

Tek-Sınıf Destek Vektör Makineleri Kullanılarak EEG İşaretlerinin Sınıflandırılması

EEG Signal Classification Using One-Class Support Vector Machines

Ercan Avşar¹, Mustafa Ersel Kamaşak², Zehra Çataltepe²

İstanbul Teknik Üniversitesi

¹Elektronik ve Haberleşme Mühendisliği Bölümü,

²Bilgisayar Mühendisliği Bölümü

{avsare,kamasak,cataltepe}@itu.edu.tr

Özetçe

Elektroensefalogram (EEG) işareti beyindeki elektriksel aktivite hakkında bilgi edinmek için yaygın olarak kullanılan bir araçtır. Epilepsi tanısında da kullanılmaktadır. Bu çalışmanın amacı, tek-sınıf destek vektör makineleri (DVM) kullanarak, epileptik EEG işaretlerini otomatik olarak tespit etmektir. Bunu yaparken, EEG işaretini analiz etmek amacıyla, literatürde geçen 28 adet öznelik kullanılmıştır. Bu öznelikler üzerinde Temel Bileşen Analizi (TBA) uygulanarak öznelik uzayının boyutu düşürülmüştür. 4 farklı eğitim kümesinde, tek-sınıf DVM'nin tüm uzay ve TBA sonrasında oluşturulan daha az boyutlu uzaydaki sınıflandırma performansı incelenmiştir. 6-7 arasında TBA bileşeni ile enerjinin %95'inin kapsandığı ve 4 eğitim kümesinin 3 tanesinde 28 öznelik kullanılmasından daha iyi sınıflandırma performansı alındığı görülmüştür.

Abstract

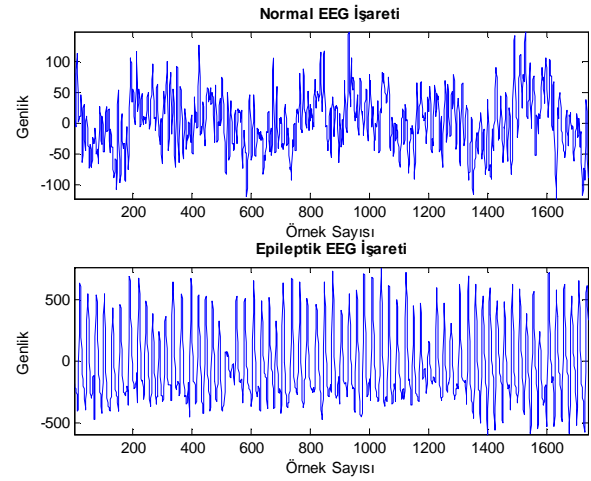
Electroencephalogram (EEG) is a widely used tool to obtain information about electrical activity of the brain. It is also used for epilepsy diagnosis. The aim of this work is to detect epileptic EEG signals automatically using one-class support vector machines (SVM). To do this, for the purpose of analyzing EEG signal, 28 features, which have been used in literature, are used. Dimension of the feature space is reduced by applying Principal Component Analysis (PCA) on these features. On 4 different training sets, classification performance of one-class SVM on the 28 dimensional space and the PCA reduced dimensional space is investigated. It is observed that 95% of energy is covered using 6-7 PCA components and in 3 out of 4 training sets, PCA features have better classification results than using all 28 features.

1. Giriş

EEG beynin elektriksel aktivitesini gösteren bir işarettir. Standart bir EEG kaydı hastanın kafasına yüzey elektrotlarının yapılandırılması ile elde edilir. Bunun dışında, girişimsel olarak doğrudan beynin yüzeyinden de EEG kayıtları alınabilir.

Epilepsi aralıklı olarak tekrar eden krizlerle tanımlanan bir nörolojik rahatsızlıktır. Bu krizler beyindeki anormal nöral aktiviteler yüzünden oluşur ve EEG işaretinde gözlemlenebilirler. Dolayısı ile epilepsi krizinin tespitinde EEG işaretleri oldukça faydalı araçlar olarak karşımıza

çıkılmaktadır. Şekil 1'de normal ve epileptik EEG işaretleri gösterilmektedir.



Şekil 1: 10'ar saniyelik Normal ve Epileptik EEG işareti örneği

Bu amaçla şimdiye kadar farklı öznelikler kullanan farklı yöntemler öne sürülmüştür. Güler ve Übeyli, Adaptive Neuro-Fuzzy Inference Sistem (ANFIS) modelinde dalgacık katsayılarını (wavelet coefficients) kullanarak %98.68'lik sınıflandırma doğruluğuna ulaşmışlardır [1]. Srinivasan, Eswaran ve Sriraam zaman ve frekans bilgisini kullanarak çıkardıkları 5 tane öznelik ve farklı sistem parametreleri ile eğittikleri yapay sinir ağlarını kullanarak doğruluk oranını %99'un üzerine çıkarmayı başarmışlardır [2]. Subaşı ise ayrık dalgacık dönüşümü (ADD) katsayılarını kullanarak eğitilen Mixture of Experts (ME) modelinin standart yapay sinir ağlarından daha iyi sonuç verdiğini göstermiştir [3]. Aşağıda Tablo 1.'de değişik çalışmalarda kullanılan öznelikler listelenmiştir.

Bu çalışmanın daha önceliklere göre katkıları şunlardır:

- Literatürde kullanılmış olan 28 tane öznelik birden çıkarılmış ve kullanılmıştır.
- Bu özneliklerden sınıflandırma için önemli olan boyutların elde edilmesi için TBA yöntemi kullanılmıştır.
- Daha önceki çalışmaların çoğunda sınıflandırıcı başarımına doğruluk kullanılarak karar verilmiştir. Fakat epileptik

hastalardan alınan EEG verileri çok sayıda hastaliksız ve az sayıda hastalıklı veri noktası içermektedir. Böyle problemler için ROC eğrisinin altında kalan alan (AUC) başarımı daha doğru ölçtüğü için, bu çalışmada başarımlar ölçümünde AUC kullanılmıştır.

d) Epilepsi tespiti için yapılmış olan birçok çalışmada kullanılan sınıflandırıcının eğitimi için hastanın epilepsi krizi geçirmedeği anda alınmış EEG kaydının (normal EEG) yanı sıra kriz anında alınmış EEG kaydına (epileptik EEG) da ihtiyaç duyulmaktadır. Bugünün yöntemleriyle, krizin ne zaman başlayacağı daha önceden kestirilemediği için hastayı uzun süreli EEG kaydına tabi tutmak gerekebilir. Bu da hem hasta için rahatsızlık verici bir durumdur, hem de kayıt süresinin uzaması ile yapılan işlemin maliyeti de artmaktadır. Bu çalışmada kullanılan tek-sınıf destek vektör makinelerinin (DVM) eğitimi için epileptik EEG verisine ihtiyaç duyulmamaktadır. Böylelikle, hasta için diğerlerinden daha hesaplı bir yöntem oluşturulması ümit edilmektedir.

Bildirinin geri kalan kısmı şu bölümlerden oluşmaktadır. Bölüm 2’de, kullanılan EEG işaretleri hakkında açıklama yapılmakta ve öznitelikler listelenmektedir. Bölüm 3’te, özniteliklerin nasıl çıkarıldığı açıklanmakta ve sınıflandırma için kullanılan tek sınıflı destek vektör makineleri ile boyut azaltımı için kullanılan temel bileşenler analizi tanıtılmaktadır. Bölüm 4’te EEG verisi üzerinde yapılan deneyler ayrıntılarıyla anlatılmakta ve bu deneylerin sonuçları verilmektedir. Son olarak bölüm 5’te yorumlar yer almaktadır.

2. Veri Kümesi ve Öznitelikler

2.1. EEG Veri Kümesi

Bu çalışmada kullanılan EEG işaretleri [6] da ayrıntılı bir biçimde açıklanan veri tabanından alınmıştır. Bu veri tabanında, A, B, C, D, E olmak üzere, 5 farklı EEG veri kümesi bulunmaktadır. A kümesi, sağlıklı kişilerden, gözler açık biçimde, yüzey elektrotları ile kaydedilmiş EEG verisini içerir. B kümesi, sağlıklı kişilerden, gözler kapalı biçimde, yüzey elektrotları ile kaydedilmiş EEG verisini içerir. C ve D kümeleri, hastalıklı kişilerden, epilepsi krizinin olmadığı anda, girişimsel olarak kaydedilmiş EEG verisini içerir. E kümesi ise hastalıklı kişilerden, epilepsi krizi anında, girişimsel olarak kaydedilmiş EEG verisini içerir. Bu EEG işaretleri 173.61 Hz’lik örnekleme frekansı ile sayısallaştırılmıştır ve her bir küme 39.33 dakikalık kayıt içermektedir.

2.2. Öznitelikler

Özniteliklerin detayları için Tablo 1’in 2. Sütununda verilen kaynaklara bakınız.

3. Kullanılan Yöntemler

3.1. Öznitelik Çıkarma:

Öznitelik çıkarma işlemi, 1 saniye uzunluğundaki pencereyi 0.5 saniyelik adımlarla EEG işareti üzerinde kaydırarak yapılmıştır. Buradaki her bir adım için Tablo-1’de listelenmiş olan 28 adet öznitelik çıkarılmıştır.

Tablo 1: Literatürde kullanılan EEG öznitelikleri

Öznitelik İsmi	Kullanıldığı Çalışmalar
Spektral Entropi	[2,7,8]
Renyi Entropisi	[7]
Varyans	[9]
Petrosian Fraktal Boyutu	[10]
Hjorth Mobilitesi	[10]
Hjorth Kompleksitesi	[10]
Ortalama Eğri Uzunluğu	[5,8]
Ortalama Enerji	[5,8]
Ortalama Teager Enerjisi	[5]
Wigner-Wille Katsayıları (4 adet)	[9]
Dalgacık Katsayıları (15 adet)	[3]

3.2. Tek-Sınıf Destek Vektör Makineleri (DVM)

DVM’nin dayandığı ana fikir, öznitelik uzayında iki sınıfı en uygun olarak ayırabilecek bir hiperdüzlem bulmaktır. Diğer birçok doğrusal öğrenme makineleri de bu mantıkla çalışmasına karşın, DVM onlardan farklı olan yönü, test edilecek nesnelere için yanlış sınıflandırma olasılığını en aza indirecek bir çözüm sunabilir (structural risk minimization)[4].

Tek-sınıf DVM durumunda, veri öncelikle uygun bir çekirdek fonksiyon kullanılarak öznitelik uzayına taşınır ve daha sonra bir hiperdüzlem vasıtasıyla orijinden en fazla olacak şekilde ayrılır. Bu hiperdüzlem parametreleri normal DVM’dekine benzer bir kuadratik problemin çözülmesi ile elde edilir [5]:

$$\min \left(\frac{1}{2} \|w\|^2 + \frac{1}{vl} \sum_{i=1}^l \xi_i - \rho \right) \quad (1)$$

$$(w \cdot \Phi(x_i)) \geq \rho - \xi_i \quad i = 1, 2, \dots, l \quad \xi_i \geq 0 \quad (2)$$

Burada w ve ρ hiperdüzlem parametreleri, Φ çekirdek fonksiyon, v yanlış sınıflandırılan değerlerin (outlier) izin verilen oranı, l eğitim kümesindeki nesnelere sayısı ve ξ ise hata parametresidir.

Bu çalışmada epilepsi olan verilerin etiketi -1 ile olmayan verilerin etiketi ise +1 ile gösterilmiştir. Test kümesindeki veriler için karar çıktıları şöyle hesaplanmaktadır:

$$f(x) = \text{sgn}(w \cdot \Phi(x) - \rho) \quad (3)$$

3.3. Temel Bileşenler Analizi (TBA)

TBA (Principal Component Analysis: PCA), yüksek boyutlu veri kümelerinde öznitelik uzayının boyutunu azaltarak benzer performansın daha az sayıda öznitelik kullanılmasıyla elde edilmesini amaçlayan bir yöntemdir. Bunu yaparken de

öznitelik vektörleri arasındaki ilişkiyi kullanır. PCA adımlarını şöyle sıralayabiliriz:

- Eğitim kümesi özniteliklerinin ortalamasının eğitim ve test kümesi özniteliklerinden çıkarılması.
- a'nın sonucunda elde edilen eğitim kümesinin kovaryans matrisinin hesaplanması.
- Kovaryans matrisinin özvektör ve bunlara karşılık gelen özdeğerlerin bulunması.
- Seçilen enerji eşik değerini içerecek en büyük özdeğerlere denk gelen özvektörlerin alınarak izdüşüm matrisinin oluşturulması.
- a'nın sonucunda elde edilen kümelerin bu izdüşüm matrisiyle çarpılarak boyutun düşürülmesi.

4. Deneysel Sonuçları

Kullanılan EEG verisindeki A, B, C, D kümeleri epilepsi krizinin olmadığı anda kaydedildiğinden bu kümelerin her biri için farklı sınıflandırıcılar eğitilmiştir. Bu eğitilen sınıflandırıcılar, epileptik veriyi içeren E kümesi üzerinde test edilmiştir. Eğitim işleminden önce ilgili kümedeki (A, B, C ya da D) verinin %10'luk kısmı test için ayrılmış, %90'luk kısmı ise sınıflandırıcının eğitiminde kullanılmıştır.

Eğitim ve test işlemleri hem tüm öznitelikler kullanılarak hem de PCA kullanılarak özniteliklerin öz uzaydaki enerjisinin %95'ini içerecek şekilde boyut azaltım işleminden sonra yapılmıştır. Performans ölçütü olarak her iki durumdaki tüm sınıflandırıcılar için ROC eğrileri çıkarılmış ve AUC değerleri hesaplanmıştır. ROC eğrisi ikili sınıflandırma sistemlerinde farklı eşik değerleri için gerçek pozitif oranının (True Positive Rate: $TPR=TP/(TP+FN)$) yanlış pozitif oranına (False Positive Rate: $FPR=FP/(FP+TN)$) göre değişimini gösterir. Bu eğrinin altında kalan alan ise bize AUC değerini vermektedir. Dolayısı ile sonuç 1'e ne kadar yakın olursa sistemin o kadar iyi çalıştığı yorumu yapılabilir.

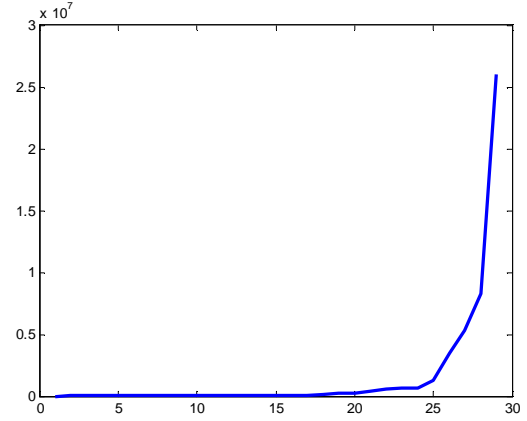
4.1. PCA

PCA yöntemi elimizdeki veri kümeleri kullanılarak oluşturulan eğitim ve test kümelerine 3.3. bölümde anlatıldığı şekilde uygulanmıştır. Bu işlem sonucunda elimizdeki 28 boyutlu öznitelik uzayının boyutu, eğitim için A kümesi kullanılması durumunda 7'ye, B, C ve D kümesi kullanılması durumunda 6'ya düşmüştür. C kümesi için özdeğerler Şekil 2'de gösterilmiştir. Özdeğerlerin çoğunun küçük olması TBA ile boyut azaltılsa da performansta düşme olmayabileceğine işaret etmektedir.

4.2. Tek-sınıf DVM

Elimizdeki örnekler DVM'de kullanılmadan önce ölçeklenmiştir ($z=(x-\mu)/\sigma$). Burada μ öznitelik ortalaması, σ öznitelik standart sapması, x ise ölçeklenecek öznitelik vektörüdür.

Uygulama MATLAB ortamında gerçekleştirilmiş ve ücretsiz bir yazılım olan LIBSVM'den [11] faydalanılmıştır. DVM çekirdek fonksiyonu olarak, $\gamma=1$ olan Gauss RBF kullanılmıştır. Tek-sınıf DVM'nin yanlış sınıflandırmaların izin verilen oranını kontrol eden parametresi olan ν 'nün iki farklı değeri için ($\nu=0.1$ ve $\nu=0.5$) bu adımlar tekrar edilerek, bu parametrenin sistemin performansı üzerindeki etkisi incelenmiştir.

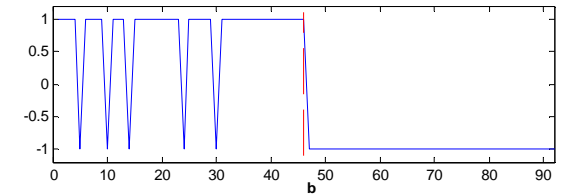
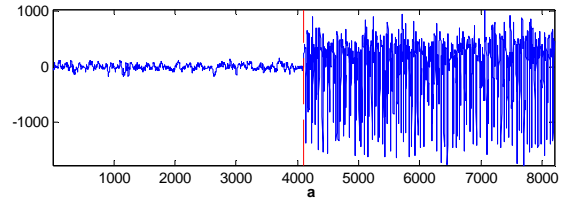


Şekil 2: C kümesi için özdeğerler

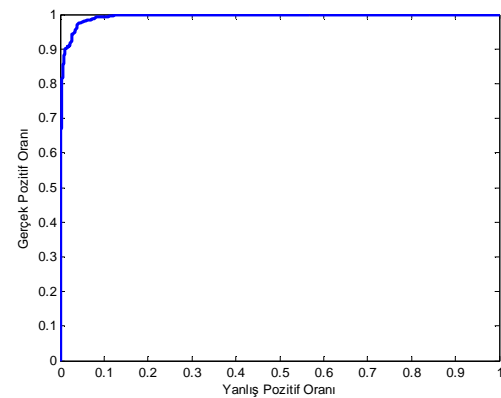
4.3. Sonuçlar

C kümesi kullanılarak eğitilen sınıflandırıcının C ve E kümesi üzerinde test edilmesi ile elde edilen sınıflandırıcı çıktılarını ($f(x)$) ve ROC eğrisi Şekil 3 ve 4'te gösterilmektedir.

Tüm sınıflandırıcılar için AUC değerleri ise Tablo 2 ve Tablo 3'te gösterilmektedir. Tablolarda görülen AUC değerlerinin çok yüksek olması, sınıflandırıcıların tüm ya da PCA uzayında, üzerlerinde eğitilmedikleri bir veri kümesinde de iyi sonuç verdiklerini göstermektedir. D hariç bütün veri kümelerinde PCA sonuçları bütün 28 öznitelik kullanılmasıyla daha iyidir.



Şekil 3: C ve E kümelerindeki test girdi ve çıktıları.



Şekil 4: C kümesi için ROC eğrisi. AUC=0.9941

Tablo 2: AUC Değerleri ($v = 0.1$)

Eğitim Kümesi	Tüm Öznitelikler	PCA
A	0.9935	0.9978
B	0.9925	0.9928
C	0.9823	0.9940
D	0.9858	0.9589

Tablo 3: AUC Değerleri ($v = 0.5$)

Eğitim Kümesi	Tüm Öznitelikler	PCA
A	0.9935	0.9978
B	0.9925	0.9933
C	0.9823	0.9942
D	0.9864	0.9840

5. Yorumlar

Bu çalışmada tek-sınıf DVM'den yararlanılarak epileptik EEG işaretleri tespit edilmiştir. Şimdiye kadar bu yönde yapılmış olan birçok çalışmada sınıflandırıcının eğitimi için edinilmesi bir hayli güç olan epileptik veriye de ihtiyaç duyulmuştur. Fakat tek-sınıf DVM'nin çekici tarafı eğitim için sadece normal veriye ihtiyaç duyması ve bu sınıflandırıcı için yanlış sınıflandırılan örnekleri epileptik olarak etiketlemesidir.

Bu amaçla 28 adet öznitelik çıkarılmış ve TBA yöntemiyle boyut azaltma işlemine gidilmiştir. Özdeğerlerin, bazı öznitelikler için diğerlerine göre daha baskın olduğu, dolayısı ile boyutun muazzam bir şekilde azaltılabileceği görülmüştür. Böylelikle sınıflandırıcının karmaşıklığının azaltıldığı gibi eğitim ve test sürelerinde de düşüş olmuştur. Bununla birlikte, Tablo 1 ve Tablo 2'deki sonuçlara baktığımızda da TBA uygulanması durumunda sistem performansının tüm öznitelikler kullanılması durumuna göre artma eğiliminde olduğu gözlemlenmiştir. Son olarak, SVM'in v parametresi farklı 2 değeri için de PCA sonuçlarının daha iyi olduğu gözlemlenmiştir.

6. Kaynakça

- [1] İ. Güler and E. D. Übeyli, "Adaptive Neuro-Fuzzy Inference System for Classification of EEG Signals Using Wavelet Coefficients," *Journal of Neuroscience Methods*, vol. 148, issue 2, pp. 113-121, Oct. 2005.
- [2] V. Srinivasan, C. Eswaran and N. Sriraam, "Artificial Neural Network Based Epileptic Detection Using Time-Domain and Frequency-Domain Features," *Journal of Medical Systems*, vol. 29, no. 6, pp. 647-660, Dec. 2005.
- [3] A. Subaşı, "EEG Signal Classification Using Wavelet Feature Extraction and a Mixture of Expert Model," *Expert Systems with Applications*, vol. 32, issue 4, pp. 1084-1093, May 2007
- [4] V. N. Vapnik, *The Nature of Statistical Learning Theory*. Springer-Verlag, New York, 1999
- [5] A. B. Gardner, A. M. Krieger, G. Vachtsevanos and B. Litt, "One-Class Novelty Detection for Seizure Analysis

from Intracranial EEG," *Journal of Machine Learning Research*, vol. 7, pp. 1025-1044, Dec 2006.

- [6] R. G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David and C. E. Elger, "Indications of Nonlinear Deterministic and Finite-Dimensional Structures in Time Series of Brain Electrical Activity: Dependence on Recording Region and Brain State," *Physical Review E*, vol. 64, Nov 2001.
- [7] N. Kannathal, U. Rajendra Acharya, C. M. Lim, and P. K. Sadasivan, "Entropies for detection of epilepsy in EEG," *Computer Methods and Programs in Biomedicine*, vol. 80, issue 1, pp. 17-23, June 2005.
- [8] M. D'Alessandro, R. Esteller, G. Vachtsevanos, A. Hinson, J. Echaus and B. Litt, "Epileptic Seizure Prediction Using Hybrid Feature Selection Over Multiple Intracranial EEG Electrode Contacts: A Report of Four Patients," *IEEE Trans. On Biomedical Engineering*, Vol.50, No.5, pp. 603-615, , May 2003
- [9] H. R. Mohseni, A. Maghsoudi and M. B. Shamsollahi, "Seizure Detection in EEG signals: A Comparison of Different Approaches," *28th Annual International Conference of IEEE EMBS*, pp. 6724-6727, Aug 2006.
- [10] F. S. Bao, D. Yu-Chun Lie and Y. Zhang, "A New Approach to Automated Epileptic Diagnosis Using EEG and Probabilistic Neural Network," *IEEE ICTAI'08*, pp. 482-486, Nov 2008.
- [11] Chih-Chung Chang and Chih-Jen Lin, LIBSVM: a library for support vector machines, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>