

Komşuluk Özellik Çıkarımı Yöntemiyle EKG Sinyallerinin Sınıflandırılması

ECG Signals Classification with Neighborhood Feature Extraction Method

Çiğdem Bakır¹

¹ Bilgisayar Mühendisliği
Yıldız Teknik Üniversitesi
cigdem@ce.yildiz.edu.tr

Özetçe

Bu çalışmada doğrusal olmayan boyut indirgeme methodları EKG sinyallerine uygulanmıştır ve bu boyut indirgeme tekniklerinin EKG sinyallerinin sınıflandırılmasında ve bölütlenmesindeki başarısı ele alınmıştır. Ayrıca EKG sinyallerinin uzanımsal birleşimi dikkate alınarak yüksek boyutlu uzaydan düşük boyutlu uzaya geçilerek verilerin yeni bir yöntem olan komşuluk özellik çıkarımı (KÖÇ) yöntemiyle bölütlenmesi sağlanmıştır. Uzanımsal birleşim yoluyla yapılan ve yeni bir yöntem olarak geliştirilen KÖÇ algoritmasının sınıflandırma sonuçları tek EKG örneği alınarak yapılan boyut indirgeme ile elde edilen EKG sinyallerinin sınıflandırılma sonuçları karşılaştırılmıştır. Komşu EKG örneklerine bakılarak yapılan KÖÇ boyut indirgeme tekniğinin sonuçları EKG sinyallerinde bölütleme üzerindeki etkisinin avantajları deneysel sonuçlar kısmında sunulmuştur ve önerilen yöntemin başarısı gösterilmiştir. Yapılan çalışmayla elde edilen sonuçlar ileride yapılacak çalışmalar için de ümit vaad etmektedir.

Anahtar Kelimeler — komşuluk özellik çıkarımı; sınıflandırma; EKG; uzanımsal birleşim.

Abstract

In this study, non-linear dimension reduction methods were applied to ECG signals and success of such dimension reduction techniques for the classification and segmentation of ECG signals were discussed. Also, segmentation of data through neighbourhood feature extraction (NFE) method were enabled by transiting from high dimensioned space to low dimension space by considering the longitudinal combination of ECG signals. Results classification results of NFE algorithm performed through longitudinal combination and as a newly developed method were compared with classification results of ECG signals obtained through dimension reduction by taking one ECG instance. Results of NFE dimension reduction technique performed by considering the neighbour ECG instances, advantage of effect on segmentation of ECG signals were presented at empirical results section and the success of suggested method was indicated. Results

obtained by performed study are promising for the studies to be conducted in further period.

Keywords — neighborhood feature extraction; classification; ECG; spatial combination.

1. Giriş

İnsan vücudunda kalbin nasıl çalıştığını ve kalp hareketlerini gösteren elektriksel işaretlere elektrokardiyogram (EKG) denir. EKG sinyallerine bakılarak kalbin ritm bozukluklarına bağlı olarak gelişen kalp hastalıkları, kardiyolojik bozukluklar, kalpte meydana gelen anormallikler belirlenir. Bu sebeple EKG sinyallerinin doğru bir şekilde analiz edilmesi, yorumlanması günümüzde oldukça önemlidir[1].

Günümüzde birçok insan kalp hastalıklarına bağlı olarak hayatını kaybetmektedir. Bundan dolayı kalp sağlığına önem verilmesi gereklidir. EKG kalpte meydana gelip tüm vücuda yayılan elektriksel işaretlerde meydana gelen değişiklikleri kaydederek kalpte meydana gelen bozuklukları belirlemeye çalışır. EKG kısacası kalp atışını gösterir ve sinyal işlemede yaygın bir şekilde kullanılmaktadır[2].

EKG sinyallerinin sınıflandırılması ile ilgili birçok çalışma yapılmıştır. Ancak yapılan çalışmalarda EKG sinyallerinin hem hızlı bir şekilde anlaşılması hem de doğru yorumlanması tam olarak sağlanmamıştır. Bu çalışmayla birlikte geliştirilen yeni yöntem tüm boyut indirgeme tekniklerine başarılı bir şekilde uygulanmıştır. Ayrıca EKG sinyallerinin sınıflandırma başarısı artırılmıştır ve diğer biyomedikal alanlarda da yapılacak olan çalışmalara yol gösterilmeye çalışılmıştır.

Yapılan çalışmada ilk olarak boyut indirgeme için bant seçimi ve özellik çıkarımı yapılmıştır. Bu çalışmada EKG verilerinde özellik çıkarımından yararlanılarak eğitici yöntemlerle sınıf etiketleri kullanılarak sınıf ayırımının yapılmasına çalışılmıştır. Özellik çıkarımının yapılması için birçok yöntem mevcuttur. Bu yöntemlerden en eski ve en bilineni doğrusal bir teknik olan PCA dir[3]. Ancak doğrusal teknikler bilgi çıkarımında yeterli değildir. Karmaşık veriyi sınıflandırmak için doğrusal olmayan tekniklere ihtiyaç duyulur. Son zamanlarda bu amaçla birçok nonlinear teknik kullanılmıştır. Çalışmamızda mevcut veri setimiz için Kernel PCA, Isomap, gibi global özellikleri koruyan doğrusal olmayan özellik çıkarımı tekniklerin yanı sıra; Laplacian Eigenmaps gibi lokal

Sinyal İşleme 3

2. Gün / 16 Ekim 2015, Cuma

özellikleri koruyan no linear teknikler de kullanılmıştır. Ayrıca doğrusal olmayan tekniklerle sınıflandırma başarısını karşılaştırmak için PCA ve LDA gibi doğrusal özellik çıkarımı teknikleri de kullanılmıştır[4].

İkinci aşama olarak ise bu özellik çıkarımı yöntemleriyle oluşturduğumuz veri setimiz için 5, 10, 15, 20 ve 50 uzanımsal komşuluklarını ele alıp EKG verilerini sınıflandırmaya çalıştık [5]. EKG verisinde her bir veri bir örnek nokta olarak ifade edilmiştir. Bu örnek nokta EKG örneği olarak ifade edilmiştir. Bu imge elamanları için oluşturulan vektörler bütün bandlar için geçerlidir ve belirli bir değere sahiptir. Eğer bir imge elamanının sınıfı belirli ise onun hemen yanındaki komşusunun sınıfı da kendisiyle aynıdır ve bu komşu EKG örneği uzanımsal komşuluğu göstermiştir. Bu uzanımsal komşuluk EKG sinyalindeki bir örnek için onun etrafındaki $n \times n$ komşusu alınıp EKG sinyali eğitilmiştir. İki nokta arasındaki uzanımsal komşuluğun yakınlığı öklid uzaklığı ile ölçülmüştür. Uzanımsal komşuluklarında birbirine yakın iki veri noktası benzerdir ve aynı sınıftan olma olasılığı yüksektir. $n \times n$ komşusu için uzaklık arttıkça benzerlik azalır.

2. Yöntem

Gerçek dünyada büyük miktarda verilerle uğraşmaktadır. Bu sebeple verileri doğru sınıflandırmak için çok fazla hafızaya, hesaplama ihtiyacı duyulur. Verinin indirgenmesi ile veri etkin bir şekilde kullanılır ve maliyet azaltılır. Bu işlemi veri kaybı olmadan yapabilmek için boyut indirgemeyi sağlayabilecek, bant sayısını azaltabilecek boyut indirgeme methodlarına ihtiyaç duyulur. Özellik çıkarımı yüksek boyutlu verilerin işlenmesinde çok önemli bir adımı oluşturur. Özellik çıkarımı bantlar üzerinde doğrusal ya da doğrusal olmayan yöntemlerle yapılır. Bu çalışmada sınıf bilgileri kullanılarak eğitici çalışma yapabilmek, sınıflandırma başarısını artırıp işlem maliyetini azaltmak için ilk olarak özellik çıkarımı yapılmıştır.

Yapılan çalışmada öncelikle tüm verinin işlenmesi için belirli özellikleri sağlayan imzalar alınmıştır. Bu spektral imzalar spektral dağılımı olabildiğince en iyi temsil eden spektral imzalar olarak belirlenmiştir. Doğrusal olmayan özellik çıkarımı teknikler tüm veri yerine bu spektral imzalara uygulanmıştır. Çalışmada kullanılan boyut azaltma teknikleriyle mevcut veri seti 5 ve 10 boyutlu özellik vektörleri çıkarılarak elde edilmiştir. Ayrıca tüm veri setini eğitebilmek için RBF ve kNN interpolasyonu kullanılmıştır. RBF ile indirgenmiş veri seti yapay sinir ağları ile öğrenilmiş ve her spektral imza bir ağ yapısı kullanılarak düşük boyutlu uzaya geçilmiştir. RBF nöronlar arasındaki bağlantılarla oluşur. Eğitim aşamasında girdi ve çıktılara bakılarak nöronların ağırlıkları güncellenir. Hata oranı minimum oluncaya kadar bu işlem devam eder[6]. kNN interpolasyon yönteminde ise k en yakın örneklerine bağlı olarak bulunmuştur. kNN interpolasyonunda normalize edilmiş verinin en yakın komşuluğu bulunarak ağırlıklar hesaplanmıştır[7]. Bu yöntemde EKG verisindeki her örneğin spektral imzaları ve uzaklıkları bulunmuştur. Daha sonra $k=9$ değeri seçilerek boyutu indirgenecek olan EKG örneklerine en yakın k spektral imzanın doğrusal olmayan boyut indirgeme yöntemleri bulunup özellik vektörleri elde edilmiştir. Buradaki amaç; doğrusal olmayan izdüşüm yöntemleri ile elde edilecek en yakın ara değerlendirmeyi bulmaktır. Bu yöntem tüm spektral imzalar üzerine uygulanarak boyut indirgenmiştir.

2.1. Temel Bileşen Analizi (Principal Component Analysis-PCA)

PCA en popüler ortogonal linear dönüşümdür. PCA en büyük varyansa sahip olan verinin düşük boyutlu uzayda gösterimidir. Yüksek varyanslı özellikler düşük varyansa tercih edilir. X veri matrisinin örneklerinin kovaryans matrisini hesaplayarak maliyet fonksiyonunu arttıran M linear haritalamayı bulur. En büyük özdeğere ait özvektörleri çıkarır.PCA x_i ve x_j veri noktaları arasında öklid uzaklığı kullanır .PCA dönüşümü; $\mu^T=X^TW$ şeklindedir. W ortogonal matris, μ^T linear dönüşümü, W ise kovaryans matrisine denk gelen özvektörleri gösterir[8]. PCA verinin doğrusal bir biçimde gösterimidir. Veri PCA ile eşitlik 1'deki gibi ayrıştırılır[8].

$$x_i = \sum_{j=1}^p w_{ij} Q_j \quad (1)$$

2.2. Doğrusal Ayrıştırma Analizi (Linear Discriminant Analysis-LDA)

PCA en az sayıda bileşenlerle veriyi yüksek çeşitlilikteki ifade eder. Ancak bu bileşenler farklı sınıflandırmak ve olası en yüksek ayrıştırmanın sağlanması için en iyi çeşitliliğe ihtiyaç duymaz. LDA ile birlikte verinin doğru sınıflandırılması için verinin farklı sınıfları arasında en yüksek ayrıştırma sağlar. Bu ayrıştırmayı sağlama için gerekli boyut indirgemeyi kovaryans matrisi hesaplayarak bulur. Sınıflar içindeki kovaryans matris (S_W) ve sınıflar arasındaki kovaryans matristen (S_B) yararlanılarak toplam kovaryans matris hesaplanılır[1].

$$S_T = S_W + S_B \quad (2)$$

İkinci aşamada ise LDA katsayılarının elde edilmesi için projeksiyon matrisi hesaplanılır.

2.3. Çekirdek Temel Bileşen Analizi (Kernel Principal Component Analysis-KPCA)

KPCA PCA methodunun genişletilmiş halidir. Ancak PCA doğrusal bir method iken; KPCA ise doğrusal teknikleri geliştiren bir doğrusal olmayan tekniktir. Kernel matrisi kullanılarak verinin boyutu indirgenir. $k_{ij}=K(x_i, x_j)$ şeklinde x_i ve x_j veri noktalarını kullanarak K kernel matrisi hesaplanılır. K kernel matrisi girişlerinin değiştirilmesiyle merkezlerin ve d özvektörleri bulunur. Böylelikle KPCA ile birlikte veri PCA boyut indirgeme tekniğine göre daha yüksek boyutlu uzaya geçirilmiş olur.KPCA'da önemli olan kernel fonksiyonunu seçmektir. Kernel fonksiyonu linear kernel, Gauss kernel ve polinomial kernel olabilir. Literatürde yapılan çalışmalarda KPCA yüz tanıma, konuşma tanımadada oldukça başarılı sonuçlar sağlamıştır[9,10].

2.4. Isomap

Isomap ağırlıklı grafik üzerinde çok boyutlu ölçekleme algoritmalarında kullanılan Öklid uzaklığı yerine geodesic

Sinyal İşleme 3

2. Gün / 16 Ekim 2015, Cuma

uzaklığı kullanan bir düşük boyutlu embedding methodtur. Gedesic uzaklık en kısa uzaklıktır ve bu uzaklığı bulmak için tüm veri noktaları arasında komşulukların bulunması gerekir. Geodesic uzaklığı veri noktaları x_i, x_j ($i=1,2,\dots,n$) arasında komşu grafiğiyle hesaplanır. Ayrıca Isomap'te komşuluk parametresinin seçimi de önemlidir Her veri noktasındaki bağlantı yüksek boyutlu uzayda en yakın öklid uzaklığı olarak bilinir. İki nokta arasında en kısa mesafe Dijkstra algoritması ile bulunur[11].

2.5. Laplacian Eigenmaps

Verinin manifold lokal özelliklerini koruyarak düşük boyutlu uzaya geçmesini sağlar. Laplacian methodu ilk önce G grafiğini k en yakın komşuluğuna bağlantılı olarak oluşturur ve ağırlığı seçer. x_i ve x_j veri noktaları için kenar ağırlıkları Gauss ile hesaplanır. y_i düşük boyutlu uzayda gösterimindeki maliyet fonksiyonunda w_i ağırlıkları x_i ve x_j veri noktaları arasındaki küçük uzaklıklara bağlıdır. Böylelikle y_i ve y_j spektral grafik teorisyle maliyet fonksiyonunu en aza indirger. M derece matrisi, W dianogal matris olmak üzere L laplacian grafiği $L=M-W$ şeklinde ifade edilir. Laplacian grafiğinin öz ayrıştırması yapılır ve düşük boyutlu embedding oluşturulur[12].

3. Sonuçlar

Bu çalışmada PTB (Physikalish-Technische Bundesanstalt) EKG veriseti kullanılmıştır. Bu veriseti Almanya Ulusal Metroloji Enstitüsü tarafından 289 kişiden 549 EKG sinyali alınarak oluşturulmuştur. Hastalara kalp hastalıklarına bağlı olarak oluşan 9 farklı tanı konulmuştur. Ayrıca herbir kayıt eş zamanlı olarak ölçülen 15 sinyali içerir. Herbir sinyal yaklaşık olarak bir saniyede 1000 örnekle sayısallaştırılmıştır[13].

Tablo 1'de 5 ve 10 boyut için RBF ile yapılmış sınıflandırma doğruluğunun sonuçları gösterilirken; Tablo 2'de ise 5 ve 10 boyut $k=9$ değeri için kNN interpolasyonu yapılmış sonuçları gösterilmiştir. Tablo 3'de ise 5 boyut için uzanımsal birleşim sağlanılarak önerilen KÖÇ tekniğinin RBF ile yapılan sınıflandırma doğruluğunun sonuçları, Tablo 4'de ise 10 boyut için yapılmış sınıflandırma doğruluğunun sonuçları gösterilmiştir. Tablo 5'de ise 5 boyut için uzanımsal birleşim sağlanılarak önerilen KÖÇ tekniğinin Knn interpolasyonu yapılan sınıflandırmasının sonuçları, Tablo 6'da ise 10 boyut için yapılmış sınıflandırma doğruluğunun sonuçları gösterilmiştir. Tablo 1 ve Tablo 2'de yapılan çalışmayla Tablo 3, 4, 5, 6'de önerilen yöntemin sonuçlarına bakıldığında önerilen yöntemin sınıflandırma doğruluğunun başarısını büyük ölçüde artırdığı gözlemlenir. Özellikle komşuluk değeri 15 en yüksek sınıflandırma başarısına sahiptir. Ancak komşuluk değerinin büyük ölçüde artması sınıflandırma başarısını düşürür.

Bu çalışmada uzanımsal birleşim sağlanılarak boyut indirgemenin sınıflandırma üzerindeki avantajları sunulmuştur. KÖÇ tekniğinin EKG sinyallerinin sınıflandırması üzerindeki başarısını her iki sınıflandırma algoritmasında da sınıflandırma doğruluğunun %10 dan fazla arttırdığı görülmüştür.

Tablo 1: 5 ve 10 boyut için RBF sınıflandırma algoritmasının sonuçları

Boyut İndirgeme Metotları	RBF	
	5 boyut	10 boyut
PCA	72.65	73.84
LDA	68.66	65.37
KPCA	78.13	80.31
Isomap	79.04	79.56
Laplacian E.	80.25	82.03

Tablo 2: 5 ve 10 boyut için kNN sınıflandırma algoritmasının sonuçları

Boyut İndirgeme Metotları	kNN(k=9 için)	
	5 boyut	10 boyut
PCA	75.83	75.25
LDA	65.72	69.84
KPCA	84.35	85.04
Isomap	85.67	83.39
Laplacian E.	86.69	88.34

Tablo 3: 5 Boyut için KÖÇ ile elde edilen verilerin RBF ile sınıflandırılma sonuçları

Boyut İndirgeme Metotları	Komşuluk Değerleri				
	5	10	15	20	50
PCA	75.87	85.34	87.62	85.34	64.37
LDA	70.12	75.82	79.87	80.12	60.72
KPCA	82.88	90.12	92.11	87.64	70.37
Isomap	83.17	88.72	95.64	87.62	71.64
Laplacian E.	84.62	90.62	97.83	94.62	74.83

Yapılan bu çalışmayla önerilen KÖÇ tekniğinin EKG sinyallerinin sınıflandırılma doğruluğu üzerindeki başarısı gösterilmiştir. İlerde yapılacak çalışmalarda KÖÇ tekniğinin geliştirilerek farklı uygulama alanlarında da uygulanması sağlanacaktır.

Sinyal İşleme 3

2. Gün / 16 Ekim 2015, Cuma

Tablo 4: 10 Boyut için KÖÇ ile elde edilen verilerin RBF ile sınıflandırılma sonuçları

Boyut İndirgeme Metotları	Komşuluk Değerleri				
	5	10	15	20	50
PCA	75.01	77.83	80.64	78.62	64.33
LDA	71.64	79.65	74.62	73.13	60.06
KPCA	85.37	87.83	91.62	85.12	73.62
Isomap	82.64	89.67	91.24	83.64	77.37
Laplacian E.	84.65	90.83	95.62	92.47	78.12

Tablo 5: 5 Boyut için KÖÇ ile elde edilen verilerin kNN ile sınıflandırılma sonuçları

Boyut İndirgeme Metotları	Komşuluk Değerleri (k=9 için)				
	5	10	15	20	50
PCA	80.62	87.64	89.63	82.52	70.03
LDA	70.62	73.52	75.64	73.27	64.87
KPCA	87.64	92.02	96.75	91.54	71.64
Isomap	89.64	93.52	98.64	97.62	78.54
Laplacian E.	89.12	95.52	99.14	96.82	71.87

Tablo 6: 10 Boyut için KÖÇ ile elde edilen verilerin kNN ile sınıflandırılma sonuçları

Boyut İndirgeme Metotları	Komşuluk Değerleri (k=9 için)				
	5	10	15	20	50
PCA	78.62	83.64	85.72	84.17	65.67
LDA	74.44	76.62	79.87	75.67	61.12
KPCA	90.67	94.86	97.61	95.62	76.15
Isomap	84.61	88.63	90.82	87.46	71.04
Laplacian E.	90.26	95.65	98.79	87.54	78.22

4. Kaynakça

[1] Kiranyaz S., Ince T. etc, "Personalized long-term ECG classification: A systematic approach", *Expert Systems with Applications*, vol.38, issue 4, pp.3220-3226, 2011.

[2] Erdoğan P., Peşçaker A., "Dalgacık Dönüşümü ile EKG Sinyallerinin Özellik Çıkarımı ve Yapay Sinir Ağları ile Sınıflandırılması", *5.Uluslararası İleri Teknolojiler Sempozyumu (IATS'09)*, 2009.

[3] Martis R., Acharya R. And Min L., "ECG beat classification using PCA, LDA, ICA and Discrete Wavelet Transform", *Elsevier Biomedical Signal Processing and Control*, vol.8, issue 5, pp.437-448, 2013.

[4] Ceylan R.,Ozbay Y., "Comparison of FCM, PCA and WT techniques for classification ECG arrhythmias using neural network", *Expert Systems with Applications*, vol.33, issue 2, pp.286-295, 2007.

[5] Bakir C., "Nonlinear Feature Extraction for Hyperspectral Images", *International Conference on Advanced Technology & Sciences (ICAT'14)*, pp.945-949, 2014.

[6] Silipo R., Bortolan G. And Marchesi C., "Design of hybrid architectures based on neural RBF pre-processing for ECG analysis", *International Journal of Approximate Reasoning*, pp.177-196, 1999.

[7] Castillo O., Melin P. ETC, "Hybrid intelligent system for cardiac arrhythmia classification with Fuzzy K-Nearest Neighbors and neural networks combined with a fuzzy system", *Expert Systems with Applications*, vol.39, issue 3, pp.2947-2955, 2012.

[8] Maglaveras N., Stamkopolous T. etc, "ECG pattern recognition and classification using nonlinear transformations and neural networks:A review", *Medical Informatics*, vol.52, issue 1-3, pp.191-208, 1998.

[9] Wen Y., He L. And Shi P., "Face recognition using difference vector plus KPCA", *Digital Signal Processing*, vol.22, issue 1, pp.140-146, 2012.

[10] Widjaja D., Varon C.etc, "Application of Kernel Principal Component Analysis for Single-Lead-ECG-Derived Respiration", *IEEE Transactions on Biomedical Engineering*, vol.59, no.4, pp.1169-1176, 2012.

[11] Shalbaf A., Alizadefani Z. And Behram H. "Echocardiography without electrocardiogram using nonlinear dimensionality reduction methods", *J.Med Ultrasonics*, vol.42, pp.137-149, 2015.

[12] Perry T., Zha H. etc, "Supervised Laplacian Eigenmaps with Applications in Clinical Diagnostics for Pediatric Cardiology", *Computer Science & Learning*, 2012.

[13] Vozda M., Cerny M., "Methods for derivation of orthogonal leads from 12-lead electrocardiogram:A review", *Biomedical Signal Processing and Control*, vol.19, pp.23-34, 2015.