



Yığılanmış Özdevinimli Kodlayıcı ile İnsan Aktivitelerinin Sınıflandırılması

Classification of Human Activity by Using a Stacked Autoencoder

Hasan BADEM¹, Abdullah CALISKAN², Alper BASTURK¹, Mehmet Emin YUKSEL²

¹Bilgisayar Mühendisliği Bölümü, Erciyes Üniversitesi, Kayseri, Türkiye
{hbadem, ab}@erciyes.edu.tr

²Biyomedikal Mühendisliği Bölümü, Erciyes Üniversitesi, Kayseri, Türkiye
{acaliskan, yuksel}@erciyes.edu.tr

Özetçe —Bu çalışmada iki özdevinimli kodlayıcıdan oluşan yığılanmış özdevinimli kodlayıcı ve bir softmax katmanına sahip derin sinir ağı yapısının insan aktivitelerini sınıflandırma amacıyla uygulanması incelenmiştir. Bu yapının performansı yaygın olarak kullanılan Human Activity Recognition Using Smartphones data seti ile test edilmiştir. Derin sinir ağının parametreleri uygun şekilde ayarlandığında bilinen klasik yöntemlerden daha iyi sınıflandırma sonuçları elde edildiği görülmüştür.

Anahtar Kelimeler—Derin Sinir Ağı, Yığılanmış Özdevinimli Kodlayıcı, Softmax, İnsan aktivitesi.

Abstract—This paper investigates the application of a deep neural network architecture that consists of stacked autoencoder with two autoencoders and a softmax layer for the purpose of human activity classification. The performance of the proposed architecture is tested on a commonly used data set known as Human Activity Recognition Using Smartphones. It is observed that the proposed method yields better classification results than the representative state-of-the-art methods provided that the parameters of the deep network are suitably optimized.

Keywords—Deep Neural Network, Stacked Autoencoder, Softmax, Human Activity Recognition.

I. GİRİŞ

Günümüzde gelişen teknoloji ile birlikte eğlence, spor ve sağlık gibi alanlarda insanların günlük yaşamlarında sıklıkla yaptığı yürüme, oturma, merdiven çıkma v.b günlük aktivitelerin sınıflandırılmasına ihtiyaç duyulmaktadır. Bu temel aktivitelerin sınıflandırılması video görüntülerinden ve giyilebilir sensörler vasıtasıyla yapılabilmektedir. Ancak bu yöntemlerin birçoğunun pahalı ve zahmetli yöntemler olması herkes tarafından ulaşılabilirliğini sınırlamaktadır. Bu yöntemlere alternatif olarak günümüz akıllı telefonlarının sahip olduğu sensör teknolojilerinin kullanılmasıyla bu temel aktivitelerin izlenebilmesi mümkün olmaktadır.

İnsan aktivitelerinin algılanması ve sınıflandırılması üzerine birçok çalışma yapılmıştır. Bu çalışmalar kişilerin video ile

izlenerek [1], vücudunun belirli yerlerine sensör yerleştirilerek [2] ve giyilebilir teknoloji sistemleri [3], [4] v.b yöntemlerle alınan verilerin değerlendirilmesi ve sınıflandırılması esasına dayanmaktadır. Ancak bu yöntemlerde teknoloji maliyetinin yüksek olması, teknolojiye erişimin zahmetli ve sınırlı olması aynı zamanda üretilen verinin anlamlandırılmasının zor olması gibi bazı dezavantajların barındığı görülmektedir [5].

Derin öğrenme bilimin muhtelif alanlarında kullanılmakta olup, nesne tanıma, ses tanımlama, yüz tanıma ve dil çevirimi gibi alanlarda oldukça başarılı sonuçlar üretmiştir. Derin öğrenme zor problemlerin çözümünde üstün performans göstermektedir. Bu da bilimin birçok alanında derin öğrenmenin uygulanabilirliğini artırmaktadır [6]. Bu gözlemden hareketle, kolay ulaşılabilir ve neredeyse herkesin yanında taşıdığı akıllı telefonların gelişmiş sensörleri aracılığıyla elde edilen verilerin derin öğrenme yöntemleri yardımıyla yorumlanarak hareketlerin sınıflandırılmasının mümkün olabileceği akla gelmektedir.

Bu çalışmada bir grup kişinin yürüme, merdiven çıkma, merdiven inme, oturma, ayakta durma ve yatma olmak üzere 6 farklı aktivitesinin sınıflandırılması yapılmıştır. Human Activity Recognition Using Smartphones (HARUSD) [5] veri setinde yer alan 30 kişinin yaptığı aktivitelerin sınıflandırılması derin öğrenme yapılarından yığılanmış özdevinimli kodlayıcı (YÖK) kullanılarak yapılmıştır. Veri setinde yer alan 561 adet öznitelik, YÖK yapısının ilk özdevinimli kodlayıcısına (ÖK) giriş olarak verilmektedir. İlk ÖK'nin gizli katman çıkışı alınarak yeni özellikler ikinci ÖK'ye giriş olarak verilmektedir. Eğer arzu edilirse tek ÖK'de kullanılabileceği gibi ikiden fazla ÖK'de kullanılarak bir ÖK'nin gizli katman çıkışı bir sonraki ÖK'ye verilmek suretiyle işleme devam edilebilir. Bu ardi ardına dizilmiş ÖK'lerin tamamı YÖK yapısı olarak tanımlanmaktadır [7]. Son ÖK'nin çıkışı sınıflandırma birimi olan softmax birimine giriş olarak verilir. Tüm bu işlemlerden sonra daha önce ayrı ayrı eğitime tabi tutulan ÖK'ler ve softmax biriminin tüm parametreleri tekrar optimize edilir.

Bu çalışmanın, ikinci bölümünde YÖK yapısı ve tasarlanan ağ yapısı hakkında bilgi verilecektir. Üçüncü bölümde deneysel sonuçlar verilmektedir. Son bölümde ise çalışmadan elde



Sinyal İşleme 5

3. Gün / 29 Ekim 2016, Cumartesi

edilen sonuçlar değerlendirilmektedir.

II. TASARLANAN YIĞINLANMIŞ ÖZDEVİNİMLİ KODLAYICI

Şekil 1'de görülen tasarlanan YÖK'de, gizli katmanlarında dörder nörona sahip iki ÖK katmanı ve sınıflandırma birimi olarak softmax katmanı yer almaktadır. Şekil 1 (a-b) görüldüğü gibi ağın eğitimine her bir ÖK eğitilerek başlanır. Sonra ikinci ÖK'nin gizli katman çıkışı Şekil 1 (c) görüldüğü gibi softmax katmanına gönderilir. Daha sonra iki ÖK ve softmax katmanı birleştirilerek Şekil 1 (d) elde edilir. Ağın tüm parametreleri bir kez daha optimize edilerek ağ son halini alır [7]–[9]. Tüm eğitim aşamalarında, Limited-memory-Broyden–Fletcher–Goldfarb–Shanno Quasi-Newton (L-BFGS) optimizasyon algoritması kullanılmıştır.

A. Özdevinimli Kodlayıcı

ÖK bir giriş katmanı, bir çıkış katmanı ve bir adet gizli katmanı olan yapay sinir ağı modelidir. ÖK'nin parametreleri kendi girişini çıkışta elde etmek için eğitilir. Bundan dolayı ÖK'nin giriş sayısı çıkış sayısına her zaman eşittir. Giriş (\mathbf{x}) ve çıkış ($\hat{\mathbf{x}}$) verisinin boyutu M , gizli katmandaki nöron sayısı N olmak üzere ÖK'nin gizli katman çıkışı aşağıdaki gibi ifade edilir [9], [10].

$$\mathbf{c} = f(\mathbf{b} + \mathbf{W}^T \mathbf{x}) \quad (1)$$

Burada $\mathbf{c} = [c_1 \ c_2 \ \dots \ c_N]^T$ gizli katman çıkışı, $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_M]^T$ girişi, $\mathbf{b} = [b_1 \ b_2 \ \dots \ b_N]^T$ biası, $\mathbf{W} = [\mathbf{w}_1 \ \mathbf{w}_2 \ \dots \ \mathbf{w}_N]$ giriş katmanı ile gizli katman arasında kalan kısmın ağırlıklandırma parametrelerini, f nöron aktivasyon fonksiyonunu ifade etmektedir.

ÖK'nin çıkışı ise aşağıdaki şekilde ifade edilir.

$$\hat{\mathbf{x}} = \hat{f}(\hat{\mathbf{b}} + \hat{\mathbf{W}}^T \hat{\mathbf{c}}) \quad (2)$$

Burada $\hat{\mathbf{x}} = [\hat{x}_1 \ \hat{x}_2 \ \dots \ \hat{x}_M]^T$ çıkışı $\hat{\mathbf{b}} = [\hat{b}_1 \ \hat{b}_2 \ \dots \ \hat{b}_M]^T$ biası $\hat{\mathbf{W}} = [\hat{\mathbf{w}}_1 \ \hat{\mathbf{w}}_2 \ \dots \ \hat{\mathbf{w}}_M]$ gizli katman ile çıkış katmanı arasında kalan kısmın ağırlıklandırma parametrelerini göstermektedir.

ÖK'ün eğitiminde aşağıdaki maliyet fonksiyonu kullanılır.

$$E_T = \frac{1}{T} \sum_{k=1}^T e_k^2 + \frac{\lambda}{2} (\|\mathbf{W}\| + \|\hat{\mathbf{W}}\|) \quad (3)$$

Burada λ ağırlıklandırma faktörüdür. Yukarıdaki denklemde ifade edilen arzu edilen çıkış ile ağın çıkışı arasındaki fark e_k aşağıdaki gibi ifade edilir.

$$e_k = \|\mathbf{x}^{(k)} - \hat{\mathbf{x}}^{(k)}\| \quad (4)$$

Burada $k = 1 \dots T$ örneklem sayısını ifade etmektedir.

Aşağıdaki düzenleme yardımıyla maliyet fonksiyonuna ayrıca bir seyreklik kısıtı eklenir. Seyreklik kısıtı ÖK'nin ilginç öznitelikler çıkarmasına sebep olur.

Tablo I: Kullanılan parametre tablosu

Veri Seti	1. ÖK	2. ÖK
HARUSD	Nöron Sayısı	4
	ρ	0.05
	β	3
	Nöron Sayısı	4
	ρ	0.1
	β	2

$$E_{sparse} = E_T + \beta \sum_{j=1}^N KL(\rho \parallel \hat{\rho}_j) \quad (5)$$

Amaç fonksiyonuna eklenen $\beta \sum_{j=1}^N KL(\rho \parallel \hat{\rho}_j)$ parametresi, ceza terimi olarak, β ise seyreklik ceza terimi olarak tanımlanmaktadır. Birer Bernoulli rastgele değişkeni olan seyreklik parametresi ρ ve (7)'den hesaplanan ortalama aktivasyon değeri $\hat{\rho}_j$ değerleri arasındaki ilişki, Kullback-Leibler (KL) diverjansı olarak tanımlanmaktadır. Özdevinimli kodlayıcı için KL diverjansı (6)'dan hesaplanmaktadır [9], [10].

$$KL(\rho \parallel \hat{\rho}_j) = \rho \log \frac{\rho}{\hat{\rho}_j} + (1 - \rho) \log \frac{1 - \rho}{1 - \hat{\rho}_j} \quad (6)$$

$$\hat{\rho}_j = \frac{1}{T} \sum_{i=1}^T f_j(\mathbf{x}(i); \mathbf{W}, \mathbf{b}) \quad (7)$$

B. Softmax

Softmax, iki ya da daha fazla sınıfın olduğu durumlarda kullanılan olasılık tabanlı bir doğrusal sınıflandırıcıdır. İki sınıfı ayırabilen mantıksal regresyonun genişletilmiş halidir. Yiğınlanmış özdevinimli kodlayıcıya softmax sınıflandırıcı birimi olarak eklenmektedir [11].

III. DENEYSEL SONUÇLAR

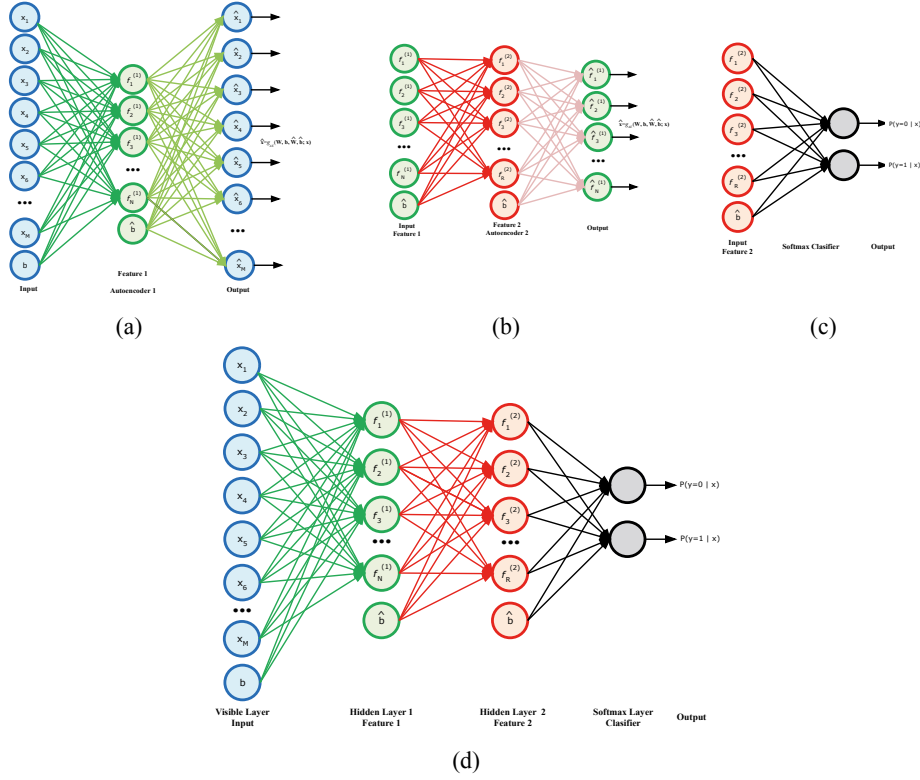
Tasarlanan ağ yapısının, eğitim aşamasında kullanılan çeşitli parametreleri Tablo I'de görülmektedir. Bunun yanında eğitim aşamalarının tamamında L-BFGS algoritması 400 iterasyon koşulurulmuştur. Ayrıca softmaxın, ÖK'nin ve YÖK'ün eğitiminde $\lambda = 0.003$ alınmıştır. Bu parametreler ile elde edilen ağ yapısı Intel i7 2600 3.4 Ghz işlemcili, 12 GB DDR3 RAM bellekli sistemde, 10-kat çapraz doğrulama ile koşulurulmuş ve deneysel sonuçlar elde edilmiştir. Elde edilen deneysel sonuçlar literatürde yer alan aynı amaca yönelik belli başlı yöntemlerle kıyaslanmıştır.

A. HARUSD Veri Seti

HARUSD veri setinin temel amacı bireyin akıllı telefonundan ivme ve jiroskop sensörlerinden alınan sinyaller üzerinden bireyin gerçekleştirmiş olduğu aktivitelerin belirlenmesidir. Bu çalışmada kullanılan veri seti Data Mining Repository of University of California, Irvine (UCI) [12] dan alınmıştır. Veri setinde 30 adet deneğin aktivitelerinin sınıflandırılması Tablo II'de görülen özniteliklere göre belirlenmeye çalışılmıştır. Bu öznitelikler akıllı telefonun ivme ve jiroskop senörlerinden alınan zaman serilerinden çıkarılmıştır. Zaman serilerinin örneklem frekansı 50 Hz olmak üzere, öznitelikler 2.56s uzunluğunda pencereler yardımıyla çıkarılmıştır [5].

Sinyal İşleme 5

3. Gün / 29 Ekim 2016, Cumartesi



Şekil 1: Yığılmış özdevimli kodlayıcı modeli

Tablo II: Çalışmada kullanılan öznitelikler

1	<i>mean</i> : ortalama değer
2	<i>std</i> : standart sapma
3	<i>mad</i> : mutlak ortanca değer
4	<i>max</i> : maksimum değer
5	<i>min</i> : minimum değer
6	<i>sma</i> : sinyalin magnitüde alanı
7	<i>energy</i> : sinyalin karelerinin toplamının ortalaması
8	<i>arCoeff</i> : otoregresyon katsayıları
9	<i>correlation</i> : korelasyon katsayısı
10	<i>maxFreqInd</i> : en geniş frekans bileşeni
11	<i>meanFreq</i> : frekans domaininde sinyalinin ağırlıklandırılmış ortalaması
12	<i>skewness</i> : frekans domaininde çarpıklık
13	<i>kurtosis</i> : frekans domaininde basıklık
14	<i>energyBand</i> : frekans aralığının enerjisi
15	<i>angle</i> : iki vektör arasındaki açı

Tablo III: HARUSD Veri Setinin Karşılaştırma Tablosu

Metot	Doğruluk Oranı (%)
Önerilen YÖK	89.64
KNN	86.98
Karar Ağacı	64.90
Navie Bayes	58.44

verisinin öznitelikleri hem azaltıp hem de altı sınıfa ait veriyi birbirinden uzaklaştırarak sınıflandırma işlemi kolaylaştırmaktadır. YÖK'nin sağladığı bu avantaj ile veri setine uygulanan bu yapının seyreklik parametresi, seyreklik ceza parametresi ve ağırlık yapısının optimum değerleri tespit edilerek kıyaslama yapılan yöntemlerden daha başarılı sonuçlar elde edilmiştir.

B. YÖK Benzetim Sonuçları

Veri setindeki 30 birey rastgele sıralandıktan sonra 10 parçaya bölünmek suretiyle bu parçalardan 27 tanesi eğitim verisi geri kalan 3 tanesi ise test verisi olarak kullanılmıştır. Eğitim aşaması tamamlandıktan sonra test verisi üzerinden doğruluk değeri elde edilmiştir. 10 kat çapraz doğrulama işlemi, 10 defa eğitim ve test işlemi gerçekleştirilmekte; ancak her seferinde eğitim ve test verisi değiştirilmektedir. Elde edilen 10 farklı doğruluk değerinin ortalaması Tablo III'de rapor edilmekte ve Şekil 2'de 10 doğruluk değerinden en yüksek olanının hata matrisi verilmiştir.

YÖK'de yer alan her bir özdevimli kodlayıcı, giriş

Bu tablolar incelendiğinde, tasarlanan YÖK yapısının sınıflandırma başarımının literatürde önerilen 10 kat çapraz doğrulama tekniğinin uygulandığı diğer yöntemlerden belirgin şekilde yüksek olduğu gözlenmiştir. Çapraz doğrulama tekniği ile test işleminde kullanılan veriden kaynaklanabilecek avantaj ve dezavantajlar ortadan kaldırılarak adil bir kıyaslama sunulmuştur.

IV. SONUÇ

Bu çalışmada derin öğrenme yöntemleri yardımıyla insan günlük aktivitelerinin sınıflandırılması problemi ele alınmış ve literatürde yer alan mevcut yöntemlere göre oldukça iyi sonuçlar elde edilmiştir. Derin öğrenme yöntemlerinin her

Sinyal İşleme 5

3. Gün / 29 Ekim 2016, Cumartesi

Hata Matrisi

Sistemin Çıkışı	Yürme	Mesken Çıkma	Mesken İniş	Çıkma	Ayakta Durma	Yatma	Toplam
Yürme	217 19.6%	3 0.3%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	98.6% 1.4%
Mesken Çıkma	0 0.0%	166 15.0%	0 0.0%	1 0.1%	0 0.0%	4 0.4%	97.1% 2.9%
Mesken İniş	0 0.0%	0 0.0%	155 14.0%	0 0.0%	0 0.0%	0 0.0%	100% 0.0%
Çıkma	0 0.0%	0 0.0%	0 0.0%	171 15.5%	2 0.2%	6 0.5%	95.5% 4.5%
Ayakta Durma	0 0.0%	0 0.0%	0 0.0%	7 0.6%	190 17.2%	0 0.0%	96.4% 3.6%
Yatma	0 0.0%	0 0.0%	0 0.0%	0 0.0%	0 0.0%	184 16.6%	100% 0.0%
Toplam	100% 0.0%	98.2% 1.8%	100% 0.0%	95.5% 4.5%	99.0% 1.0%	94.8% 5.2%	97.9% 2.1%

Beklenen Çıkış

Şekil 2: Önerilen YÖK yapısının 10 kat çapraz doğrulamada en iyi olduğu durumun hata matrisi

geçen gün yeni bir uygulaması sunulmakta ve bu yöntemler klasik yöntemlere göre daha başarılı sonuçlar vermektedir. Buradan bu yöntemlerin biyomedikal alanındaki uygulamalarının ilerleyen dönemlerde daha da yaygınlaşması beklenebilir. Bu çalışmanın devamında aktivite sınıflandırma yardımıyla çeşitli hastalıkların teşhis ve takibi konusu çalışılabilir.

KAYNAKÇA

- [1] C. Thureau, V. Hlavac, "Pose primitive based human action recognition in videos or still images", *IEEE Conference on Computer Vision and Pattern Recognition-CVPR 2008.*, Anchorage, AK, 2008, pp. 1-8.
- [2] R. Nishkam, D. Nikhil, M. Preetham, M.L. Littman, "Activity recognition from accelerometer data", *In Proceedings of the Seventeenth Conference on Innovative Applications of Artificial Intelligence*, 2005, pp. 1541-1546.
- [3] P. Lukowicz, J.A. Ward, H. Junker, M. Stager, G. Troster, A. Atrash, T. Starner, "Recognizing work-shop activity using body worn microphones and accelerometers", *Proceedings of the 2nd Int Conference Pervasive Computing*, 2004, pp. 18-22.
- [4] D.M. Karantonis, M.R. Narayanan, M. Mathie, N.H. Lovell, B.G. Celler. "Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring", *IEEE Transactions on Information Technology in Biomedicine*, vol.10, no.1, pp.156-167, 2006.
- [5] D. Anguita, A. Ghio, L. Oneto, X. Parra and J. L. Reyes-Ortiz, "A Public Domain Dataset for Human Activity Recognition Using Smartphones", *Proc. ESANN*
- [6] Y. LeCun, Y. Bengio, G. Hinton, "Deep Learning", *Nature*, vol.521, pp.436-444, 2015.
- [7] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, "Greedy Layer-Wise Training of Deep Networks", *19. Advances in Neural Information Processing Systems*, 2006, pp. 153-160.
- [8] P. Vincent, H. Larochelle, Y. Bengio, P.A. Manzagol "Extracting and Composing Robust Features with Denoising Autoencoders", *Proceedings of the 25th international conference on Machine learning*, 2008, pp. 1096-1103.
- [9] J. Ngiam, A. Coates, A. Lahiri, B. Prochnow, Q. V. Le, A. Y. Ng, "On optimization methods for deep learning", *in: Proceedings of the 28th International Conference on Machine Learning (ICML-11)*, 2011, pp. 265-272.

- [10] Y. Bengio, "Practical recommendations for gradient-based training of deep architectures", *in: Neural Networks: Tricks of the Trade*, Springer, pp. 437-478, 2012.
- [11] Y. Zhang, E. Zhang, W. Chen, "Deep neural network for halftone image classification based on sparse auto-encoder", *Engineering Applications of Artificial Intelligence*, vol.50, pp.245-255, 2016.
- [12] Lichman, M. UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science, 2013.