kitaplarda ve maillerde nasıl dağıldığını bulmak için kullanmıştır. Duygu kelimelerinin yoğunluğu kavramını, romanları ve peri masallarını çalışarak tanıtmıştır. Peri masallarının romanlardan daha fazla duygu kelimesi yoğunluğu dağılımına sahip olduklarını ispatlamıştır.

Kaynak inşası

Kaynak İnşası (Kİ) sözcükleri, sözlükleri ve koporayı oluşturmayı hedefler ve burada bu fikir ifadeleri polaritelerine göre belirtilmiştir. Kİ, bir DA görevi değildir fakat DA'yı ve DT'yi geliştirmeye yardım edebilir. Kelimelerin belirsizliği, çok dillilik, ayrıntıcılık ve metinsel tarzlar arasındaki ifade farkları karşılaşılan zorluklarıdır (Montoyo ve ark., 2012).

Veri sözlüğü inşası, Tan ve Wu (2011) tarafından gösterilmiştir. Çalışmalarında aynı anda eski alandan ve hedef alandan aldıkları duygu kelimelerinden ve dökümanlardan faydalanarak alan-odaklı duygu veri sözlüğü yapmak için rastgele bir yürüme algoritması önermişlerdir. Denemelerini üç adet alan özellikli duygu veri setlerinde yapmışlardır. Alan-odaklı duygu veri sözlüğünün otomatik inşasında önerdikleri algoritma iyi sonuçlar elde etmiştir.

Korpus inşası ise Robaldi ve Di Caro (2013) tarafından tanıtılmıştır. Fikir madenciliği-MÖ isminde XML tabanlı bir teknik önermişlerdir. Duygu-MÖ ve WordNet'in ötesinde yeni bir tekniktir. Çalışmaları iki bölümden oluşur. İlk kısmında metin içindeki etkili ifadelerin belirtilmesinde standart bir metodoloji sunulmuştur ve bu herhangi bir uygulama alanından kesin olarak bağımsızdır. İkinci olarak, alan-bağımlı olan ontoloji desteğine dayanan bir alan odaklı adaptasyon düşünülmüştür. Restoran yorumları üzerine çalışmışlardır. Farklı

yorumcular arasındaki anlaşmazlıkların ayrıntılı bir analizi ile birlikte önerilerini değerlendirmişlerdir. Elde ettikleri sonuçlar gayet iyidir.

Efstratios ve ark (2013) çalışmalarında tweetlerde tartışılan konuların ne olduğuna karar verecek ve her bir tweeti konularına göre özellik kümelerine ayıracak ontoloji tabanlı teknikler önermişlerdir. Sonuç her bir ayrı özelliğe bir duygu skorunun atanmasıdır. Popüler bir ürün olan akıllı telefonlar üzerine çalışmışlardır.

Shi ve ark (2013) Çin mikro bloglarındaki DA çalışmalarının bazı eksiklikleri üzerine yeni bir çalışma yapmışlardır. Bir ontoloji tabanlı model Çin mikro bloglarındaki DA için önerilmiştir. 23 Temmuz ve 1 Ağustos 2011 günleri arasında Sina micro blogging kullanıcıları tarafından '7.23 Wenzhou Tren Çarpışması' konusunda yapılan gönderilerin DA'sı yapılmıştır. Her bir günde sekiz boyutta duygu (neşe, sevgi, sürpriz, anksiyete, üzüntü, kızgınlık ve nefret) çıkarılmıştır ve sonra HowNet tabanlı bulanık duygu ontoloji kurulmuştur. Bununla birlikte yeni microblog metinlerdeki DA için yeni bir hesaplama metodu geliştirmişlerdir. Sonuçta '7.23 Wenzhou Tren Çarpışması' konusunda halkın duygusunu keşfetmişlerdir.

Boldrini ve ark (2012) geleneksel olmayan metin tarzlarındaki subjektifliğin etiketlenmesi için ayrıntılı bir şema olan EmotionB-log üzerine çalışmışlardır. Doküman, cümle ve element seviyeleri için yapılan açıklamalara odaklanmışlardır. İspanyolca, İngilizce ve italyanca dillerinde olan üç farklı konu üzerine bu dillerde gönderilen 270,000 örnekten oluşan blog gönderilerinden meydana gelmiş bir EmotiBlog korpusu oluşturmuşlardır. Modelin gürbüzlüğünü ve DDİ görevindeki uygulanabilirliğini test etmişlerdir. ISEAR gibi birçok korpora üzerinde kendi modellerini test etmiş ve başarılı sonuçlar elde etmişlerdir. Emotiblog'u duygu polarite sınıflandırmasında ve duygu tespitinde uygulamışlardır. Kendi kaynaklarının bu işlem için sistem inşası performansını geliştirdiğini ortaya koymuşlardır.

Sözlük İnşası Steinberger ve ark (2012) tarafından sunulmuştur. Çalışmalarında birçok dilde duygu sözlükleri oluşturmak için yarı-otomatik bir yaklaşım önerilmiştir. İlk olarak iki dil için yüksek seviyeli altın standard duygu sözlükleri oluşturmuşlardır ve sonra bunları otomatik olarak üçüncü bir dile çevirmişlerdir.

Bu kelimeler hedef dil kelime listesinde bulunabilir çünkü bu kelimelerin kelime duyguları iki kaynak dildekilerle benzerdir. Çalışmaları sırasında insanların açıklama ve değerlendirme çabalarındaki biçimsel kırılma ve subjektiflik gibi iki konuya dikkat çekmişlerdir. Kendi üçgensel listelerini üçgensel olmayan makine çeviri kelime listeleri ile karşılaştırmış ve kendi yaklaşımlarını onaylamışlardır.

Haeng-Jin ve ark (2013) yaptıkları çalışmada Kişisel Değer Tutum adında yeni bir model oluşturmuşlardır. Bu metot Youtube'daki bilgi teknolojisi çözümleri üzerine yapılan aktüel yorumlar kullanılarak test edilmiştir. Elde edilen sonuçlar oldukça umut verici olmuştur.

Transfer öğrenmesi

TÖ hedef alandaki öğrenme sürecini geliştirmek için yardımcı alandan bilgi çıkarır. Örneğin Wikipedia dokümanlarından bilgi transfer eder. TÖ alan farklarının çeşitli yönlerini belirten yeni bir çapraz öğrenme tekniği olarak düşünülebilir. Metin sınıflandırması (Joachims, 2002), DA (Pang ve Lee 2008), adlandırılmış varlık tanıma (Zhang ve Johnson, 2003), konuşmanın bir bölümünü etiketleme (Ratnaparthy, 1996) gibi birçok metin madenciliği görevlerini geliştirmede kullanılmıştır.

DA'da; TÖ bir alandan diğerine duygu sınıflandırması transferinde (Tang ve Wang, 2011) veya iki alan arasında bir köprü kurmada kullanılabilir (Wu ve Songbo, 2011). Tan ve Wang (2011) yüksek sıklıklı alana özgü yapıları ortaya çıkarmak için entropi tabanlı bir algoritma önermişlerdir. Ufak bir ağırlığı HFDS özelliklerine atamışlardır ve bir büyük ağırlığı aynı etikete sahip örnekler atamışlardır. Alana özgü Çince veri setlerinden gelen eğitim, stok ve bilgisayar yorumları üzerine çalışmışlardır. YSAÖ özelliklerinin olumsuz etkisini önerdikleri model aşmıştır. Aynı zamanda önerdikleri modelin DA uygulamalarında iyi bir seçenek olduğunu göstermişlerdir.

Wu ve Songbo (2011) çapraz-alan duygu sınıflandırması için iki aşamalı bir çerçeve önermişlerdir. İlk aşamada hedef alan ile kaynak alan arasında hedef alandaki bazı en güvenli etiketlenmiş dokümanları almak için bir köprü inşa edilir. Etiketlenmiş bu dokümanlar tarafından hedef alandaki verileri etiketleyerek ikinci aşamada bu esas yapıyı sömürmüşlerdir. Kitaplar, oteller ve el bilgisayarı yorumları

üzerine çalışmışlardır. Bunlarda alan odaklı Çince veri setlerinden alınmıştır. Çapraz alan duygu sınıflandırmasında etki sonuçlar elde etmişlerdir.

Stokastik Anlaşma Düzenleme (SAD) algoritması, çapraz alan polarite sınıflandırması (Ganchev ve ark., 2008) ile ilgilenir. İki farklı görüşle eğitilmiş modeller arasındaki Bhattacharyya mesafesinin minimizasyonunu temel alan bir olasılıksal çerçevedir. Her bir görüşteki modeli düzenler. SAD algoritması Lambov ve ark (2011) tarafından sunulan çalışmaya dayanak olmuştur. Bu çalışmada çapraz alan metin subjektivitesi sınıflandırması tartışılmıştır. Çoklu-görüş öğrenmesine dayalı üç yeni algoritma önermişlerdir (Wan, 2009). Üç popüler veri setinden gelen film yorumları ve soru-cevap verileri üzerine çalışmışlardır. SAD'dan daha iyi sonuçlar elde etmişlerdir.

Çeşitli veri kaynakları arasındaki farklar çoklu veri kaynaklarının ortak modellemesinde bir problem oluşturur. Ortak modelleme TÖ için önemlidir. Gupta ve ark (2012) bu problemi çözmeye çalışmışlardır. Çalışmalarında düzenlenmiş paylaşılan bir alt uzay öğrenme çerçevesi önermişlerdir. Blogspot, Flicker Youtube ve CNN, BBC haber sitelerinden aldıkları veriler üzerine çalışmışlardır. Yaptıkları çalışma, önceki çalışmalara göre daha iyi sonuçlar vermiştir.

SONUÇ

Bu çalışmada sosyal medyanın ve genel olarak internet sosyal ağlarının günümüzde giderek önem kazanması ile beraber popüler bir konu haline gelmiş DA genel olarak anlatılmış ve bu konuda yapılan çalışmalarda ortaya konulan DA algoritmaları ve onun uygulamaları gösterilmiştir. Güncel bir sonuç çıkarmak amacı ile son yıllarda yapılan çalışmalar ele alınmıştır. Daha sonra bu çalışmalar kategorize edilerek gösterilmiş ve kısaca özetlenmişlerdir. Bu çalışmalar gerçek dünya problemlerine DA tekniklerini uygulayarak DA ile ilgili birçok alana katkı sunmuştur. Tüm bu çalışmaların incelenmesinden sonra duygu sınıflandırması ve özellik seçimi algoritmalarının geliştirilmesinin halen önemli bir çalışma alanı olduğu açıkça görülmektedir. Ayrıca duygu sınıflandırma problemlerinin çözümünde Naive Bayes ve Destek Vektör Makineleri'nin en çok kullanılan

makine öğrenme algoritmaları olduğu sonucu da ortaya çıkmıştır.

Yapılan analizler sonucunda İngilizce dilinde birçok çalışma ve bunların doğal sonucu olarak birçok kaynak olduğu ortaya çıkmıştır. İngilizce dışındaki dillerde ise çalışma ve kaynak eksikliğinin olduğu görülmüştür. İngilizcenin dışında en çok kullanılan veri kaynağı WordNet olmuştur.

DA çalışmaları için son zamanlarda yeni bilgi kaynakları ortaya çıkmıştır. Bu kaynaklar mikro-blog, bloglar ve forumlar olmuştur. Tüm bu kaynaklar insanların belli güncel konular hakkında ne düşündükleri ve bu konularla ilgili duygularının anlaşılması için önemli bir rol oynamaktadır. Daha derin ve detaylı analizler için sosyal ağ siteleri ve mikro-bloglar önemli kaynaklar olmaktadır. Ayrıca algoritmaların test edilmesi için birçok veri seti de mevcuttur.

Yapılan bu çalışmada DA konusunda halen tartışmaya açık alanlar olduğu görülmüştür. Genel çalışma alanlarına ek olarak Türkçe konusunda yapılan çalışmalar halen yetersizdir. Son zamanlarda bu konuya araştırmacılar eğilmiş olsa da daha birçok çalışmanın yapılacağı öngörülmektedir. Türkçe DA, DS ve ÖS konularında yeni ve etkili çalışmalara ihtiyaç duyulmaktadır.

KAYNAKLAR

- Agarwal, A., Mittala, N.,** 2014. Semantic Feature Clustering for Sentiment Analysis of English Reviews, IETE Journal of Research. 60:6, 414-422.
- Aggarwal, C. C., Zhai, C.,** 2012. A survey of text classification algorithms. In Mining text data, Springer US. pp. 163-222.
- Aizerman, M., Braverman, E., Rozonoer, L.,** 1964. Theoretical foundations of the potential function method in pattern recognition learning. Autom Rem Cont. 821-37.
- Al-Rowaily, K., Abulaish, M., Haldar, N.A.H., Al-Rubaian, M.,** 2015. BiSAL- A bilingual sentiment analysis lexicon to analyze Dark Web forums for cyber security. Digital Investigation 14 53-62.
- Alvaro, O., José, M.M, Rosa, M.C.,** 2013, Sentiment analysis in Facebook and its application to e-learning, Computers in Human Behavior. 31;527-541.
- Appel, O., Chiclana, F., Carter, J., Fujita, H.,** 2016. A hybrid approach to the sentiment analysis problem at the sentence level. Knowledge-Based Systems, 108;110-124.
- Asher, N., Benamara, F., Mathieu, Y.,** 2008. Distilling Opinion in Discourse: A Preliminary Study. In COLING (Posters) (pp. 7-10).
- Ayoub, B., Mohamad, S., Franciska, de J.,** 2013. Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews, Knowledge-Based Systems. 52;201-213.
- Bai, X.,** 2011. Predicting consumer sentiments from online text. Decis Support Syst. 50:732-42.
- Balahur, A., Hermida J.M., Montoyo, A.,** 2012, Detecting implicit expressions of emotion in text: a comparative analysis. Decis Support Syst, 53:742-53.
- Balahur, A., Turchi, M.,** 2014. Comparative experiments using supervised learning and machine translation for multilingual sentiment analysis, Computer Speech and Language. 28, 56-75.
- Boldrini, E., Balahur, A., Martínez-Barco, P., Montoyo, A.,** 2012. Using EmotiBlog to annotate and analyse subjectivity in the new textual genres. Data Min Knowl Discov. 25: 603-34.
- Bolshakov, I.A., Gelbukh, A.,** 2004, Synonymous paraphrasing using wordnet and internet. In: International Conference on Application of Natural Language to Information Systems. Springer Berlin Heidelberg, p. 312-323.
- Bravo-Marquez, F., Mendoza, M., Poblete, B.,** 2014. Meta-level sentiment models for big social data analysis, Knowledge-Based Systems. 69, 86-99.
- Cambria, E., Benson, T., Eckl, C., Hussain, A.,** 2012. Sentic PROMs: application of sentic computing to the development of a novel unified framework for measuring health-care quality. Expert Syst Appl. 39:10533-43.
- Cambria, E., Havasi C., Hussain, A.,** 2012. SenticNet 2: A Semantic and Affective Resource for Opinion Mining and Sentiment Analysis." FLAIRS conference. p.202-207.
- Cambria, E., Hussain, A., Havasi, C., Eckl, C., Munro, J.,** 2010. Towards crowd validation of the UK national health service. WebSci10, 1-5.
- Hu, Y., Wenjie, L.,** 2011. Document sentiment classification by exploring inopinion mining and sentiment analysis. IEEE Intell Sys. 28:15-21.
- Camp, M., Bosch, A.,** 2012. The socialist network. Decis Support Syst. 53:761-9.
- Cao, Q., Duan, W., Gan, Q.,** 2011. Exploring determinants of voting for the ‘‘helpfulness’’ of online user reviews: a text mining approach. Decis Support Syst. 50:511-21.
- Chakrabarti, S., Roy, S., Soundalgekar, M.V.,** 2003. Fast and accurate text classification via multiple linear discriminant projections. VLDB J.2:172-85.

- Chan, S.W.K., Chong, M.C.V.**, 2017. Sentiment analysis in financial texts. *Decision Support Systems*. 94:53–64.
- Chen, T., Xu, R., He, Y., Wang, X.**, 2017. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. *Expert Systems With Applications*, 72; 221–230.
- Chenlo, J., Hogenboom, A., Losada, D.**, 2013. Sentiment-based ranking of blog posts using rhetorical structure theory. In: *International Conference on Application of Natural Language to Information Systems*. Springer Berlin Heidelberg, p. 13-24.
- Chien, C. C., You-De, T.**, 2011. Quality evaluation of product reviews using an information quality framework. *Decis Support Syst*. 50:755–68.
- Cortes, C., Vapnik, V.**, 1995, Support-vector networks. *Machine learning*, 20(3), 273-297.
- Cover TM, Thomas JA. 1991. *Elements of information theory*. New York: John Wiley and Sons.
- Deerwester, S., Dumais, S., Landauer, T., Furnas, G., Harshman, R.**, 1990. Indexing by latent semantic analysis. *JASIS*. 41:391–407.
- Deng, S., Sinha, A.P., Zhao, H.**, 2017. Adapting sentiment lexicons to domain-specific social media texts, *Decision Support Systems* 94;65–76.
- Di Caro, L., Grella, M.**, 2013. Sentiment analysis via dependency parsing. *Computer Standards & Interfaces*, 2013, 35.5: 442-453.
- Diana, M., Adam, F.**, 2011. Automatic detection of political opinions in tweets. In: *Proceedings of the 8th international conference on the semantic web, ESWC'11*, p. 88–99.
- Duric, A., Song, F.**, 2012. Feature selection for sentiment analysis based on content and syntax models. *Decis Support Syst*. 53:704–11.
- Efstratios, K., Christos, B., Theologos, D., Nick, B.**, 2013, Ontology-based sentiment analysis of twitter posts, *Expert Systems with Applications*. 40;4065-4047.
- Emma, H., Xiaohui, L., Yonnim, Lee., Yong, Shi.**, 2013. The Role of Text Pre-processing in Sentiment Analysis, *Procedia Computer Science*. 17; 26-32.
- Fahrni, A., Klenner, M.**, 2008. Old wine or warm beer: target-specific sentiment analysis of adjectives. In: *Proceedings of the symposium on affective language in human and machine*, AISB. p. 60–3.
- Fan, T. K., Chang, C. H.**, 2011. Blogger-centric contextual advertising. *Expert Syst Appl*. 38:1777–88.
- Fattah, M.A.**, 2015. New term weighting schemes with combination of multiple classifiers for sentiment analysis. *Neurocomputing*, 167: 434-442.
- Feldman, R.**, 2013. Techniques and applications for sentiment analysis. *Commun ACM*. 56:82–9.
- Fermin, LC., Troyano Jose, A., Enriquez F., Javier Ortega F, Vallejo, C.G.**, 2013. Long autonomy or long delay? The importance of domain in opinion mining. *Expert Syst Appl*. 40:3174–84.
- Ganchev, K., Graca, J., Blitzer, J., Taskar, B.**, 2008. Multi-view learning over structured and non-identical outputs. In: *Proceedings of the 24th conference on Uncertainty in Artificial Intelligence (UAI'08)*. p. 204–11.
- Giatsoglou, M., Vozalis, M.G., Diamantaras, K., Vakali, A., Sarigiannidis, G., Chatzisavvas, K.C.**, 2017. Sentiment analysis leveraging emotions and word embeddings. *Expert Systems With Applications*, 69 (2017) 214–224.
- Griffiths, T.L., Steyvers, M., Blei, D.M., Tenenbaum, J.B.**, 2005. Integrating topics and syntax. *Adv Neural Inform Process Syst*. 537–44.
- Guo, K., Sun, Y., Qian, X.**, 2017. Can investor sentiment be used to predict the stock price? Dynamic analysis based on China stock market. *Physica A*, 469;390–396.
- Gupta, S. K., Phung, D., Adams, B., Venkatesh, S.**, 2012 Regularized nonnegative shared subspace learning. *Data Min Knowl Discov*. 26:57–97.
- Habernal, I., Ptacek, T., Steinberger, J.**, 2014. Supervised sentiment analysis in Czech social media, *Information Processing and Management*. 50, 693-707.
- Haeng-Jin, J., Jaemoon, S., Yonnim, L., Ohbyung, K.**, 2013. Ontology Deep sentiment analysis: Mining the causality between personality-value-attitude for analyzing business ads in social media, *Expert Systems with Applications*. 40;7492-7503.
- Hagenau, M., Liebmann, M., Neumann, D.**, 2013. Automated news reading: Stock price prediction based on financial news using context-capturing features. *Decision Support Systems*, 55(3), 685-697.
- Hatzivassiloglou, V., McKeown, K.**, 1997. Predicting the semantic orientation of adjectives. In: *Proceedings of the eighth conference on European chapter of the Association for Computational Linguistics*. Association for Computational Linguistics, p. 174-181.
- He, Y., Zhou, D.**, 2011. Self-training from labeled features for sentiment analysis. *Inf Process Manage*. 47:606–16.
- Heerschop, B., Goossen, F., Hogenboom, A., Frasincar, F., Kaymak, U., de Jong, F.**, 2011. Polarity analysis of texts using discourse

- structure. In: Proceedings of the 20th ACM international conference on Information and knowledge management. ACM, p. 1061-1070.
- Hogenboom, A., Heerschop, B., Frasincar, F., Kaymak, U., & de Jong, F.,** 2014. Multi-lingual support for lexicon based sentiment analysis guided by semantics, Decision Support Systems. 62;43-53.
- Hu, M., Liu B.,** 2004. Mining and summarizing customer reviews. In: Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, p. 168-177.
- Hu, N., Bose, I., Koh, N. S., Liu, L.,** 2012. Manipulation of online reviews: an analysis of ratings, readability, and sentiments'. Decis Support Syst. 52:674-84.
- Hu, Y., Wenjie, L.,** 2011. Document duygu classification by exploring description model of topical terms. Comput Speech Lang. 25:386-403.
- Hur, M., Kang, P., Cho, S.,** 2016. Box-office forecasting based on sentiments of movies reviews and Independent subspace method. Information Sciences 372 ;608-624.
- Isa, N., Puteh, M., Kamarudin, R. M. H. R.,** 2013. Sentiment Classification of Malay Newspaper Using Immune Network (SCIN), Proceedings of the World Congress on Engineering 2013 Vol III, WCE 2013, July 3 – 5. London, U.K.
- Jebaseeli, A.N., Kirubakara, E.,** 2013. Genetic Optimized Neural Network Algorithm to Improve Classification Accuracy for Opinion Mining of M-Learning Reviews, International Journal of Emerging Trends & Technology in Computer Science (IJETTCS). 2, 345-349.
- Jiao, J., Zhou, Y.,** 2011. Sentiment Polarity Analysis based multi-dictionary. Physics Procedia, 22:590-596.
- Joachims, T.,** 2002. Learning to classify text using support vector machines: methods, theory and algorithms. MA, USA: Norwell.
- Joachims, T.,** 1996, A Probabilistic Analysis of the Rocchio Algorithm with TFIDF for Text Categorization. Carnegie-mellon univ pittsburgh pa dept of computer science.
- Jolliffe, I.T.,** 2002. Principal component analysis. Springer.
- Jurado, F., Rodriguez, P.,** 2015. Sentiment Analysis in monitoring software development processes: An exploratory case study on Git Hub's Project issues, The Journal of Systems and Software. 104, 82-89.
- Kang, H., Yoo, S. J., Han, D., Dongil, H.,** 2012. Senti-lexicon and improved Naive Bayes algorithms for sentiment analysis of restaurant reviews. Expert Syst Appl. 39:6000-10.
- Katz, G., Ofek, N., Shapira, B.,** 2015. ConSent: Context-based sentiment analysis, Knowledge-Based Systems. 84, 162-178
- Kaufmann, J.M.,** 2012. JMaxAlign: A Maximum Entropy Parallel Sentence Alignment Tool. In: Proceedings of COLING'12: Demonstration Papers, Mumbai. p. 277-88.
- Keshtkar, F., Inkpen, D.,** 2013. A bootstrapping method for extracting paraphrases of emotion expressions from texts. Computational Intelligence, 29(3), 417-435.
- Khan, F.H., Qamar, U., Bashir, S.,** 2016. SWIMS: Semi-supervised subjective feature weighting and intelligent model selection for sentiment analysis. Knowledge-Based Systems 100;97-111.
- Khan, F.H., Qamar, U., Bashir, S.,** 2016. eSAP: A decision support framework for enhanced sentiment analysis and polarity classification. Information Sciences 367-368;862-873.
- Kim, S., Hovy, E.,** 2004. Determining the sentiment of opinions. In: Proceedings of the 20th international conference on Computational Linguistics. Association for Computational Linguistics. p. 1367
- Ko, Y, Seo, J.,** 2000. Automatic text categorization by unsupervised learning. In: Proceedings of the 18th conference on Computational linguistics-Volume 1. Association for Computational Linguistics, 453-459.
- Kontopoulos, E., Berberidis, C., Dergiades, T., Bassiliades, N.,** 2013. Ontology-based sentiment analysis of twitter posts. Expert Syst Appl. 40(10), 4065-4074.
- Kranjc, J., Smailovic, J., Podpecan, V., Grcar, M., Znidaršic, M., Lavrac, N.,** 2014. Active learning for sentiment analysis on data streams: Methodology and workflow implementation in the ClowdFlows platform, Information Processing and Management. 51,187-203.
- Lafferty, J., McCallum, A., Pereira, F.,** 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: Proceedings of the eighteenth international conference on machine learning, ICML, p. 282-289.
- Lambov, D., Pais, S., Dias, G.,** 2011 Merged agreement algorithms for domain independent sentiment analysis. Procedia-Social and Behavioral Sciences, 27: 248-257.
- Lane, P.C., Clarke, D., Hender, P.,** 2012. On developing robust models for favourability analysis: Model choice, feature sets and imbalanced data. Decision Support Systems, 53(4), 712-718.
- Lau, R. Y., Li, C., Liao, S. S.,** 2014. Social analytics: Learning fuzzy product ontologies for aspect-

- oriented sentiment analysis, *Decision Support Systems*, 2014; 65:80-94.
- Lewis, D.D., Ringuette, M.**, 1994. A comparison of two learning algorithms for text categorization. In: *Third annual symposium on document analysis and information retrieval*. p. 81-93.
- Li, S., Tsai, F.**, 2011. Noise control in document classification based on fuzzy formal concept analysis. In: *Fuzzy Systems (FUZZ)*, 2011 IEEE International Conference on. IEEE, p. 2583-2588.
- Li, S.T., Tsai F.C.**, 2013. A fuzzy conceptualization model for text mining with application in opinion polarity classification. *Knowl-Based Syst.* 39: 23-33.
- Li, Y., Jain, A.**, 1998. Classification of text documents. *Comput J.* 41: 537-46.
- Li, Y.M, Li T.Y.**, 2013 Deriving market intelligence from microblogs. *Decision Support Systems*, 55(1), 206-217.
- Liu, B., Hsu, W., Ma, Y.**, 1998. Integrating classification and association rule mining. In: *Proceedings of the 4th*.
- Liu, B.**, 2012. *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.
- Liu, Y., Bi, J.W, Fan, Z.P.**, 2017. Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory. *Information Fusion* 36;149-161.
- Lu, C.Y., Lin, S.H., Liu, J.C., Cruz-Lara, S., Hong, J.S.**, 2010. Hong Jen-Shin. Automatic event-level textual emotion sensing using mutual action histogram between entities. *Expert Syst Appl.* 37:1643-53.
- Lund, K., Burgess, C.**, 1996. Producing high-dimensional semantic spaces from lexical co-occurrence. *Behav Res Methods*.28:203-8.
- Maks, I., Vossen, P.**, 2012. A lexicon model for deep sentiment analysis and opinion mining applications. *Decis Support Syst.* 53:680-8.
- Mann, W., Thompson, S.**, 1998. Rhetorical structure theory: toward a functional theory of text organization. *Text.* 8, 243-28.
- Martin-Valdivia, M.T., Martinez-Camara, E., Perea-Ortega, J.M., Urena-Lopez, A.L.**, 2013, Sentiment polarity detection in Spanish reviews combining supervised and unsupervised approaches. *Expert Syst Appl.* 40(10), 3934-3942.
- Medhat, W., Hassan, A., Korashy, H.**, 2008. Combined algorithm for data mining using association rules. *Ain Shams J Electric Eng.* 1(1).
- Medhat, W., Hassan, A., Korashy, H.**, 2014. Sentiment analysis algorithms and applications: A survey. *Ain Shams Engineering Journal*, 5:1093-1113.
- Meire, M., Ballings, M., Van den Poel, D.**, 2016. The added value of auxiliary data in sentiment analysis of Facebook posts. *Decision Support Systems.* 89:98-112.
- Meral, M., Diri, B.**, 2014. Sentiment analysis on Twitter. In: *Signal Processing and Communications Applications Conference (SIU)*, 2014 22nd. IEEE, p. 690-693.
- Miller, G., Beckwith, R., Fellbaum, C., Gross, D., Miller, K.**, 1990. *WordNet: an on-line lexical database*. Oxford Univ. Press.
- Min, H. J., Park, J. C.**, 2012. Identifying helpful reviews based on customer's mentions about experiences. *Expert Syst Appl.* 39:11830-8.
- Mohammad, S., Dunne, C., Dorr, B.**, 2009. Generating high-coverage semantic orientation lexicons from overtly marked words and a thesaurus. In: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 2-Volume 2*. Association for Computational Linguistics, p. 599-608.
- Mohammad, S.M.**, 2012. From once upon a time to happily ever after: tracking emotions in mail and books. *Decis Support Syst.* 53:730-41.
- Montoyo, A., MartiNez-Barco, P., Balahur, A.**, 2012. Subjectivity and sentiment analysis: an overview of the current state of the area and envisaged developments. *Decis Support Syst.* 53:675-9.
- Moraes, R., Valiati J.F., Gaviao Neto, W.P.**, 2013. Document-level sentiment classification: an empirical comparison between SVM and ANN. *Expert Syst Appl.* 40: 621-33.
- Moreo, A., Romero, M., Castro, J.L., Zurita, J.M.**, 2012. Lexicon-based comments-oriented news sentiment analyzer system. *Expert Syst Appl.* 39:9166-80.
- Mudinas, A., Zhang, D., Levene, M.**, 2012. Combining lexicon and learning based approaches for concept-level sentiment analysis. In: *Proceedings of the First International Workshop on Issues of Sentiment Discovery and Opinion Mining*. ACM, p. 5.
- Neppalli, V.K., Caragea, C., Squicciarini, A., Tapia, A., Stehle, S.**, 2017. Sentiment analysis during Hurricane Sandy in emergency response. *International Journal of Disaster Risk Reduction* 21;213-222.
- Neviarouskaya, A., Prendinger, H., Ishizuka, M.**, 2010. Recognition of Affect, Judgment, and Appreciation in Text. In: *Proceedings of the 23rd international conference on computational linguistics*, Beijing. p. 806-14.
- Neviarouskaya, A., Prendinger, H., Ishizuka, M.**, 2007. Recognition of affect conveyed by text messaging in online communication. *Online Communities and Social Computing*, 141-150.

- Neviarouskaya, A., Prendinger, H., Ishizuka, M.,** 2009, Compositionality Principle in Recognition of Fine-Grained Emotions from Text. In: ICWSM.
- Neviarouskaya, A., Tsetserukou, D., Prendinger, H., Kawakami, N., Tachi, S., Ishizuka, M.,** 2009. Emerging system for affectively charged interpersonal communication. In: ICCAS-SICE, 2009. IEEE, p. 3376-3381.
- Ng, H.T., Goh, Wei, Low, Kok.,** 1997. Feature selection, perceptron learning, and a usability case study for text categorization. In: ACM SIGIR Forum. ACM, p. 67-73.
- Nizam, H., Akin, S.S.,** 2014. Sosyal Medyada Makine Öğrenmesi ile Duygu Analizinde Dengeli ve Dengesiz Veri Setlerinin Performanslarının Karşılaştırılması, XIX. Türkiye'de İnternet Konferansı.
- Ortigosa-Hernández, J., Rodríguez, J. D., Alzate, L., Lucania, M., Inza, I., Lozano, J.A.,** 2012. Approaching sentiment analysis by using semi-supervised learning of multi-dimensional classifiers. *Neurocomputing*. 92:98–115.
- Pai, M.Y., Chu H.C., Wang, S.C., Chen, Y.M.,** 2013. Electronic word of mouth analysis for service experience. *Expert Syst Appl*. 40:1993–2006.
- Pang, B., Lee, L.,** 2008. Opinion mining and sentiment analysis. *FoundTrends Inform Retrieval*. 2:1–135.
- Pirayani, R., Madhavi, D., Singh, V.K.,** 2017. Analytical mapping of opinion mining and sentiment analysis research during 2000-2015. *Information Processing and Management*, 53;122–150.
- Plutchik, R.,** 1980. A general psychoevolutionary theory of emotion. *Emotion: Theory Res Exp*. 1:3–33.
- Poria, S., Cambria, E., Howard, N., Huang, G.B., Hussain, A.,** 2016. Fusing audio, visual and textual clues for sentiment analysis from multimodal content. *Neurocomputing*. 174. 50–59.
- Priss, U.,** 2006. Formal concept analysis in information science. *Arist*, 40.1: 521-543.
- Pröllochs, N., Feuerriegel, S., Neumann, D.,** 2016. Negation scope detection in sentiment analysis: Decision support for news-driven trading. *Decision Support Systems*, 88;67–75.
- Ptaszynski, M., Dokoshi H., Oyama, S., Rzepka, R., Kurihara, M., Araki, K., Momouchi, Y.,** 2013. Affect analysis in context of characters in narratives. *Expert Syst Appl*. 40:168–76.
- Qiu, G., He, X., Zhang, F., Shi, Y., Bu, J., & Chen, C.,** 2010. DASA: dissatisfaction-oriented advertising based on sentiment analysis. *Expert Syst Appl*. 37:6182–91.
- Quinlan, J.R.,** 1986. Induction of decision trees. *Machine Learn*. 1:81–106.
- Ratnaparkhi, A.,** 1996. A maximum entropy model for part-of-speech tagging. In: *Proceedings of the conference on empirical methods in natural language processing*. p. 133-142.
- Read, J., Carroll, J.,** 2009 Weakly supervised techniques for domain-independent sentiment classification. In: *Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion*. ACM, p. 45-52.
- Reyes, A., Rosso, P.,** 2012. Making objective decisions from subjective data: detecting irony in customer reviews. *Decis Support Syst*. 53:754–60.
- Robaldo, L., and Di Caro, L.,** 2013. Opinionmining-ml. *Computer Standards & Interfaces*, 35.5: 454-469.
- Rui, H., Liu, Y., Whinston, A.,** 2013. Whose and what chatter matters? The effect of tweets on movie sales. *Decis Support Syst*. 55(4), 863-870.
- Ruiz, M., Srinivasan, P.,** 1999. Hierarchical neural networks for text categorization (poster abstract). In: *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, p. 281-282.
- Schumaker, R.P., Jarmoszko, A.T., Labedz, C.S.,** 2016. Predicting wins and spread in the Premier League using a sentiment analysis of twitter. *Decision Support Systems* 88;76–84.
- Shi, W., Wang, H., He, S.,** 2013. Sentiment analysis of Chinese microblogging based on sentiment ontology: a case study of '7.23 Wenzhou Train Collision'. *Computer Science*. 25: 4, 161-178.
- Smailovic, J., Grcar, M., Lavrac, N., Znidaršic, M.,** 2014. Stream-based active learning for sentiment analysis in the financial domain, *Information Sciences*. 285, 181-203.
- Somasundaran, S., Namata, G., Wiebe, J., Getoor, L.,** 2009. Supervised and unsupervised methods in employing discourse relations for improving opinion polarity classification. In: *Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing: Volume 1-Volume 1*. Association for Computational Linguistics, p. 170-179.
- Somasundaran, S., Wiebe, J., Ruppenhofer, J.,** 2008. Discourse level opinion interpretation. In: *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*. Association for Computational Linguistics, p.801-808.
- Somasundaran, S., Wiebei, J.,** 2009 Recognizing stances in online debates. In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of*

- the AFNLP: Volume 1-Volume 1. Association for Computational Linguistics, p. 226-234.
- Steinberger, J., Ebrahim, M., Ehrmann, M., Hurriyetoglu, A., Kabadjov, M., Lenkova, P., Zavarella, V.,** 2012. Creating sentiment dictionaries via triangulation. *Decis Support Syst.* 53: 689–94.
- Sun, X., Li, C., Ren, F.,** 2016. Sentiment analysis for Chinese microblog based on deep neural networks with convolutional extension features. *Neurocomputing.* 210:227–236.
- Tan, S., Wang, Y.,** 2011. Weighted SCL model for adaptation of sentiment classification. *Expert Syst Appl.* 38: 10524–31.
- Tan, S., Wu, Q.,** 2011. A random walk algorithm for automatic construction of oriented sentiment lexicon. *Expert Syst Appl.* 12094–100.
- Tsytsarau, M., Palpanas, T.,** 2012. Survey on mining subjectivedata on the web. *Data Min Knowledge Discovery.* 24: 478–514.
- Turney, P.,** 2002. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. In: *Proceedings of the 40th annual meeting on association for computational linguistics.* Association for Computational Linguistics, p. 417-424.
- Van de Kauter, M., Breesch, D., Hoste, V.,** 2015. Fine-grained analysis of explicit and implicit sentiment in financial news articles, *Expert Systems with Applications*, 42, 4999-5010.
- Vapnik, V.,** 1995, *The nature of statistical learning theory*, New York.
- Walker, M.A, Anand, P., Abbott, R., Fox Tree, J.E., Craig, M., King, J.,** 2012. That is your evidence?: Classifying stance in online political debate. *Decis Support Syst.* 53: 719–29.
- Wan, X.,** 2009. Co-training for cross-lingual sentiment classification. In: *Proceedings of the Joint Conference of the 47th Annual Meeting of the ACL and the 4th International Joint Conference on Natural Language Processing of the AFNLP: Volume 1-volume 1.* Association for Computational Linguistics, p. 235-243.
- Wang, Q., Jin, Y., Yang, T., Cheng, S.,** 2017. An emotion-based independent cascade model for sentiment spreading. *Knowledge-Based Systems*, 116:86–93.
- Whitelaw, C., Garg, N., Argamon, S.,** 2005. Using appraisal groups for sentiment analysis. In: *Proceedings of the ACM SIGIR Conference on Information and Knowledge Management (CIKM).* p. 625–31.
- Wille, R.,** 1982. Restructuring lattice theory: an approach based on hierarchies of concepts. In: *I. Rival, Reidel, Dordrecht-Boston.* p. 445–70.
- Williams, L., Bannister, C., Arribas-Ayllon, M., Preece, A., Spasic, I.,** 2015. The role of idioms in sentiment analysis. *Expert Systems with Applications* 42:7375–7385.
- Wilson, T., Wiebe, J., Hoffman, P.,** 2005. Recognizing contextual polarity in phrase-level sentiment analysis. In: *Proceedings of the conference on human language technology and empirical methods in natural language processing.* Association for Computational Linguistics, p. 347-354
- Wu, F., Huang, Y., Yuan, Z.,** 2017. Domain-specific sentiment classification via fusing sentiment knowledge from multiple sources. *Information Fusion* 35:26–37.
- Wu, Q., Songbo, T.,** 2011. A two-stage framework for crossalan sentiment classification. *Expert Syst Appl.* 38:14269–75.
- Xianghua, F., Guo., L., Yanyan, G., Zhiqiang, W.,** 2013. Multiaspect sentiment analysis for Chinese online social reviews based on topic modeling and HowNet lexicon. *Knowl-Based Syst.* 37:186–95.
- Xu, K., Liao, S. S., Li, J., Song, Y.,** 2011. Mining comparative opinions from customer reviews for competitive intelligence. *Decis Support Syst.* 50: 743–54.
- Xu, T., Peng, Q., Cheng, Y.,** 2012. Identifying the semanticorientation of terms using S-HAL for sentiment analysis. *Knowl-Based Syst.* 35:279–89.
- Yan-Yan, Z., Bing, Q., Ting, L.,** 2010. Integrating intra-and inter-document evidences for improving sentence sentiment classification. *Acta Automatica Sinica*, 36.10: 1417-1425.
- Yelena, M., Padmini, S.,** 2011. Exploring Feature Definition and Selection for Sentiment Classifiers. In: *ICWSM.*
- Yu, L.C., Wu, J.L., Chann, P.C., Chu, H.S.,** 2013. Using a contextual entropy model to expand emotionwords and their intensity for the sentiment classification of stockmarket news. *Knowl-Based Syst.* 41:89–97.
- Yu, Y., Wang. X.,** 2015. World Cup 2014 in the Twitter World: A big data analysis of sentiments in U.S. sports fans' tweets, *Computers in Human Behavior.* 48, 392-400.
- Zhang, T., Johnson, D.,** 2003. A robust risk minimization based named entity recognition system. In: *Proceedings of the seventh conference on Natural language learning at HLT-NAACL 2003-Volume 4.* Association for Computational Linguistics, p. 204-207.
- Zhang, W., Xu, H., Wan, W.,** 2012. Weakness finder: find product weakness from Chinese reviews by using aspects based sentiment analysis. *Expert Syst Appl.* 39:10283–91.
- Zhao, Y., Niu, K., He, Z., Lin, J., Wang, X.,** 2013. Text sentiment analysis algorithm optimization

and platform development in social network. In: Computational Intelligence and Design (ISCID), 2013 Sixth International Symposium on. IEEE, p. 410-413.

Zhou, L., Li, B., Gao, W., Wei, Z., Wong, K., 2011. Unsupervised discovery of discourse relations for eliminating intra-sentence polarity ambiguities. In: Proceedings of the Conference on Empirical Methods in Natural Language Processing. Association for Computational Linguistics, p. 162-171.

Zirn, C., Niepert, M., Stuckenschmidt, H., Strube, M., 2011. Fine-Grained Sentiment Analysis with Structural Features. In: IJCNLP. p. 336-344.