

EEG İşaretlerinden Matematiksel İşlemlerin k-NN Algoritması İle Sınıflandırılması

Classification of Mathematical Tasks from EEG Signals Using k-NN Algorithm

Server Göksel ERALDEMİR¹, Esen YILDIRIM², Yakup KUTLU³

¹ İskenderun Meslek Yüksekokulu
Mustafa Kemal Üniversitesi
Hatay/Türkiye
servergoksel@hotmail.com

^{2,3} Bilgisayar Mühendisliği
Mustafa Kemal Üniversitesi
Hatay/Türkiye
eyildirim@mku.edu.tr², yakupkutlu@gmail.com³

Özet

Bu çalışmada matematiksel işlemlerden toplama-çıkartma ve çarpma-bölme gruplarının birbirinden ayırt edilmesi hedeflenmiştir. Çalışma sırasında kullanılan veriler 17 sağlıklı denek kullanılarak gönüllülük esasına göre 22 elektrot kullanılarak toplanmıştır. Çalışmada kayıt edilen veriler Ayırık Dalgacık Dönüşümü ile analiz edilerek öznitelik vektörleri oluşturulmuştur. Oluşturulan öznitelik vektörleri k-En yakın Komşuluk algoritmasına göre k=1, 3, 5 ve 10 değerleri için sınıflandırılmıştır. Ortalama doğru pozitif sonuçları 1,3,5 ve 10 en yakın komşuluk değerlerine göre sırası ile %79.3, %74.9, %72.4 ve %68.6 olarak elde edilmiştir.

Abstract

The purpose of this work is classification of summation-subtraction and multiplication-division groups. Database employed for the analysis includes EEG signals collected from 17 voluntary subjects from 22 electrodes. Discrete Wavelet Transform is used to extract the features. Features are classified using k-Nearest Neighbors algorithm for k values of 1, 3, 5 and 10. True positive rates are obtained as 79.3%, 74.9%, 72.4% and 68.6% for k=1, 3, 5, and 10 respectively.

1. Giriş

Beynin her türlü işlevi sonucu oluşan elektriksel sinyallerinin kayıt edilmesine Elektroensefalogram (EEG) denir. Yapılan çalışmalar kişinin zihinsel her türlü aktivitesinde genlik değeri 1-100µV, frekans değeri ise 0.5-100 Hz arasında değişen elektriksel sinyallerin oluştuğunu göstermiştir [1].

EEG kayıtları içerdiği frekans değerine, oluşum zamanına ve genliğine göre değerlendirilir. EEG işaretlerinin frekans, genlik ve zaman bilgisi ani ve sürekli olarak değişmektedir [1]. EEG işaretleri, bu ani değişimler nedeniyle durağan olmayan sinyallerdir. Durağan işaretlerin analizinde genellikle Fourier dönüşümü (FD) kullanılırken EEG sinyallerinin analizinde FD kullanıldığında sinyalin zaman bilgisi kaybolduğu için FD yetersiz kalmaktadır [2]. Literatürde bulunan birçok çalışmada Ayırık Dalgacık Dönüşümü (ADD) kullanılmaktadır. ADD analizinde esnek zaman-frekans pencereleri kullanılarak sinyalin farklı

zamanlarda sergilediği ani değişimler zaman bilgisi kaybedilmeden analiz edilebilir [2,3,4]. ADD, EEG işaretlerinin analizinde sıkça kullanılmaktadır.

Federice ve Carole yaptıkları çalışmada öğrencilerin matematik sorularını çözerken alınan EEG verilerinden doğru cevap verip vermediğini öğrencileri izleyen bireylerden daha iyi şekilde tahmin etmişlerdir [5]. Çalışmalarında öğrencilerin kolay problemleri çözerken doğru tahmin oranı %87 olurken, zor problemlerin çözümleri sırasında doğru tahmin oranı %78 olarak gerçekleşmiştir. Bir diğer çalışmada Sakkalıs ve arkadaşları zor problemlerin çözümü sırasında kayıt edilen EEG verilerinden, problemlerin çözümleri sırasında ön bölge (frontal lob) ve merkez bölge (central lob)'de teta (4-8Hz), alpha1 (8-10Hz), gamma1 (30-45Hz) ve gama2 (45-90Hz) dalgalarında artış olduğunu tespit ederek bu iki bölgenin zor problem çözümlerinde aktif olarak kullanıldıkları sonucuna varmışlardır [6].

Yazarlar tarafından yapılan önceki çalışmalarda beynin metinsel ve matematiksel işlemler sırasında oluşan farklılıkların sınıflandırılması amacıyla gönüllü deneklerden veriler toplanmış, toplanan veriler farklı sınıflandırma algoritmaları ile sınıflandırılmışlardır. Ayrıca öznitelik seçimi yoluyla kanal analizi yapılarak etkili kanallar tespit edilmiş, 26 kanallı sınıflandırma sonuçlarına yakın doğru pozitif oranları elde edilmiştir. Çalışma sonucunda 26 kanal kullanılarak yapılan analizlerde elde edilen doğru pozitif oranları BayesNet, Random Forest ve k-en yakın komşuluk (k-NN) algoritmaları için sırasıyla %88.8, %90 ve %89.5'dir [7,16]. 10 kanallı doğru pozitif sonuçları ise BayesNet, Random Forest ve k-NN algoritmaları için sırasıyla %90, %90.1 ve %86.9'dır [7].

Bu çalışmada, sayısal işlemlerden toplama, çıkartma bölme ve çarpma işlemlerinden oluşan görseller esnasında kaydedilen EEG işaretlerinin toplama-çıkartma ile çarpma-bölme gruplarının ayırt edilmesi hedeflenmektedir. Bu amaçla 17 gönüllü denekten toplanan veriler gürültüden temizlenerek bölütlere ayrılmış ve ADD ile öznitelikler çıkarılarak k-NN algoritmasına göre sınıflandırılmıştır.

2. Materyal ve Yöntem

Çalışmada kullanılan veriler, tarafımızdan 22 elektrot ile 26 kanal üzerinden, Nihon Kohden 1200 marka EEG kayıt cihazı

ile toplanmıştır. Bu veriler ilk olarak ön işleminden geçirilmiş daha sonra Dalgacık Dönüşümü analizinde uzunluğunu 1 saniye olan kayan pencere kullanılarak öznitelikler çıkartılmış 1-NN, 3-NN, 5-NN ve 10-NN değerleri kullanılarak en yakın komşuluk algoritmasına göre sınıflandırılmıştır.

2.1. Verilerin Toplanması

EEG verileri 18 gönüllü denekten uluslararası 10-20 sistemine uygun olarak kafa derisi üzerine yerleştirilen elektrotlar yardımı ile toplanmıştır. Toplanan bu veriler içerisinde 3. deneğe ait veriler tüm sınıflandırıcılarda kötü sonuç verdiği için bu deneğin deney sırasında dikkatli davranmadığı düşünülmektedir [7]. Bundan dolayı bu çalışmada bu deneğe ait veriler analiz dışı bırakılmıştır.

Çekimden önce denekler saçlarının temiz olması, jöle gibi saç şekillendiricilerinin olmaması konusunda bilgilendirilmişlerdir. Çekim sırasında deneklerden işlemleri yapmaya çalışmaları istenmiş, heyecan faktörüne karşılık, işlemlerin yetiştirilmesinin gerekli olmadığı vurgulanmış ve sadece ekrandaki işleme odaklanmaları gerektiği ifade edilmiştir. Ayrıca çekim esnasında olabildiğince hareketsiz durmaları gerektiği söylenmiş ve başlangıçta rahat pozisyonda oturmaları sağlanmıştır. Denekler 19-24 yaş arasında yükseköğrenim gören ve gözlük kullanmayan gençler arasından seçilmiştir. Çalışmada deneklere, Türkçe metinler içeren sözel ve dört işlem içeren sayısal slaytlar gösterilmiştir. Slaytların her biri 13,25 saniyelik olup toplam 60 slayttan oluşmaktadır. Slaytlardan ilk 30'u sayısal işlemlerden, sonraki 30'u ise metinlerden oluşmaktadır. Bu çalışmada sadece matematiksel işlemler esnasındaki EEG işaretleri analiz edilmiş olup, veritabanı hakkında ayrıntılı bilgi [7]'de verilmektedir.

Deneklere gösterilen matematiksel işlemlerin örnekleri Şekil 1 ve Şekil 2' de gösterilmektedir.

EEG kayıtları 1 kHz örnekleme frekansıya, 22 elektrot kullanılarak alınmıştır. Bu elektrotlardan referans noktasına göre 26 kanallık veri kaydı oluşturulmuştur; Fp1-A1, Fp2-A2, F3-A1, F4-A2, C3-A1, C4-A2, P3-A1, P4-A2, O1-A1, O2-A2, F7-A1, F8-A2, T3-A1, T4-A2, T5-A1, T6-A2, Fp2-O2, Fp1-O1, Fp1-Fp2, F7-F8, F3-F4, T3-T4, C3-C4, T5-T6, P3-P4 ve O1-O2.

$$\begin{array}{r} 3762 \\ + 1255 \\ \hline 5017 \end{array} \quad (a)$$

$$\begin{array}{r} 6493 \\ - 5631 \\ \hline 862 \end{array} \quad (b)$$

Şekil 1: (a)Toplama İşlemi Örnek Slaytı ve (b) Çıkartma İşlemi Örnek Slaytı

$$\begin{array}{r} 5312 \quad 9 \\ \hline \end{array} \quad (a)$$

$$\begin{array}{r} 42650 \\ * 7 \\ \hline \end{array} \quad (b)$$

Şekil 2: (a) Bölme İşlemi Slayt Örneği ve (b) Çarpma işlemi Slayt örneği

2.2. Ön İşlem

Kayıt sırasında oluşan 50 Hz'lik şebeke gürültüsü EEG cihazı tarafından doğrudan temizlenmiş ve 120 Hz kesme frekansı ile alçak geçiren filtre uygulanmıştır. Böylelikle veriler yüksek frekanslı gürültülerden temizlenmiştir. Elde edilen verileri bölütlere ayırma işleminde ilk slaytlar ve son slayt senkronizasyondan dolayı oluşabilecek sıkıntıları gidermek için kullanılmamıştır. Geriye kalan slaytlardan 6 tanesi çıkartma, 8 tanesi toplama, 7 tanesi bölme ve 7 tanesi çarpma işlemini içermektedir. Analiz edilen işlemlerden hepsinin eşit sayıda olmasını sağlamak amacıyla her işlemde 6'şar adet alınmış ve sonuç olarak veriler, 12'şer adetlik iki guruba ayrılmıştır. Bu gruplardan birincisi 6 adet toplama ve 6 adet çıkartma verisinden oluşurken, ikincisi 6 adet bölme ve 6 adet çarpma verisinden oluşmaktadır. Böylelikle bu iki grubun birleştirilmesi ile 24 adet slayta ait veriler gruplandırılmıştır.

2.3. Özniteliklerin Elde edilmesi

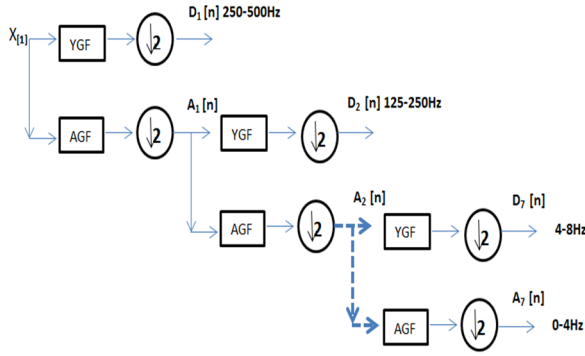
Bu çalışmada öznitelik çıkarımı için literatürde bir çok çalışmada kullanılan Ayrık Dalgacık Dönüşümü (ADD) yöntemi kullanılmıştır [3,6,7,8,12].

2.3.1 Ayrık Dalgacık Dönüşümü

ADD, dalgacık analizi için çeşitli dalgacıkları kullanır bu dalgacıklara filtre bankaları adı verilir. Dalgacıklar, işareti frekans bantlarına ayırarak filtreler [9]. Bu bantlardan, Şekil 4'te görüldüğü gibi orijinal işareti yeniden elde edebileceğimiz şekilde yaklaşım (A: Approximation) ve detay (D: Detail) katsayıları elde edilmektedir. Dalgacık dönüşümü belirlenen çözünürlüğe ulaşıncaya kadar genişletilebilir.

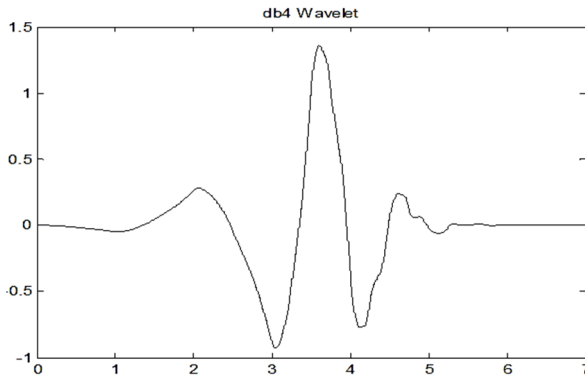
2.3.2 Özniteliklerin Çıkartılması

Bu çalışmada veriler 1 saniye uzunluğunda 0,25 sn'lik kaydırmalardan (%75 örtüşmeli) elde edilen pencerelere bölünmüştür. Her pencere ADD ile analiz edilerek öznitelikler çıkartılmıştır. Analiz için Şekil 4' de görülen db4 ana dalgacık fonksiyonu kullanılmıştır.



Şekil 3: DB4 Dalgacık Grafiği

Çıkarılan özneliklerde 1 kHz'lik örnekleme frekansı göz önüne alındığında, Şekil 3'te görüldüğü gibi ilk üç detay katsayısı yüksek frekans aralığına sahip olduğundan D1, D2 ve D3 katsayıları kullanılmamıştır. Öznelik vektörü oluşturulurken D4, D5, D6 ve D7 detay katsayıları ve A7 yaklaşım katsayısı kullanılmıştır. Öznelikler bu katsayıların standart sapması ve katsayıların mutlak değerlerinin ortalaması, maksimumu ve minimumu kullanılarak oluşturulmuştur.



Şekil 4: 7. Seviye Ayrık Dalgacık Dönüşümü

2.4. Sınıflandırma

Bu çalışmada sınıflandırma işlemi için literatürde EEG işaretlerinin sınıflandırmasında kullanılan k-NN algoritması kullanılmıştır [10,11,12]. Çalışmada k değeri 1, 3, 5 ve 10 olarak belirlenmiş ve bu komşuluk değerlerine göre sınıflandırma sonuçları verilmiştir.

2.4.1. k-NN Algoritması

k-En yakın komşuluk (k-NN) algoritması örnekler arasındaki uzaklık değerine göre çalışan bir teoremdir. Bu teoreme göre veriye en yakın olan k adet nokta seçilir sınıflandırma işlemi bu seçilen k sayısına göre yapılır. K sayısının küçük olması benzerlik oranı en yüksek olan örnekleri bir sınıfta toplamaya yardımcı olurken, k sayısının artışı ile birbirine benzemeyen veriler de aynı sınıfa dahil edilerek yanlış sınıflandırma sonuçları ile karşılaşılabilir [13]. Bu nedenle k değerinin seçilmesi için kullanılan yöntemlerden biri, zaman almasına rağmen, çeşitli k değerleri için sınıflandırma sonuçlarının karşılaştırılmasıdır. Örneğin, Özer ve Amasyalı'nın

çalışmasında verilerin sınıflandırılması 1 en yakın komşuya göre yapıldığında ortaya çıkan sınıflandırma başarısı, 3, 5, 7, 9 ve 10 en yakın komşuluk sınıflandırmalarına göre daha iyi sonuç vermiştir [14].

Bu algoritmada veri setinden bir bölüm öğrenme kümesi olarak ayrılır ve eğitim seti olarak kullanılır, bu eğitim seti dışında kalan veri seti ise test için kullanılır. Test için kullanılan verilerin içerisinde bir örnek, eğitim seti içerisinde kendisine en yakın olan k adet örneğin öklid değerine göre en yakın olan eğitim verisini bulur. Bulunan bu veriye göre sınıfı belirlenir, bu işlem test verisindeki tüm verilerin sınıflandırması bitene kadar tekrarlanır [11,15].

2.5. Değerlendirme

Çalışma sonuçlarının değerlendirilmesinde 10 kat çapraz doğrulama kullanılmıştır. Bu yöntemde veri kümesi iki parçaya ayrılır. Birinci aşamada veriler rastgele 10 eşit parçaya ayrılır veri kümelerinin her defa bir tanesi test için, geri kalan dokuz tanesi ise eğitim için kullanılır. Bu işlem on defa tekrarlanır ve bu on tekrarın ortalaması alınır. Bu işlemde her veri bir defa test verisi olarak kullanılır.

3. SONUÇLAR

Çalışma sonucunda toplama-çıkartma işlem grubu ile çarpma-bölme işlem grubu için k-NN sınıflandırma sonuçları doğru pozitif değerlerinin ortalaması 1, 3, 5 ve 10 komşuluk değerlerine göre Tablo 1'de verilmiştir.

Tablo 1: 1, 3, 5 ve 10 Değerlerine Göre Ortalama Sonuçlar

k-NN	Ortalama Doğru Pozitif Değeri	Ortalama Yanlış Pozitif Değeri	Ortalama Kesinlik
1-NN	0,793	0,207	0,793
3-NN	0,749	0,252	0,749
5-NN	0,724	0,276	0,725
10-NN	0,686	0,314	0,691

Bu sonuçlar incelendiğinde en yüksek doğru pozitif değerinin 1-NN en yakın komşuluk değeri ile elde edildiği görülmektedir. Tablo 1 incelendiğinde 1, 3, 5 ve 10 en yakın komşuluk değerlerine göre doğru pozitif değerleri komşuluk değeri arttıkça azalmakta iken yanlış pozitif oranları ise yükselmektedir. Literatürde de komşuluk değeri arttıkça doğru pozitif değerlerinin azaldığı çalışmalar mevcuttur [7, 14].

Tablo 2'de en iyi doğru pozitif değerini veren 1 en yakın komşuluk değerine göre sonuçlar verilmiştir. Sonuçlar incelendiğinde en yüksek doğru pozitif değeri %86.9 olurken en kötü doğru pozitif değeri %64.2 olarak gerçekleşmiştir.

Ayrıca Tablo 2 incelendiğinde yanlış pozitif değeri olarak en düşük %13.1 elde edilirken, en kötü sonuç %35.8 olarak gerçekleşmiştir.

Tablo 2: 1, 3, 5 Ve 10 Değerlerine Göre Ortalama Sonuçlar

1-NN			
Sıra No	Doğru Pozitif	Yanlış Pozitif	Kesinlik
1	0,812	0,188	0,815
2	0,842	0,158	0,842
3	0,818	0,182	0,819
4	0,777	0,223	0,777
5	0,778	0,222	0,779
6	0,642	0,358	0,642
7	0,807	0,193	0,807
8	0,736	0,264	0,739
9	0,837	0,163	0,839
10	0,724	0,276	0,724
11	0,753	0,247	0,754
12	0,869	0,131	0,869
13	0,805	0,195	0,806
14	0,804	0,196	0,804
15	0,820	0,180	0,820
16	0,817	0,183	0,817
17	0,833	0,167	0,834
Ortalama	0,793	0,207	0,793

Bu sonuçlara göre toplama-çıkartma işlem grubu ile çarpma bölme işlem gurubu ortalama %79.3 oranında doğru pozitif, %20.7 yanlış pozitif değerleriyle ile ayrıt edilebilmektedir.

4. TARTIŞMA

Daha önce yapmış olduğumuz çalışmada sayısal ve sözel işlemler sırasındaki EEG sinyallerinin sınıflandırılması %89.1 doğru pozitif oranı ile gerçekleşmiştir [3].

Bu çalışma ile sayısal işlemlerden toplama-çıkarma ve çarpma bölme gruplarının kendi arasında %79.3 doğru pozitif oranı ile sınıflandırılabilirliği gösterilmiştir.

Bu çalışma daha sonra yapılması düşünülen kolay ve zor metinlerin EEG sinyalleri ile belirlenmesi çalışmasına ışık tutmaktadır.

5. KAYNAKÇA

[1] Yazgan, E., ve Korürek, M., Tıp Elektronik. İTÜ Matbaası, 1996.

[2] Xizheng, Z., Ling, Y. and Weixiong, W., “Wavelet Time-frequency Analysis of Electro-encephalogram (EEG) Processing”, International Journal of Advanced Computer Science and Applications, 2010, 1-5.

[3] Eraldemir, S.G., ve Yıldırım, E., “EEG Verilerinden Sayısal Ve Sözel Düşüncenin Sınıflandırılması”, IEEE 22. Sinyal İşleme ve İletişim Uygulamaları Kurultayı, 2014,1-4.

[4] Batar, H., EEG İşaretlerinin Dalgacık Analiz Yöntemleri Kullanılarak Yapay Sinir Ağları İle Sınıflandırılması, Yüksek Lisans Tezi, Kahramanmaraş Üniversitesi, Fen Bilimleri Enstitüsü, 2005.

[5] Federico Ciret Galán, Carole R. Beal, “EEG Estimates of Engagement and Cognitive Workload Predict Math Problem Solving Outcomes”, *User Modeling, Adaptation, and Personalization(UMAP)*, 2012, pp. 51-62.

[6] Sakkalis, V., Zervakis, M., and Micheloyannis, S., “Significant EEG Features Involved in Mathematical Reasoning: Evidence from Wavelet Analysis”, *Brain Topography, Volume 19*, Numbers 1/2, pp. 53-60.

[7] Eraldemir, S.G., EEG Verileri Kullanılarak Metinsel Okuma Ve Matematiksel İşlemlerin Analizi. Yüksek Lisans Tezi, Mustafa Kemal Üniversitesi, Fen Bilimleri Enstitüsü, 2014.

[8] Übeyli, E.D. ve Güler, İ., “Dalgacık Dönüşümü İle EEG İşaretlerinden Çıkarılan Öznitelik Vektörleri Üzerinde İstatistiksel İşlemlerin Gerçekleştirilmesi”, *Eleco 2004*,s. 230-234.

[9] Strang, G., Nguyen, T., Wavelets and Filter Banks, Wellesley-Cambridge Press, 1997.

[10] Qu , H. & Gotman, J. A, “Patient-Specific Algorithm for the Detection of Seizure Onset in Long-Term EEG Monitoring: Possible Use as a Warning Device”, *IEEE Transactions on Biomedical Engineering*, 44, 115-122p, 1997.

[11] Oliveria, I., Grigori, O., Guimaraes, N., “EEG Signal Analysis for Silent Visual Reading Classification”, *International Journal Of Circuits, Systems And Signal Processing*, Issue 3, Volume 3, 2009:119-126p.

[12] Murugappan, M., “Human emotion classification using wavelet transform and K-NN”, *Pattern Analysis and Intelligent Robotics (ICPAIR) International Conference on*, (Volume:1), 28-29 June 2011, 148 – 153p.

[13] Mitchell, T., 1997, Machine Learning, McGraw Hill, 1997.

[14] Özer, Z.B., ve Amasyalı, M.F., “Meta Öğrenme ile K-NN Parametre Seçimi”, *21. Sinyal İşleme Ve Uygulamaları Konferansı*, 2013.

[15] Parvinnia, E., Sabeti, M.,Zolghadri Jahromi, M.,Boostani, R., “Classification Of Eeg Signals Using Adaptive Weighted Distance Nearest Neighbor Algorithm”, *Journal of King Saud University - Computer and Information Sciences*, Volume 26, Issue 1, January 2014, 1-6p

[16] Eraldemir, S.G., Yıldırım, E., Kutlu, Y. ve Yıldırım, S., “Kognitif EEG İşaretleri İçin Öznitelik Seçimi Tabanlı Kanal Seçimi ve Sınıflandırma”, *Akıllı Sistemlerde Yenilikler ve Uygulamaları 2014 (ASYU 2014)*, 1-6.