

EKG Sinyallerinin Kalitesinin Destek Vektör Makinaları ve İkili Karar Ağaçları Tabanlı Olarak Değerlendirilmesi

Quality Assessment of ECG Signals Based on Support Vector Machines and Binary Decision Trees

Berken Utku Demirel
Middle East Technical University
Ankara, Turkey
berken.demirel@metu.edu.tr

Yeşim Serinağaoğlu
Middle East Technical University
Ankara, Turkey
yserin@metu.edu.tr

Özetçe —Elektrokardiyografi (EKG) sinyallerinin doğru bir şekilde değerlendirilmesi, kalp hastalıklarının otomatik teşhisi için çok önemlidir. Özellikle çeşitli gürültü kaynaklarıyla güvenilir bir sonuç almaya imkan vermeyecek kadar bozulmuş sinyallerin analizlerinde başarı düşmektedir. Bu nedenle, otomatik teşhis aşamasına geçmeden önce EKG sinyal özelliklerinden, makine öğrenme sınıflandırıcılarına veya sezgisel karar kurallarına dayalı birçok sinyal kalite değerlendirme (signal quality assessment (SQA)) tekniği önerilmiştir. Bu çalışmada da, EKG sinyallerinde harekete bağlı yapay sinyal ve gürültünün çevrimdışı tespiti için doğruluğu yüksek bir algoritma sunulmaktadır. Sinyalleri derecelendirme yaklaşımımız iki aşamadan oluşmaktadır. İlk aşama, EKG'den bazı güvenilir özelliklerin çıkarılmasını içermektedir. Yaklaşımımızın ikinci aşamasında sinyalleri derecelendirmek için Destek Vektör Makineleri ve İkili Karar Ağaçları kullanılmıştır. İkinci aşamada, sinyalleri derecelendirmek için makine öğrenme teknikleri kullanılmıştır. Önerilen yöntem, FizyoNet / Computing in Cardiology Challenge 2011 veri tabanında iki farklı sınıflandırıcıyı karşılaştırarak test edilmiştir. DVM (Destek Vektör Makineleri), etiketli verilerde yaklaşık en iyi sınıflandırma doğruluklarını (94%) sağlamıştır. (Etiketler EKG sinyallerinin klinik yorum için kabul edilebilir veya kabul edilmez olduğunu belirtmektedir).

Anahtar Kelimeler—EKG, Sinyal kalite değerlendirmesi, Destek Vektör Makineleri, RMS

Abstract—Quality assessment of ECG signals has crucial importance for the automatic diagnosis of heart diseases, especially for signals which are heavily contaminated with several artifacts. Therefore, several SQA (signal quality assessment) techniques were presented based on ECG signal features and the machine learning classifiers or heuristic decision rules. This study presents an algorithm for accurate and offline detection of motion artifacts and noise in ECG signals. Our grading algorithm involves two stages. The first stage involves extracting several non-fiducial features from ECG. The second stage of our approach uses Support Vector Machines and Binary Decision Trees to grade signals. The proposed method was tested using the PhysioNet/Computing in Cardiology Challenge 2011 Database by comparing two different classifiers. The SVM provided the best classification accuracies of nearly 94% on the labeled data. (Labels which indicate the ECG signal is acceptable or not for clinical interpretation).

Keywords—ECG, Signal quality assessment, SVM, RMS

I. GİRİŞ

Elektrokardiyografi (EKG) sinyalleri, kardiyovasküler hastalık tanısı, aritmi, fizyolojik geri bildirim, uyku apnesi tespiti, kronik hasta gözetimi, ani kalp durması tahmini, biyometrik, duygusal ve fiziksel aktivite tanıma sistemleri gibi birçok sağlık sorunu için yaygın olarak kullanılmaktadır [1]. Bu EKG uygulama sistemlerinin doğru analizi için yüksek kalitede sinyaller temel gereksinimdir. Çünkü birçok algoritma için ilk adım olan temel özellikleri belirlemek ve bu temel özellikleri doğru şekilde işlemek, çoğunlukla EKG'nin kalitesine bağlıdır. Bu nedenle, özellikle uzun kayıtlarda, EKG sinyallerinin sinyal kalitesi değerlendirmesi (SQA) çok önemlidir ve bu önem, mHealth uygulamalarının gün geçtikçe yaygınlaşması ile artmaktadır.

Mevcut son teknoloji SQA yöntemleri iki gruba ayrılabilir. İlk kategoride EKG sinyallerinin kalite değerlendirmesi için referans (fiducial) (QRS kompleksi, P, Q, R, S, T morfolojileri, vb) özellikleri kullanılır [2-3]. İkinci kategoride ise sinyalleri derecelendirmek için EKG'nin referans olmayan (non-fiducial) özellikleri kullanılır [4-6]. Referans özellikleri otomatik olarak tespit etmek, özellikle mobil EKG kayıt cihazlarının eğitimsiz personel ya da hastanın kendisi tarafından kullanılmasına uygun geliştirildiği dikkate alındığında ve patolojik durumlarda, zorluklar içermektedir. Bu nedenle EKG'nin referans olmayan özelliklerini kullanarak referans bulma adımını ortadan kaldıran, bununla birlikte SQA başarımını iyileştiren EKG özellik tanımlarına ve sınıflandırma algoritmalarına halen ihtiyaç duyulmaktadır. Özellikle mHealth uygulamalarında, seçilen algoritmaların gerçek zamanlı uygulanabilir olması önemlidir. Bunun için de SQA için kullanılacak özelliklerin basitçe tanımlanabilir olması avantaj sağlar. Bir başka önemli unsur da kullanılacak sınırlandırma yöntemidir. Yüksek doğruluk oranlarında SQA değerlendirmesi, EKG'den asıl çıkarılmak istenen aritmi tanımlaması başarımını da artıracaktır.

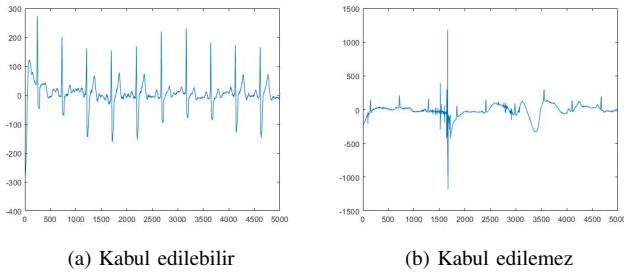
Bu çalışmada, referans özellik bulmadan da basit bir işlemle elde edilebilen ortalama karekök (root mean square (RMS)) değerine bağlı bir SQA yöntemi önerilmektedir. İncelenecek uzun kayıtlarda 10 s'lik pencerelerde her lead için

RMS hesaplanmaktadır. Bu RMS değerlerinden ortalama ve standart sapma olmak üzere iki özellik türetilmektedir. Ardından bu özellikler, Destek Vektör Makineleri (SVM) ve Karar Ağaçları (DT) olmak üzere iki farklı sınıflandırma yöntemine verilerek ilgili 10 s'lik sinyaller 'kabul edilebilir' veya 'kabul edilemez' olarak sınıflandırılmaktadır.

II. YÖNTEMLER

A. Veri setleri

Bu çalışma için PhysioNet / Kardiyoloji Bilgisayar Bilişim (CinC) Challenge 2011'in EKG kayıtları kullanılmıştır. Bu veri set-a ve set-b kümelerinden oluşmaktadır. Set-b için sinyal işaretlemeleri kamuya açıklanmadığından algoritmamızı eğitmek ve test etmek için set-a'yı kullandık. Eğitim verileri 998 kayıt arasından rastgele seçilmiştir. PhysioNet / CinC set-a veri tabanı, 16-bit çözünürlükte 500 Hz'de örneklenmiş, tam teşhis bant genişliğine sahip (10 - 100Hz) 10 sn standart 12 kanallı EKG kayıtlarını içermektedir. Eğitim veri seti, derecelendirme için uzman bir grup tarafından birbirlerinden haberdar olmaksızın etiketlenmiştir. Verilerin üçte birinden azı 'kabul edilemez' olarak sınıflandırılmıştır. Şekil 1 SQA sonucunda 'kabul edilebilir' ((a) paneli) ve 'kabul edilemez' ((b) paneli) sinyal örneklerini göstermektedir.



Şekil 1: Kabul edilebilir ve edilemez sinyallerin ilk kanallarının örnekleri

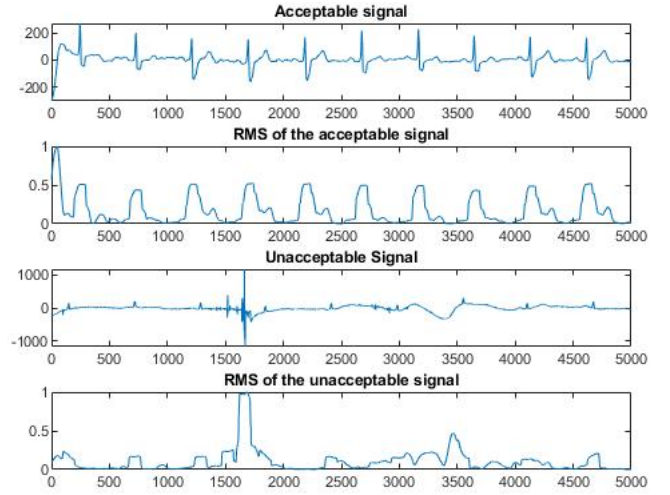
B. Moving Root Mean Square

Her kanalda ayrı ayrı 0,7 - 100 Hz kesmeli üçüncü dereceden bir Butterworth bant geçiren filtre uygulanmıştır. Daha sonra, hareketli RMS değeri, sinyalin her bir kanalı üzerinde kayan bir pencere kullanılarak hesaplanmıştır:

$$X_{RMS} = \sqrt{\frac{1}{N} \sum_{i=1}^N |x_n|^2}, \quad (1)$$

Burada N pencere uzunluğu ve x_n o pencereadaki zaman dizisidir. Sinyalin uzunluğu örnekleme oranının onda biri olarak belirlenmiştir. Şekil 2'de 'kabul edilebilir' ve 'kabul edilemez' sinyal örnekleri ile, bu sinyallerden elde edilen RMS değerleri sunulmaktadır. Büyük bir genlik kayması veya sinyalde cihaz doygunluğu olduğunda, Şekil 2'de de görülebildiği gibi RMS artar.

Sinyalin RMS değerleri 0 ve 1 arasına normalize edildikten sonra, her pencere için ortalama ve standart sapma değerleri elde edilmiştir.



Şekil 2: Kabul edilebilir ve kabul edilemez sinyal örnekleri ve bunlara karşılık gelen RMS değerleri.

C. SQA için makine öğrenmesi

Bir önceki bölümde sunulan normalize-RMS'ten elde edilmiş ortalama ve standart sapma özellikleri SVM ve DT olmak üzere iki farklı sınıflandırma yöntemine verilmiştir. SVM ve DT yöntemlerini eğitmek için CinC Challenge veri seti için etiketlenmiş verilerden rastgele 30 kayıt seçilmiştir. Bu veriler 12 kanalın 10'ar saniyelik kayıtlarından oluşmaktadır. Sınıflandırıcıların eğitim başarımını değerlendirmek için 5-kat çapraz doğrulama kullanılmıştır.

SVM için aşağıdaki denklem kullanılmıştır [3]

$$k(x_n, x_m) = \exp\left(-\frac{\|x_n - x_m\|^2}{2\sigma^2}\right) \quad (2)$$

Burada x_n ve x_m , ilk özellik alanında ifade edilen iki vektördür. Gauss çekirdeğine sahip SVM'in iki parametresi vardır: C ve σ , C , eğitim hatalarını minimuma indirmek ve modeli kontrol etmek arasındaki dengeyi kontrol eden bir sabittir ve σ Gaussian'ın genişliğini kontrol eder ve derecesine benzer bir rol oynar. Polinom çekirdeği sonuçtaki sınıflandırıcının esnekliğini kontrol eder. En iyi hiperparametreyi elde etmek için bu parametreler üzerinde bir ayar yapılmıştır.

Bir ağaç için sınıflandırma görevleri dikkate alındığında, bir karar ağacı T , t ve kenar düğümlerinden oluşan ağaç şeklinde bir sınıflandırıcıdır [8]. Herhangi bir ağaç, kök düğümü adı verilen herhangi bir kenarı olmayan bir düğümden köken alır. Terminal düğümlere, yani herhangi bir dış kenara sahip olmayan düğümlere yapraklar denir. Kalan düğümlere iç düğüm denir. Her yaprağa bir sınıf hatta bir sınıf olasılığı atanır. Her bir yaprak olmayan düğüm, giriş alanına ilişkin bir bölünmeyi temsil eder. Böyle bir bölünme $\Phi(\cdot)$ olarak adlandırılan bir kararla temsil edilir. Çoğu zaman, tek değişkenli karar olan $\Phi(\cdot) = \Phi(x)$ ($x \geq$ eşik değeri veya $x \in$ set) kullanılır. Burada optimizasyon parametresi ağaçlar (eşikler) için de uygulanmıştır; gerçek zamanlı uygulamalar için de geçerli olması için derinlik 5, çocuk sayısı da 11 olarak optimize edilmiştir.

III. SONUÇLAR VE TARTIŞMA

Önerilen SQA yöntemlerinin performansı aşağıdaki denklemde ifade edilen Hassasiyet (Se), Doğruluk (Ac) ve Özgünlük (Sp), karşılaştırma ölçütleri kullanılarak değerlendirilmiştir.

$$Se = \frac{TP}{TP + FN} \quad Sp = \frac{TN}{TN + FP} \quad (3)$$

$$Ac = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Gerçek pozitif (TP), bir SQA yöntemiyle doğru şekilde tanımlanmış kabul edilemez veya gürültülü EKG sinyalini belirtir. Yanlış negatif (FN), yöntem tarafından yeterli sinyal kalitesi var olarak tanımlanan ancak kabul edilebilir olmayan EKG sinyaline işaret eder. Gerçek negatif (TN), doğru şekilde tanımlanmış kabul edilebilir EKG sinyalini belirtir ve yanlış pozitif (FP), SQA yöntemi tarafından kabul edilemez kalitedeki tanımlanmış ancak gerçekte kabul edilebilir olan EKG kalite sinyalini belirtir.

Tüm sonuçlar ve bunlara göre iki sınıflandırma yönteminin karşılaştırması Tablo I'de verilmektedir. Test veri setinde 12 lead EKG kaydının kabul edilebilirliği, önerilen her iki sınıflandırma yöntemiyle de yaklaşık olarak 0.94 doğruluk oranı ile elde edilmiştir. Ayrıca, önerilen algoritmaların özgünlüğü çok yüksektir (0.99). Yüksek özgünlüğün nedeni, düşük sayıda yanlış pozitif olmasıdır, yani gürültü ve harekettten kaynaklanan sinyallerin varlığında, algoritma bu sinyalleri yüksek hassasiyet ile algılayarak yorumlamadan önce kabul edilemez bir sinyal olarak işaretleyecektir. Algoritmanın bu özelliği, gürültü ve insan kaynaklı yanlış alarmların sayısını azaltabildiğinden, mHealth uygulamaları için çok önemlidir.

TABLO I: Karşılaştırma Ölçütlerinin Sonuçları

Sınıflandırma yöntemi	Hassasiyet (Se)	Özgünlük (Sp)	Doğruluk (Ac)
SVM	0.925	0.99	0.941
DT	0.933	0.99	0.947

Literatürde, aynı veri tabanında uygulanan ve benzer veya daha yüksek doğruluklar sağlayan birkaç algoritma vardır. Bununla birlikte, bu algoritmalar çoğunlukla bir algortmada birkaç sinyal kalite indexinin (SQI) kombinasyonuna bağlıdır. Bu nedenle, gerçek zamanlı uygulamalar için oldukça pahalıdır. Örneğin, Behar ve ark. [3], yedi SQI kullanarak yanlış aritmi alarmının azaltılması için EKG kalitesini 99% doğrulukla değerlendirmek için bir yöntem geliştirmişlerdir. Fakat bu SQI'lerin çoğu daha yüksek hesaplama maliyeti gerektirmektedir. Zaunseder [8] EKG sinyallerini derecelendirilmek için, topluluk (ensemble) ağaçlarının bir özelliği olarak farklı frekans bantlarındaki güçleri kullanan, 90.4% doğrulukta bir algoritma önermişlerdir. Fakat frekans alanına dayalı özellikler hesaplama maliyeti için pahalıdır ve birkaç karar ağacını birleştiren topluluk yöntemini kullanmak verimli değildir.

IV. SONUÇ

En son teknoloji sinyal kalitesi değerlendirme algoritmalarının çoğunun aksine, yöntemimiz hiçbir R-tepesi tespiti, tanımlama,

hizalama, QRS kompleks tespiti [9] veya daha yüksek seviyeli EKG işleme görevlerini içermemektedir. Sonuç olarak, algoritmamızın başarısı bu tür görevlerin başarısına bağlı değildir. Buna ek olarak, sinyalden ayıkladığımız özellik sayısı, mevcut algoritmalara göre daha düşüktür. Ayrıca, algoritmamızın hesaplama maliyeti, herhangi bir frekans alanı işlemi gerektirmediği için diğer sinyal kalitesi değerlendirmelerinden [6-7] daha iyidir. Bu nedenlerden dolayı, önerilen algoritma, özellikle mHealth uygulamaları için EKG sinyal kalitesi değerlendirmesi için güçlü bir çözüm sunmaktadır.

KAYNAKLAR

- [1] U. Satija, B. Ramkumar and M. S. Manikandan, "A Review of Signal Processing Techniques for Electrocardiogram Signal Quality Assessment," in *IEEE Reviews in Biomed. Eng.*, vol. 11, pp. 36-52, 2018.
- [2] P. X. Quesnel, A. D. C. Chan, and H. Yang, "Real-time biosignal quality analysis of ambulatory ECG for detection of myocardial ischemia," in *Proc. IEEE Int. Symp. Med. Meas. Appl.* May 2013, pp. 1-5.
- [3] J. Behar, J. Oster, Q. Li, and G. D. Clifford, "ECG signal quality during arrhythmia and its application to false alarm reduction," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 6, pp. 1660-1666, Jan. 2013.
- [4] M. Vaglio et al., "Use of ECG quality metrics in clinical trials," in *Proc. Comput. Cardiol. Conf.*, Sep. 2010, pp. 505-508.
- [5] E. Morgado et al., "Quality estimation of the electrocardiogram using cross-correlation among leads," *Biomed. Eng. Online*, vol. 14, no. 1, pp. 1-19, 2015.
- [6] Q. Li and G. D. Clifford, "Signal quality and data fusion for false alarm reduction in the intensive care unit," *J. Electrocardiol.*, vol. 45, no. 6, pp. 596-603, 2012.
- [7] J. Schumm, "Quality assessment of physiological signals during ambulatory measurements," Doctoral dissertation, Eidgenössische Technische Hochschule Zurich, Nr, Zurich, Switzerland, 2010.
- [8] S. Zaunseder, R. Huhle, and H. Malberg, "CinC Challenge: Assessing the usability of ECG by ensemble decision trees," *Comput. Cardiol.*, Sep. 2011, pp. 277-280.
- [9] P. X. Quesnel, A. D. C. Chan, and H. Yang, "Real-time biosignal quality analysis of ambulatory ECG for detection of myocardial ischemia," in *Proc. IEEE Int. Symp. Med. Meas. Appl.*, May 2013, pp. 1-5.