



Evrişimli Sinir Ağı Mimarilerine Dayalı Türkçe Duygu Analizi

Aytuğ Onan^{1*}

¹ İzmir Katip Çelebi Üniversitesi, Mühendislik ve Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye (ORCID: 0000-0002-9434-5880)

(Bu yayın 26-27 Haziran 2020 tarihinde HORA-2020 kongresinde sözlü olarak sunulmuştur.)

(DOI: 10.31590/ejosat.780609)

ATIF/REFERENCE: Onan A. (2020). Evrişimli sinir ağı mimarilerine dayalı Türkçe duygu analizi. *Avrupa Bilim ve Teknoloji Dergisi*, (Special Issue), 374-380.

Öz

Duygu analizi (görüş madenciliği), metin belgeleri içerisinde yer alan nesnelere, ürünlere, servislere ya da organizasyonlara ilişkin görüş, duygu, tutum gibi öznel bilgilerin, makine öğrenmesi, istatistik ve doğal dil işleme gibi alanlardan teknik ve yöntemlerin kullanılması ile çıkarılmasını amaçlayan bir araştırma alanıdır. Duygu analizi, yapısal olmayan bilgidir, yapısal, anlamlı ve kullanışlı bilgiler çıkarılmasını olanaklı hale getirir. Bu bilgi, karar destek sistemleri ve bireysel karar vericiler için önemli bir kaynak olarak işlev görür. Evrişimli sinir ağları, veriyi ızgara benzeri bir topoloji ile işleyen bir tür derin öğrenme yöntemidir. Bu çalışmada, Türkçe metin belgeleri üzerinde duygu analizi için, evrişimli sinir ağı tabanlı üç temel derin öğrenme mimarisinin etkinliği değerlendirilmektedir. Çalışma kapsamında önerilen birinci mimaride, gömme katmanında, metin belgesinde yer alan kelimeler için, kelime gömme yöntemleri tabanlı temsil elde edilmektedir. Ardından, evrişim katmanları yığını (1-gram, 2-gram ve 3-gram) kullanılarak 1-gram, 2-gram ve 3-gram tabanlı özniteliklerin çıkarımı gerçekleştirilmektedir. Her bir katmanda, öznitelik haritalarının oluşturulması için sabit sayıda 80 filtre uygulanmaktadır. İncelenen ikinci evrişimli sinir ağı tabanlı mimaride gömme katmanı sonucu elde edilen metin temsili, yinelenen evrişim ve maksimum havuzlama katmanlarına tabi tutulmaktadır. İncelenen üçüncü mimari ise, evrişim tabanlı piramit mimarisidir. Metin belgelerinin temsiliinde, word2vec, fastText, GloVe ve LDA2vec olmak üzere dört temel kelime gömme yöntemi incelenmektedir. İncelenen evrişimli sinir ağı tabanlı mimarilerin, Türkçe duygu analizi için, geleneksel makine öğrenmesi sınıflandırıcılarına (k-en yakın komşu algoritması, destek vektör makineleri, lojistik regresyon ve Naive Bayes algoritması) ve temel derin öğrenme mimarilerine (tekrarlayan sinir ağları, uzun kısa süreli bellek birimleri ve geçitli tekrarlayan birim) kıyasla daha yüksek başarımlar elde ettiği görülmektedir. Word2vec (Skip-gram modeli) kelime kodlaması yöntemi ile evrişimli sinir ağı tabanlı mimari ile %92.53 doğru sınıflandırma oranı elde edilmiştir.

Anahtar Kelimeler: Duygu Analizi, Derin Öğrenme, Evrişimli Sinir Ağları.

Sentiment Analysis in Turkish Based on Convolutional Neural Network Architectures

Abstract

Sentiment analysis is a research field that aims to identify the sentiment orientation (as, positive, negative or neutral) of a particular topic in text documents through machine learning, statistics and natural language processing techniques. Convolutional neural networks are a type of deep learning methods, which process data with a grid-like topology. In this paper, we present empirical results for three deep learning architectures based on convolutional neural network for sentiment analysis on Turkish. The first architecture initially employs word-embedding schemes to represent text documents. Then, a stack of convolution layers (i.e., 1-gram, 2-gram and 3-gram) has been employed to extract 1-gram, 2-gram and 3-gram based features. For each layer, constant number of filters have been employed to construct feature maps. The second examined architecture employs recurrent convolution and maximum pooling schemes on word embedding based representation. The third architecture is a convolution based pyramid configuration. To represent text corpus, word2vec, fastText, GloVe and LDA2vec word embedding schemes have been utilized. The empirical results on Turkish sentiment classification indicate that convolutional deep learning based architectures outperform conventional machine learning

* Sorumlu Yazar: İzmir Kâtip Çelebi Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü, İzmir, Türkiye, ORCID: 0000-0002-9434-5880, aytug.onan@ikcu.edu.tr

methods (such as, k-nearest neighbor algorithm, support vector machines, logistic regression and Naïve Bayes algorithm) and conventional deep learning architectures (such as, recurrent neural networks, long short term memory architecture and gated recurrent unit). We obtained a classification accuracy of 92.53% with convolutional neural network based architecture in conjunction with word2vec (skip-gram model) based word embedding scheme.

Keywords: Sentiment Analysis, Deep Learning, Convolutional Neural Networks.

1. Giriş

Duygu analizi (görüş madenciliği), metin belgeleri içerisinde yer alan nesnelere, ürünlere, servislere ya da organizasyonlara ilişkin görüş, duygu, tutum gibi öznel bilgilerin, makine öğrenmesi, istatistik ve doğal dil işleme gibi alanlardan teknik ve yöntemlerin kullanılması ile çıkarılmasını amaçlayan bir araştırma alanıdır [1]. Günümüzde, Web, yapısal olmayan çok sayıda bilgi içeren önemli bir kaynak durumundadır. Elektronik ticaret siteleri, değerlendirme siteleri, tartışma forumları gibi birçok farklı platformda görüş içeren önemli ölçüde bilgi yer almaktadır. Duygu analizi, yapısal olmayan bilgidir, yapısal, anlamlı ve kullanışlı bilgiler çıkarılmasını olanaklı hale getirir. Bu bilgi, karar destek sistemleri ve bireysel karar vericiler için önemli bir kaynak olarak işlev görür [1].

Duygu analizi yöntemleri, temel olarak, sözlüğe dayalı yöntemler ve makine öğrenmesine dayalı yöntemler olmak üzere iki temel sınıf altında ele alınmaktadır [2]. Sözlüğe dayalı duygu analizi yöntemleri, duygu analizi gerçekleştirme işlemini, metinde geçen sözcük ve tümcelerin anlamsal yönelimlerini, sözlüğe dayalı olarak hesaplayarak belirler [3]. Makine öğrenmesi tabanlı duygu analizi yöntemlerinde ise, görüş kutbu etiketli veri üzerinde Naive Bayes, destek vektör makineleri, k-en yakın komşu algoritması gibi sınıflandırıcılar eğitilerek, görüş sınıflandırma modeli oluşturulmaktadır [4].

Literatürde son yıllarda gerçekleştirilen çalışmalara göre, duygu analizi, dil modelleme, makine çevirisi gibi doğal dil işleme ve metin madenciliği problemlerinde, metin belgelerinde yer alan tümcelerin derin öğrenme yöntemleri kullanılarak işlenmesi başarımları önemli ölçüde artırmaktadır [5]. Derin öğrenme tabanlı metin madenciliği uygulamalarında, özyinelemeli sinir ağları, geçitli yineleyen birim ve uzun kısa süreli bellek mimarileri başarıyla uygulanmaktadır [6]. Evrişimli sinir ağları, veriyi ızgara benzeri bir topoloji ile işleyen bir tür derin öğrenme yöntemidir. İnsan görme sisteminin modellenmesine dayalı olan evrişimli sinir ağları, başta görüntü sınıflandırma, nesne tanıma, ses işleme, parmak izi tanıma ve doğal dil işleme gibi problemler olmak üzere, birçok alanda başarıyla uygulanmaktadır [7]. Özellikle büyük hacimde verinin varlığında, evrişimli sinir ağları, yüksek başarımlar gösteren ve geliştirme yeteneği yüksek sınıflandırma modelleri oluşturmaktadır.

Metin madenciliği alanında, özyinelemeli sinir ağları, geçitli yineleyen birim ve uzun kısa süreli bellek gibi mimarilerin başarımları sıklıkla değerlendirilmekle birlikte, evrişimli sinir ağı tabanlı mimari kullanan çalışma sayısı kısmen sınırlı kalmaktadır [8]. Bu çalışmada, Türkçe metin belgeleri üzerinde duygu analizi için, evrişimli sinir ağı tabanlı üç temel derin öğrenme mimarisinin etkinliği değerlendirilmektedir. Metin madenciliği uygulamalarındaki en temel aşama, metin belgelerinin uygun bir veri temsili ile temsil edilmesidir. Geleneksel makine öğrenmesine dayalı duygu analizi yaklaşımlarında, metin belgesinde yer alan sözcükler, sözcük torbası yöntemi ile temsil edilmektedir. Derin öğrenme tabanlı duygu analizi çalışmalarında ise sözcüklerin sabit uzunluklu vektörler kullanılarak yoğun uzaylarda temsil edilmesi söz konusudur. Bu sabit uzunluklu vektörler, ya eğitim esnasında ya da eğitim öncesinde word2vec, fastText ve GloVe gibi yöntemler kullanılarak elde edilmektedir [9]. Çalışma kapsamında, metin belgelerinin temsilinde, word2vec, fastText, GloVe ve LDA2vec olmak üzere dört temel kelime gömme yönteminin etkinliği değerlendirilmektedir. Dört kelime gömme yöntemi ile elde edilen veri temsilleri, geliştirilen üç evrişimli sinir ağı tabanlı mimari üzerinde incelenmektedir. İncelenen evrişimli sinir ağı tabanlı mimarilerin, Türkçe duygu analizi için, geleneksel makine öğrenmesi sınıflandırıcılarına ve derin öğrenme mimarilerine kıyasla daha yüksek başarımlar elde ettiği görülmüştür.

Çalışmanın geri kalanı şu şekildedir: Bölüm II’de ilgili çalışmalar, Bölüm III’te metodoloji, Bölüm IV’te ise deneysel süreç, deneysel sonuçlar ve tartışma sunulmaktadır.

2. İlgili Çalışmalar

Collobert ve Weston [10] tarafından gerçekleştirilen çalışma, metin madenciliği alanında evrişimli sinir ağlarını kullanan temel çalışmalardan biridir. Bu çalışmada, evrişimli sinir ağı tabanlı mimari kullanılarak metin belgelerinde, varlık ismi tanıma ve sözcük türü belirleme gibi işlemler gerçekleştirilmiştir. Benzer şekilde, Collobert vd. [5] tarafından gerçekleştirilen çalışmada, sözcük türü belirleme, bölümlendirme, varlık ismi belirleme, anlamsal rol etiketleme gibi temel doğal dil işleme görevleri için, evrişimli sinir ağı tabanlı bir mimari geliştirilmiştir. dos Santos ve Gatti [11] tarafından gerçekleştirilen çalışmada, metin belgesinde yer alan her bir karakter, sabit uzunluklu bir karakter vektörü şeklinde temsil edilmektedir. İlgili karakter tabanlı gömme, sözcükler içerisinde yer alan morfolojik örüntülerin çıkarılması için kullanılmıştır. Bunun yanı sıra, word2vec kelime gömmesi kullanılarak özellikler elde edilmiştir. Elde edilen veri temsili, evrişim katmanı ve maksimum havuzlama içeren bir mimari kullanılarak duygu analizinde uygulanmıştır. Kim [12] tarafından gerçekleştirilen çalışmada, temel evrişimli sinir ağı tabanlı mimari, tümce seviyesi duygu analizi için kullanılmıştır. Zhang vd. [13] tarafından gerçekleştirilen çalışmada, 9-29 katmanlı evrişimli sinir ağları kullanılarak metin belgelerinden temel karakter tabanlı öznitelikler çıkarılmıştır. Geliştirilen derin öğrenme tabanlı mimari, çok sayıda veri içeren büyük veri setlerinde yüksek başarımlar elde ederken, küçük veri setlerinde, geleneksel yapay sinir ağlarından daha düşük başarımlar elde edildiği gözlemlenmiştir. Johnson ve Zhang [14] tarafından gerçekleştirilen çalışmada ise metin sınıflandırma için sözcük seviyesi evrişimli sinir ağı mimarisini geliştirilmiştir. Bu çalışmada, metin sınıflandırma için sözcük seviyesi evrişimli sinir ağları ile karakter seviyesi sinir ağlarının başarımları karşılaştırılmış; sözcük seviyesi evrişimli sinir ağı mimarileriyle daha yüksek başarımlar elde edildiği gözlemlenmiştir. Çano ve Morisio [15] tarafından gerçekleştirilen çalışmada, şarkı sözü duygu belirleme, tümce seviyesi duygu analizi, film, telefon ve otel değerlendirmeleri gibi farklı problemler üzerinde önceden eğitilmiş kelime gömmelerine tabi tutulan metin belgeleri, evrişimli sinir ağı mimarileriyle bir arada değerlendirilmiştir.

3. Metodoloji

Bu bölümde çalışmada kullanılan kelime gömme yöntemleri, temel derin öğrenme mimarileri ve önerilen evrişimli sinir ağı tabanlı mimariler tanıtılmaktadır.

3.1. Kelime Gömme Yöntemleri

Doğal dil işleme problemlerinde sıklıkla kullanılan kelime torbası yöntemi, iki temel probleme sahiptir. Kelime torbası tabanlı metin temsilinde kelimelerin sırası göz ardı edilir. Bunun yanı sıra, vektör uzunluğu tüm metin derleminde yer alan sözcüklerin boyutu kadardır. Metin belgelerinde yer alan sözcüklerin, sabit uzunluklu vektörler biçiminde yoğun uzaylarda temsil edilmesini olanaklı kılan kelime gömme yöntemleri, verinin daha az boyutla temsil edilmesini sağlamakta ve derin öğrenme tabanlı metin madenciliği uygulamalarında sıklıkla kullanılmaktadır.

3.1.1. Word2vec modeli

Word2vec modeli, girdi katmanı, çıktı katmanı ve gizli katmandan oluşan yapay sinir ağı tabanlı bir kelime gömme yöntemidir [16]. Belirli bir kelimenin, diğer kelimeler ile çevrili olma olasılığını belirleyerek kelime gömmelerini öğrenmeyi hedefler. Model, skip-gram (SG) ve continuous-bag-of words (CBOW) isimli iki temel mimariye sahiptir. CBOW mimarisi, her kelimenin içeriğini girdi olarak alarak hedef sözcüğü tanımlar; SG mimarisi ise, hedef sözcüğü girdi olarak alarak hedef kelimeyi çevreleyen kelimeleri tahmin eder. CBOW mimarisi, az miktarda veriyle düzgün bir biçimde çalışabilir. SG mimarisi ise büyük veri setlerinde daha iyi çalışır.

3.1.2. GloVe modeli

GloVe modeli, metin belgelerinden kelime gömmelerini etkin bir şekilde öğrenmek için geliştirilmiş word2vec tabanlı bir modeldir. Model, word2vec modelinin yerel bağlama dayalı öğrenmesini, küresel matris çarpanlarına ayırma ile birleştirir [17]. Modelde, hata fonksiyonun hesaplanmasında kelimelerin olasılık oranları da dikkate alınır. Metin belgesi içerisinde birbirlerine yakın gözlemlenen ve birlikte görülmeye olasılıkları yüksek sözcükler, öğrenme sürecinde diğer kelimelere kıyasla daha önemlidir [16].

3.1.3. FastText modeli

FastText modeli, metin belgelerinden kelime gömmelerini elde etmek için kullanılan bir başka etkili yöntemdir. Bu modelde, her kelime, karakter n-gram şeklinde bölerek temsil edilmektedir. Eğitim setinde yer alan her bir n-gram için kelime vektörleri oluşturulur. fastText modeli, morfolojik olarak zengin diller ve nadir görülen kelimeler için daha verimli bir kelime gömme şeması sağlar [6].

3.1.4. LDA2vec modeli

LDA2vec modeli, word2vec modeline ve gizli Dirichlet ayırma yöntemine dayanan bir başka kelime gömme yöntemidir. Bu yöntemde, gizli belge düzeyi karışım vektörlerinden yoğun sözcük vektörleri, Dirichlet dağılımına bağlı olarak birlikte elde edilmektedir. Model, metin koleksiyonlarının konularını ve konuya göre ayarlanmış sözcük vektörlerini tanımlamayı sağlar. Bu bağlamda, model her bir kelimeyi bir konuya bağlayarak konuyla zenginleştirilmiş bir kelime gömme şeması sunar. Modelde, skip-gram negatif örnekleme objektif fonksiyon olarak kullanılmaktadır [6].

3.2. Temel Derin Öğrenme Mimarileri

3.2.1. Tekrarlayan Sinir Ağı Mimarisi

Tekrarlayan sinir ağı (RNN), sıralı verileri işlemek için kullanılan bir derin ağı mimarisidir. Bu mimaride, sinir hücreleri arasındaki bağlantılar yönlendirilmiş bir çizge oluşturur. RNN mimarisi, herhangi bir uzunluktaki girdinin işlenmesini olanaklı kılar. Hesaplama esnasında, geçmiş bilgiler de dikkate alınır [11]. RNN mimarilerinde, genellikle kaybolan gradyan problemine rastlanır. RNN mimarilerinde, katman sayısına dayalı olarak artan ya da azalan çarpımsal gradyan nedeniyle uzun vadeli bağımlılıkların modellenmesi zordur [6].

3.2.2. Uzun Kısa Süreli Bellek Mimarisi

Uzun kısa süreli bellek birimleri (LSTM), geleneksel RNN mimarilerinde karşılaşılan kaybolan gradyan problemini önlemeye yönelik, tekrarlayan bir sinir ağı mimarisidir. LSTM sınırlı sayıda zaman adımı ile hatanın geriye doğru yayılmasına izin verir. Temel bir LSTM birimi, bir hücre ve giriş geçidi, çıkış geçidi ve unutmaya geçidi olmak üzere üç temel kapıdan oluşur. Kapıların durumuna göre, hangi bilgilerin korunması gerektiği ve birimlere ne zaman erişileceği belirlenir [5].

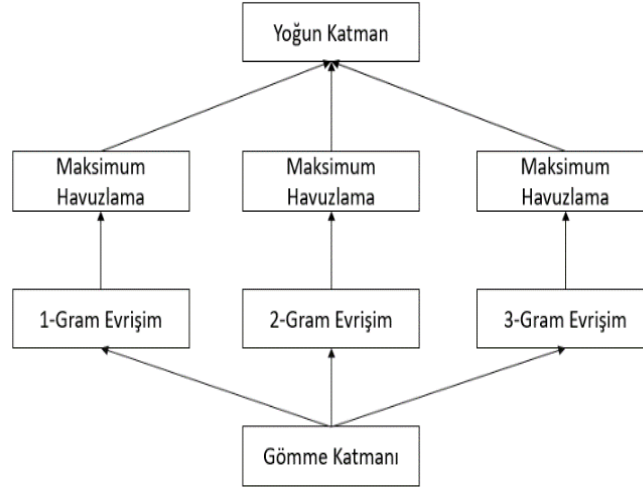
3.2.3. Geçitli Tekrarlayan Birim Mimarisi

Geçitli tekrarlayan birim (GRU), LSTM'e kıyasla daha az karmaşık bir mimariye sahip ve benzer deneysel sonuçlar veren bir tekrarlayan sinir ağı mimarisidir [6, 13]. Temel bir GRU mimarisinde, sıfırlama kapısı ve güncelleme kapısı olmak üzere iki kapı yer alır.

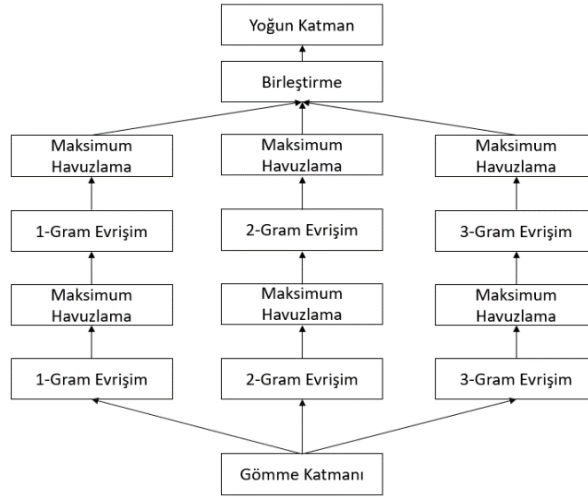
3.3. Evrişimli Sinir Ağı Tabanlı Metin Madenciliği Mimarileri

Bu bölümde, evrişimli sinir ağı tabanlı üç temel mimari sunulmaktadır. Tanıtılan evrişimli sinir ağı mimarilerinin gömme katmanlarında, metin belgeleri kelime gömme yöntemleri kullanılarak temsil edilmektedir.

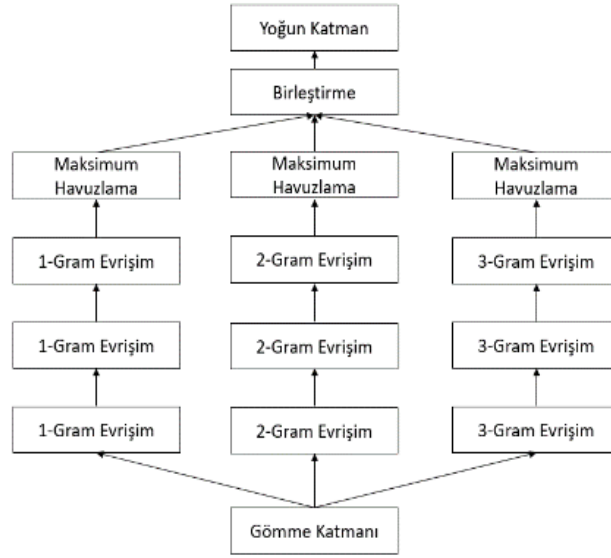
Çalışma esnasında, daha önceden öğrenilmiş sözcük vektörleri kullanmanın daha yüksek başarımlar verdiği gözlemlendiğinden, sözcük vektörlerinin rastgele olarak eğitim esnasında öğrenildiği modeller göz ardı edilmiştir. Kelime gömmeleri için, vektör boyutu 300 olarak alınmıştır. Şekil 1'de evrişimli sinir ağı tabanlı birinci mimari özetlenmektedir.



Şekil. 1. Evrişimli sinir ağı tabanlı temel mimari (CNN1)



Şekil. 2. Yinelemeli evrişim ve maksimum havuzlama tabanlı mimari (CNN2)



Şekil 3. Evrişim tabanlı piramit mimari (CNN3)

Mimarinin gömme katmanında, metin belgesinde yer alan kelimeler için, kelime gömme yöntemleri tabanlı temsil elde edilmektedir. Ardından, evrişim katmanları yığını (1-gram, 2-gram ve 3-gram) kullanılarak 1-gram, 2-gram ve 3-gram tabanlı özniteliklerin çıkarımı gerçekleştirilmektedir. Her bir katmanda, öznitelik haritalarının oluşturulması için sabit sayıda 80 filtre uygulanmaktadır. Katmanlarda, aktivasyon fonksiyonu olarak, yüksek deneysel başarımından ötürü, $relu(x)=max(0, x)$ fonksiyonu kullanılmıştır. Havuzlama katmanlarında, bölgesel maksimum havuzlama, [8]'deki parametre setine göre alınmıştır. Yoğun katman, daha önceki katmanlarda elde edilen öznitelik haritalarının girdi olarak verildiği ileri beslemeli bir yapay sinir ağıdır. Yoğun katman, 80 yapay sinir ağı hücresi içermektedir. Modelde, aşırı öğrenmeyi engellemek amacıyla L2 düzenleme (0.1 düzenleme katsayısı ile) ve bırakma (0.30 oranı ile) uygulanmıştır. Kayıp değerini hesaplamak ve eğitim sürecini optimize etmek için, ikili çapraz entropi ve Adam algoritması kullanılmıştır. Deneysel analizde, veri seti 70/10/20 şeklinde bölümlenerek eğitim, geliştirme ve test aşamalarında kullanılmıştır. Şekil 2'de evrişimli sinir ağı tabanlı ikinci mimari sunulmaktadır. Bu mimari, temelde Şekil 1'de sunulan mimariyle aynı özellikleri taşımak ile birlikte, gömme katmanı sonucu elde edilen metin temsili, yinelenen evrişim ve maksimum havuzlama katmanlarına tabi tutulmaktadır. Şekil 3'te ise evrişim tabanlı piramit mimarinin genel yapısı özetlenmektedir.

4. Deneysel Süreç ve Sonuçlar

Bu bölümde, deneysel analizlerde kullanılan veri seti, gerçekleştirimde kullanılan yöntemlere ve izlenen deneysel sürece ilişkin ayrıntılar ve derin öğrenme mimarileri ile elde edilen deneysel sonuçlar sunulmaktadır.

4.1. Deneysel Süreç

Deneysel analizlerde, geliştirilen evrişimli sinir ağı mimarilerinin başarımlarını değerlendirmek için Türkçe Twitter mesajları içeren bir veri seti oluşturulmuştur. Veri seti, Twitter API kullanılarak Python dilinde yazılmış bir uygulamayla iki aylık bir süreçte elde edilmiştir. Oluşturulan duygu analizi setinde, 10500'ü pozitif ve 10500'ü negatif olmak üzere, toplam 21000 Twitter mesajı yer almaktadır. Veri seti üzerinde, dizgi parçalama, durak kelimeleri çıkarma, kök bulma gibi ön işleme aşamaları uygulanmıştır. Dizgi parçalama aşamasında, hem pozitif, hem de negatif bir ifade bildiren Twitter mesajları veri setinden kaldırılmıştır. Bunun yanı sıra, mesajlarda yer alan harflerin her biri küçük harfe dönüştürülmüş, noktalama işaretleri, rakamlar ve '@', '#' gibi özel karakterler ortadan kaldırılmıştır. Metin mesajları, terim sayısı ve karakter uzunluğuna göre filtrelenmiş, yinelenen harfler çıkarılmıştır. Durak kelimelerinin çıkarılması için, Lucene uygulama geliştirme ara yüzü, kök bulma aşamasında ise Zemberek kütüphanesinden yararlanılmıştır. Geliştirilen evrişimli sinir ağı tabanlı mimarilerin performansının değerlendirmesinde doğru sınıflandırma oranı kullanılmıştır.

4.2. Deneysel Sonuçlar

Veri setinin, geleneksel makine öğrenmesi sınıflandırıcıları ve veri temsil yöntemleriyle performansını ölçümlenmek adına, veri seti, terim sıklığı (TF), terim varlığı (TP), TF-IDF ölçütü ve 1-gram, 2-gram ve 3-gram modelleri kullanılarak temsil edilmiş; böylelikle, 9 farklı veri temsili elde edilmiştir. Bu veri temsil yöntemleri ile temsil edilen veri setinin başarımı, dört temel sınıflandırıcı olan, k-en yakın komşu algoritması (KNN), destek vektör makineleri (SVM), lojistik regresyon (LR) ve Naive Bayes (NB) ile değerlendirilmiştir. Makine öğrenmesi tabanlı deneysel analizlerde, 10-kat çapraz geçişleme kullanılmıştır. Gerçekleştirim, WEKA 3.9 ile varsayılan parametre değerleri aracılığıyla yapılmıştır. Derin öğrenme tabanlı modellerin gerçekleştirimi, Tensorflow ve Keras kullanılarak yapılmıştır. Derin öğrenme modellerinde uygun hiperparametrelerin belirlenmesi için Gauss işlemi kullanılarak

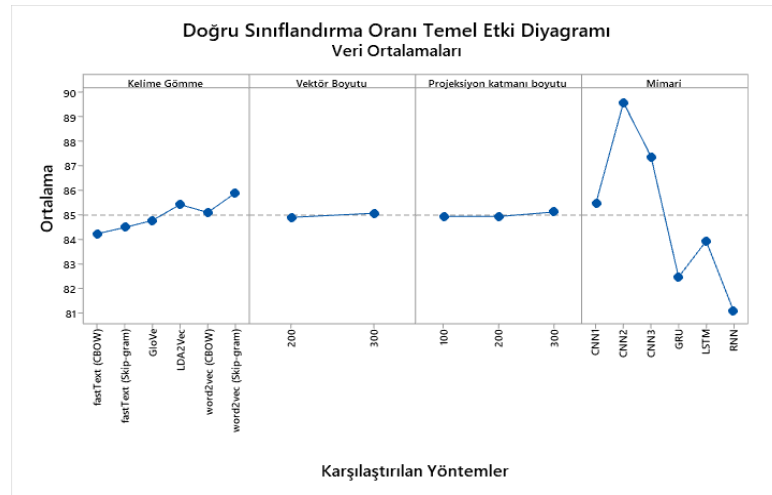
Bayes optimizasyonu uygulanmıştır. LDA2vec kelime gömme yöntemi için, konu sayısı (N=25) ve negatif örnekleme katsayısı ($\beta \in 0.75$) olarak alınmıştır. Tablo 1’de temel makine öğrenmesi tabanlı yöntemler ile Tablo 2’de ise derin öğrenme mimarileri ile elde edilen doğru sınıflandırma oranları sunulmuştur. Temel kelime gömme yöntemleri için farklı vektör boyutları ve projeksiyon katmanı boyutları incelenmiş olup Tablo 2’de en yüksek başarımın alındığı vektör boyutu 300 ve projeksiyon katmanı boyutu 300 için sonuçlar verilmiştir.

Tablo 1. Makine Öğrenmesi Yöntemlerine İlişkin Deneysel Sonuçlar

Veri Temsil Yöntemleri	K-en yakın komşu algoritması	Destek vektör makineleri	Lojistik regresyon	Naive Bayes algoritması
1-gram +TP	76.10	79.58	78.25	80.71
1-gram +TF	76.63	79.83	78.61	80.94
1-gram +TF-IDF	76.40	79.69	78.41	80.85
2-gram +TP	74.65	79.16	77.64	80.38
2-gram +TF	75.72	79.47	78.07	80.61
2-gram +TF-IDF	75.29	79.35	77.92	80.50
3-gram +TP	71.36	78.76	76.89	79.98
3-gram +TF	73.93	79.07	77.38	80.26
3-gram +TF-IDF	72.93	78.93	77.19	80.08

Tablo 2. Derin Öğrenme Yöntemlerine İlişkin Deneysel Sonuçlar

Kelime Gömme	RNN	LSTM	GRU	CNN1	CNN2	CNN3
word2vec (Skip-gram)	81.82	84.69	83.16	86.29	92.53	88.44
word2vec (CBOW)	81.34	84.18	82.67	85.72	89.62	87.77
fastText (Skip-gram)	80.90	83.64	82.16	85.15	88.96	86.85
fastText (CBOW)	80.44	83.41	81.97	84.94	88.69	86.54
GloVe	81.13	83.88	82.40	85.41	89.26	87.27
LDA2Vec	81.54	84.40	82.90	86.00	90.16	88.16



Şekil. 4. Karşılaştırılan yöntemlere ve faktörlere ilişkin temel etki diyagramı

Şekil 4’te karşılaştırılan yöntemlere ilişkin temel etki diyagramı sunulmuştur. Kelime gömme yöntemleri incelendiğinde en yüksek başarımların word2vec (skip-gram) modeli ile elde edildiği, bu yöntemi, LDA2vec modelinin izlediği görülmektedir. Tüm modeller içerisinde, CBOW modelleri, skip-gram modellerine kıyasla daha düşük başarımlar vermektedir. Derin öğrenme tabanlı mimarilerin, geleneksel makine öğrenmesi tabanlı yöntemlere kıyasla daha yüksek başarımlar elde ettiği gözlenmektedir. Önerilen

evrişimli sinir ağı tabanlı üç mimarinin (CNN1, CNN2 ve CNN3), metin madenciliğinde sıklıkla kullanılan, RNN, LSTM ve GRU gibi mimarilerden daha yüksek başarımla çalıştığı görülmektedir.

5. Sonuç

Duygu analizi (görüş madenciliği), metin belgeleri içerisinde yer alan nesnelere, ürünlere, servislere ya da organizasyonlara ilişkin görüş, duygu, tutum gibi öznel bilgilerin, makine öğrenmesi, istatistik ve doğal dil işleme gibi alanlardan teknik ve yöntemlerin kullanılması ile çıkarılmasını amaçlayan bir araştırma alanıdır. Duygu analizi, yapısal olmayan bilgiden, yapısal, anlamlı ve kullanışlı bilgiler çıkarılmasını olanaklı hale getirir. Bu çalışmada, Türkçe metin belgeleri üzerinde duygu analizi için, evrişimli sinir ağı tabanlı üç temel derin öğrenme mimarisinin etkinliği değerlendirilmektedir. Metin belgelerinin temsili, word2vec, fastText, GloVe ve LDA2vec olmak üzere dört temel kelime gömme yöntemi incelenmektedir. Çalışma kapsamında önerilen birinci mimaride, gömme katmanında, metin belgesinde yer alan kelimeler için, kelime gömme yöntemleri tabanlı temsil elde edilmektedir. Ardından, evrişim katmanları yığını (1-gram, 2-gram ve 3-gram) kullanılarak 1-gram, 2-gram ve 3-gram tabanlı özniteliklerin çıkarımı gerçekleştirilmektedir. Her bir katmanda, öznitelik haritalarının oluşturulması için sabit sayıda 80 filtre uygulanmaktadır. İncelenen ikinci evrişimli sinir ağı tabanlı mimaride gömme katmanı sonucu elde edilen metin temsili, yinelenen evrişim ve maksimum havuzlama katmanlarına tabi tutulmaktadır. İncelenen üçüncü mimari ise, evrişim tabanlı piramit mimarisidir. Metin belgelerinin temsili, word2vec, fastText, GloVe ve LDA2vec olmak üzere dört temel kelime gömme yöntemi incelenmektedir. İncelenen evrişimli sinir ağı tabanlı mimarilerin, Türkçe duygu analizi için, geleneksel makine öğrenmesi sınıflandırıcılarına (k-en yakın komşu algoritması, destek vektör makineleri, lojistik regresyon ve Naive Bayes algoritması) ve temel derin öğrenme mimarilerine (tekrarlayan sinir ağları, uzun kısa süreli bellek birimleri ve geçitli tekrarlayan birim) kıyasla daha yüksek başarımla çalıştığı görülmektedir. Word2vec (Skip-gram modeli) kelime kodlaması yöntemi ile evrişimli sinir ağı tabanlı mimari ile %92.53 doğru sınıflandırma oranı elde edilmiştir. İncelenen evrişimli sinir ağı mimarileri arasındaki en yüksek başarımla çalışılan evrişim ve maksimum havuzlama tabanlı mimari ile alınmaktadır.

Kaynakça

1. Fersini, E., Messina, E., & Pozzi, F. A. (2014). Sentiment analysis: Bayesian ensemble learning. *Decision support systems*, 68, 26-38.
2. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey. *Ain Shams engineering journal*, 5(4), 1093-1113.
3. Taboada, M., Brooke, J., Tofiloski, M., Voll, K., & Stede, M. (2011). Lexicon-based methods for sentiment analysis. *Computational linguistics*, 37(2), 267-307.
4. Tan, S., & Zhang, J. (2008). An empirical study of sentiment analysis for chinese documents. *Expert Systems with applications*, 34(4), 2622-2629.
5. Collobert, R., Weston, J., Bottou, L., Karlen, M., Kavukcuoglu, K., & Kuksa, P. (2011). Natural language processing (almost) from scratch. *Journal of machine learning research*, 12(Aug), 2493-2537.
6. Zhang, L., Wang, S., & Liu, B. (2018). Deep learning for sentiment analysis: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4), e1253.
7. Şeker, A., Diri, B., & Balık, H. H. (2017). Derin öğrenme yöntemleri ve uygulamaları hakkında bir inceleme. *Gazi Mühendislik Bilimleri Dergisi (GMBD)*, 3(3), 47-64.
8. Çano, E., & Morisio, M. (2019). A data-driven neural network architecture for sentiment analysis. *Data Technologies and Applications*, 53(1), 2-19.
9. Güngör, O., Üsküdarlı, S., & Güngör, T. (2018, May). Recurrent neural networks for Turkish named entity recognition. In *2018 26th Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE.
10. Collobert, R., & Weston, J. (2008, July). A unified architecture for natural language processing: Deep neural networks with multitask learning. In *Proceedings of the 25th international conference on Machine learning* (pp. 160-167).
11. Dos Santos, C., & Gatti, M. (2014, August). Deep convolutional neural networks for sentiment analysis of short texts. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers* (pp. 69-78).
12. Kim, Y. (2014). Convolutional neural networks for sentence classification. *arXiv preprint arXiv:1408.5882*.
13. Zhang, X., Zhao, J., & LeCun, Y. (2015). Character-level convolutional networks for text classification. In *Advances in neural information processing systems* (pp. 649-657).
14. Johnson, R., & Zhang, T. (2017, July). Deep pyramid convolutional neural networks for text categorization. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)* (pp. 562-570).
15. Çano, E., & Morisio, M. (2018, March). Role of data properties on sentiment analysis of texts via convolutions. In *World Conference on Information Systems and Technologies* (pp. 330-337). Springer, Cham.
16. Mikolov, T., Chen, K., Corrado, G., & Dean, J. (2013). Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781*.
17. Pennington, J., Socher, R., & Manning, C. D. (2014, October). Glove: Global vectors for word representation. In *Proceedings of the 2014 conference on empirical methods in natural language processing (EMNLP)* (pp. 1532-1543).