



EKG Vurularının Tip-2 Bulanık Öbekleştirme Algoritması ile Öbekleştirilmesi

Type-2 Fuzzy Clustering of ECG Beats

İpek Toker^{1,2}, Berat Doğan², Sedef Kent Pınar²

¹ Biyomedikal Mühendisliği Enstitüsü
Boğaziçi Üniversitesi
İstanbul, Türkiye
ipek.toker@boun.edu.tr

² Elektronik ve Haberleşme Mühendisliği Bölümü
İstanbul Teknik Üniversitesi
bdogan@itu.edu.tr , kents@itu.edu.tr

Özetçe

Bu çalışmada tip-2 bulanık öbekleştirme algoritması kullanılarak altı farklı sınıfa ait EKG vuruları öbekleştirilmiştir ve sonuçlar bulanık c-ortalamalar algoritması ile karşılaştırılmıştır. Öbekleştirme başarımını ölçmek için algoritmalar tarafından bulunan nihai öbek merkezleri kullanılarak danışmanlı öğrenmeye dayalı bir sınıflama yöntemi kullanılmıştır. Yapılan deneyler sonucunda tip-2 bulanık öbekleştirme algoritmasının farklı hacim ve yoğunluktaki gruplara sahip veri kümelerini öbekleştirmede daha iyi sonuçlar verdiği gözlemlenmiştir.

Abstract

In this study, six different types of ECG beats are clustered by using type-2 fuzzy c-means algorithm and the results are compared with fuzzy c-means algorithm. By using the final cluster centers obtained with the clustering algorithms, a supervised classification method is then used to evaluate the performance of the each algorithm. After several experiments it is shown that, the type-2 fuzzy c-means algorithm performs better than the standard fuzzy c-means algorithm for the datasets that forms clusters with different volumes and densities on the feature space.

1. Giriş

Elektrokardiyogram (EKG) kalbin elektriksel aktivitesi hakkında bilgi veren girişimsiz bir yöntemdir. Kalbin elektriksel aktivitesinde bir anormallik olması durumunda ölçülen EKG işareti normal şekildedir sapmaktadır. EKG'de meydana gelen bu anormalliklerin otomatik teşhisine yönelik bazı yöntemler geliştirilmiştir. Bu yöntemler çoğunlukla işaret işleme ve örüntü tanıma dayalı yöntemlerdir.

Örüntü tanıma yöntemleri öğrenme prosedürlerine göre danışmanlı ve danışmansız yöntemler olmak üzere iki ana başlıkta incelenmektedir. Danışmanlı öğrenme yöntemleri, eğitim kümesinin etiketlenmesi esasına dayanır. Danışmansız öğrenme yöntemlerinde ise eğitim kümesinin etiketlenmesine ihtiyaç duyulmaz. Bu yöntemler bir benzerlik kriterini (örneğin Öklid uzaklığını) baz alarak eğitim verilerini gruplara

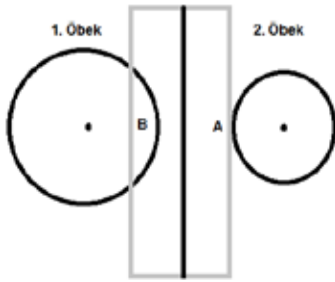
(öbeklere) ayırırlar. Böylece eğitim sonunda eğitim kümesi içerisinde birbirine benzeyen veriler bir öbek içerisinde, benzemeyenler ise farklı öbeklerde toplanırlar.

Literatürde sıklıkla kullanılan öbekleştirme algoritmalarından biri bulanık c-ortalamalar algoritmasıdır (Fuzzy c-means algorithm, FCM) [1]. Bulanık c-ortalamalar algoritmasında herhangi bir eğitim verisi birden çok öbeğe farklı üyelik değerleriyle ait olabilmektedir. Özellikle verilerin öznelik uzayında birbirleri içine geçtiği durumda bulanık c-ortalamalar algoritmasının bu özelliği katı öbekleştirme algoritmalarına kıyasla (k-ortalamalar algoritması) avantajlı olmaktadır.

Bulanık c-ortalamalar algoritması birçok farklı alanda çeşitli problemlerin çözümünde başarıyla kullanılmıştır. Fakat, algoritmanın iki temel problemi mevcuttur. Bunlardan ilki, algoritmanın başlangıç öbek merkezlerine oldukça bağımlı olması ikincisi ise, algoritmanın yerel minimum noktalara takılmasıdır. Bu sorunların önüne geçilmesi için literatürde bulanık c-ortalamalar algoritmasının performansını iyileştirmeye yönelik çeşitli hibrid yöntemler öne sürülmüş ve bu yöntemler EKG vurularının öbekleştirmede de kullanılmıştır [2][3]. Literatürde kullanılan bu hibrid yöntemler çoğunlukla bulanık c-ortalamalar algoritmasının bir optimizasyon algoritması ile kombinasyonundan oluşmaktadır. Bu yöntemler iyi sonuçlar vermesine rağmen temel problemleri bulanık c-ortalamalar algoritmasına göre hesapsal yük açısından dezavantajlı olmalarıdır. Hesapsal yük açısından avantajlı ve hibrid yöntemler kadar başarılı bir diğer yöntem ise son yıllarda önerilen tip-2 bulanık öbekleştirme algoritmasıdır [4]. Tip-2 bulanık öbekleştirme algoritmasında, bulanık c-ortalamalar algoritmasından farklı olarak iki adet farklı bulanıklaştırma katsayısı kullanılmaktadır. Bulanık c-ortalamalar algoritmasında bulanıklaştırma katsayısı tek ve değeri sabit olmaktadır. Bu katsayı algoritmanın bulanık davranışını etkileyen önemli bir parametredir ve ileriki bölümlerde bu parametrenin önemine daha detaylı değinilecektir. Özellikle veri kümesinin farklı hacim ve yoğunlukta öbeklere sahip olduğu durumlarda tip-2 bulanık öbekleştirme algoritması bulanık c-ortalamalar algoritmasına göre daha iyi sonuçlar verdiği bilinmektedir.

Sinyal İşleme 1

2. Gün 26 Eylül 2014 Cuma (09.00-10.00)



Şekil 2: Farklı hacimdeki iki öbek için bulanıklaştırma katsayısının etkisi.

Veri kümesi öznelik uzayında farklı hacim ve yoğunlukta öbeklere sahipse bulanık c-ortalamar algoritması genellikle yetersiz kalmaktadır. Çünkü böyle bir durumda bulanıklaştırma katsayısının seçilen değeri bütün öbekleri aynı oranda etkilemektedir. Bulanıklaştırma katsayısının değeri öbeklerden bir ya da birkaçı için uygunken diğerleri için uygun olmayabilir [4]. Şekil-2'de bu duruma örnek bir öbek dağılımı gösterilmiştir.

$$\bar{u}_{ij} = \begin{cases} \frac{1}{\sum_{l=1}^c \left(\frac{d_{ij}^2}{d_{lj}^2} \right)^{m_1-1}} & \text{eger } \frac{1}{\sum_{l=1}^c \left(\frac{d_{ij}^2}{d_{lj}^2} \right)^{m_1-1}} > \frac{1}{\sum_{l=1}^c \left(\frac{d_{ij}^2}{d_{lj}^2} \right)^{m_2-1}} \text{ ise} \\ \frac{1}{\sum_{l=1}^c \left(\frac{d_{ij}^2}{d_{lj}^2} \right)^{m_2-1}} & \text{aksi taktirde} \end{cases} \quad (6)$$

$$\underline{u}_{ij} = \begin{cases} \frac{1}{\sum_{l=1}^c \left(\frac{d_{ij}^2}{d_{lj}^2} \right)^{m_2-1}} & \text{eger } \frac{1}{\sum_{l=1}^c \left(\frac{d_{ij}^2}{d_{lj}^2} \right)^{m_1-1}} > \frac{1}{\sum_{l=1}^c \left(\frac{d_{ij}^2}{d_{lj}^2} \right)^{m_2-1}} \text{ ise} \\ \frac{1}{\sum_{l=1}^c \left(\frac{d_{ij}^2}{d_{lj}^2} \right)^{m_1-1}} & \text{aksi taktirde} \end{cases} \quad (7)$$

Şekil-2'de B noktası 1. öbeğin sınırları içerisinde yer almasına rağmen, maksimum bulanıklık bölgesine dâhil olmasından dolayı, 1. öbeğe üyelik değeri 1'den uzaklaşacaktır ve öbek merkezi hesaplanırken katkısı daha az olacaktır. Bulanıklaştırma katsayısının değeri küçültülürse, bu defa da A noktası maksimum bulanıklık bölgesinin dışında kalacak ve 2. öbeğe 1'e yakın bir üyelik değeriyle bağlı olacaktır. Görüldüğü gibi, her iki kümeye de uygun bir bulanıklık bölgesini tek m değeri ile oluşturmak mümkün değildir [4]. Tip-2 bulanık öbekleştirme algoritmasında iki farklı m değeri kullanılarak bu sorunların önüne geçilmek istenmiştir. Tip-2 bulanık öbekleştirme algoritması ile iki farklı m değeri kullanılarak her bir veri için iki farklı (alt ve üst üyelik değerleri) üyelik değeri Eşitlik-6 ve Eşitlik-7 ile hesaplanmaktadır.

Alt ve üst üyelik değerlerinin ortalaması alınarak (Eşitlik-8) öbek merkezlerini hesaplamada kullanılacak üyelik matrisi

bulunur. Daha sonra Eşitlik-1 ile bulanık c-ortalamar algoritmasında olduğu gibi öbek merkezleri hesaplanır.

$$u = \frac{\bar{u} + \underline{u}}{2} \quad (8)$$

İki farklı m değeri kullanıldığından öbek merkezleri de üyelik değerlerinde olduğu gibi bir aralık ile temsil edilecektir. Bulanık c-ortalamar algoritmasında olduğu gibi öbek merkezlerinin iteratif olarak güncellenebilmesi için tip azaltma işlemi yapılmalıdır. Herhangi bir tip-2 veri seti, n veri için 2^n tane alt veri setinden (tip-1) oluşur ve tip azaltma işlemi yapılabilmesi için hesaplamalarda bunların hepsi dikkate alınmalıdır. Çok sayıda veri mevcutken bu mümkün olmadığından, küme merkezi aralığının sınırlarını (c_R ve c_L) hesaplayıp tip azaltma işlemi yapmak için Karnik ve Mendel [7] tarafından önerilen yöntem kullanılır. Karnik-Mendel algoritmasına ait adımlar Tablo-2'de verilmiştir.

Tablo 2: Karnik-Mendel algoritmasının adımları

c_R 'yi bulmak için şöyle bir yol izlenir;

1. Veri artan şekilde sıralanır ve $x_k \leq c \leq x_{k+1}$ şartını sağlayan k değeri bulunur. ($1 \leq k \leq n-1$)

2. $i > k$ ise $u' = \bar{u}$

$i \leq k$ ise $u' = \underline{u}$

3. Yeni u' değeri kullanılarak Eşitlik-1 ile c' değeri hesaplanır ve c ile c' değeri karşılaştırılır. Aradaki fark, seçilen ε değerinden küçükse işlem sonlandırılır, aksi taktirde Adım-1'e dönülür.

Not: c_L 'yi bulmak için de aynı yol izlenir. Sadece Adım-2 aşağıdaki gibi değiştirilmelidir.

$i \leq k$ ise $u' = \bar{u}$

$i > k$ ise $u' = \underline{u}$

Karnik-Mendel algoritması kullanılarak öbek merkezlerinin sınır değerleri bulunduktan sonra, Eşitlik-9 ile tip azaltma işlemi tamamlanmış olur.

$$c'' = \frac{c_R + c_L}{2} \quad (9)$$

Yeni hesaplanan öbek merkezi bir önceki adımda hesaplanan ile karşılaştırılır. Aradaki fark seçilen ε değerinden küçükse işlem sonlandırılır. Şart sağlanmıyorsa alt ve üst üyelik değerleri yeniden hesaplanarak aynı işlemler tekrarlanır.

Algoritmanın sonlandırma şartı sağlandıktan sonra, bölütleme işlemi yapılabilmesi için katı-sınıflama işlemi gerekir. Ancak veri seti artık tip-2 olduğundan bu aşama da klasik c-ortalamar algoritmasındaki gibi uygulanmaz ve burada da tip azaltma işlemi gerekir [4]. Tip azaltma Eşitlik-10'daki gibi yapılır.

$$u_{ij} = \frac{u_{ij}^R + u_{ij}^L}{2} \quad (10)$$

u_{ij}^R ve u_{ij}^L sırasıyla Eşitlik-11 ve Eşitlik-12 ile bulunabilir.

Sinyal İşleme 1

2. Gün 26 Eylül 2014 Cuma (09.00-10.00)

$$u_{ij}^R = \frac{\sum_{j=1}^n u_{ij}(x_j)}{n}$$

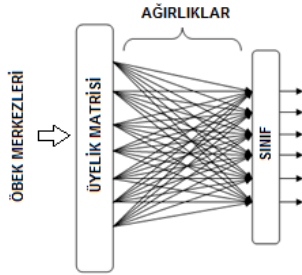
$$u_{ij}(x_j) = \begin{cases} \bar{u}_{ij} & \text{eger } l.\text{vericinin } c_i^R \text{ hesaplanırken } \bar{u}_{ij} \text{ kullanılmışsa} \\ u_{ij} & \text{eger } l.\text{vericinin } c_i^R \text{ hesaplanırken } u_{ij} \text{ kullanılmışsa} \end{cases} \quad (11)$$

$$u_{ij}^L = \frac{\sum_{j=1}^n u_{ij}(x_j)}{n}$$

$$u_{ij}(x_j) = \begin{cases} \bar{u}_{ij} & \text{eger } l.\text{vericinin } c_i^L \text{ hesaplanırken } \bar{u}_{ij} \text{ kullanılmışsa} \\ u_{ij} & \text{eger } l.\text{vericinin } c_i^L \text{ hesaplanırken } u_{ij} \text{ kullanılmışsa} \end{cases} \quad (12)$$

3. Sonuçlar

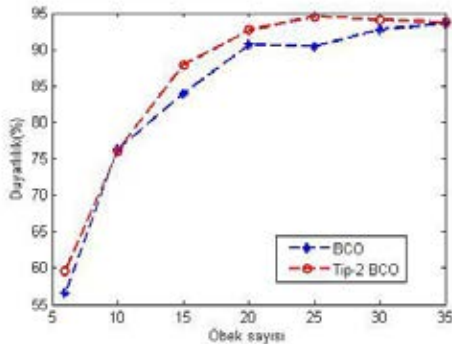
Eğitim sırasında algoritmalar koşturularak öbek merkezleri ve üyelik matrisi bulunmuştur. Algoritmaların performansını test etmek için danışmanlı bir sınıflama yöntemi kullanılmıştır. Şekil-3'te kullanılan sınıflama yöntemi görülmektedir.



Şekil 3: Algoritmaların performansını değerlendirmede kullanılan sınıflama yöntemi.

Şekil-3 ile verilen sınıflama yönteminde bir test verisini sınıflamadan önce eğitim esnasında elde edilen üyelik matrisi kullanılarak Eşitlik-13 yardımıyla ağırlıklar bulunmalıdır.

$$W = (U^T U)^{-1} U^T * T \quad (13)$$



Şekil 4: Uygulanan yöntemlerin başarımları.

Burada W ağırlıkları, U eğitim verisine ait üyelik matrisini, T ise istenen çıkışı temsil etmektedir. Daha sonra test verileri sınıflandırılırken eğitim aşamasında elde edilen öbek merkezleri kullanılarak test verileri için üyelik matrisi hesaplanır. Elde edilen üyelik matrisi Eşitlik-14'te görüldüğü

gibi daha önce bulunan ağırlıklar ile çarpılarak test verisi sınıflandırılır.

$$o = U_t * W \quad (14)$$

Burada, o sınıflandırılmış test verisini U_t ise test verilerine ait üyelik matrisini temsil etmektedir.

Şekil-4'te 6, 10, 15, 20, 25, 30 ve 35 öbek merkezi için algoritmaların performansı verilmiştir. Verilen sonuçlar her bir algoritma için 10 deneme sonunda elde edilen ortalama duyarlılık sonuçlarıdır. Bulanıklaştırma katsayılarının değeri bulanık c-ortalamalar algoritması için $m=1.5$ ve tip-2 bulanık c-ortalamalar algoritması için $m=1.5$, $m_1=1.1$, $m_2=1.9$ olarak seçilmiştir.

Tablo-3'de 25 küme için tip-2 bulanık c-ortalamalar algoritmasıyla elde edilen en iyi sınıflama başarımları verilmiştir.

Tablo 3: 25 küme için sınıflama sonuçları.

	N	PVC	F	A	R	f	Duy.(%)	Özg.(%)
N	815	0	8	2	11	0	97,49	98,54
PVC	0	574	22	8	0	1	94,88	99,87
F	14	0	231	0	2	1	93,15	97,77
A	2	1	2	160	0	0	96,97	99,44
R	3	0	0	1	146	0	97,33	99,29
f	0	1	10	0	1	118	90,77	99,9
ORT							95,10	99,14

4. Kaynakça

- [1] Bezdek, J.C., Ehrlich, R. ve Full, W., "FCM: The fuzzy c-means clustering algorithm", *Computers & Geosciences*, vol. 10., 191-203, 1984.
- [2] Korürek, M. ve Doğan, B., "ECG beat classification using particle swarm optimization and radial basis function neural network", *Expert Systems with Applications*, vol.37, no.12, 7563-7569, 2010.
- [3] Doğan, B. ve Korürek, M., "A new ECG Beat Clustering Method Based on Kernelized Fuzzy C- Means and Hybrid Ant Colony Optimization for Continuous Domains", *Applied Soft Computing*, vol.12 no.11, 3442-3451, 2012.
- [4] Hwang, C. ve Rhee, F.C.H., "Uncertain Fuzzy Clustering: Interval Type-2 Fuzzy Approach to C-Means", *IEEE Transactions on Fuzzy Systems*, vol. 15 no.1, 107-120, 2007.
- [5] Ceylan, R. ve Özbay, Y. ve Karlık, B., "A novel approach for classification of ECG arrhythmias: Type-2 fuzzy fuzzy clustering neural network", *Expert Systems with Applications*, vol.36, no.3, 6721-6726, 2009.
- [6] Özbay, Y., Ceylan, R. ve Karlık, B., "Integration of type-2 fuzzy clustering and wavelet transform in a neural network based ECG classifier", *Expert Systems with Applications*, vol.38, no.1, 1004-1010, 2011.
- [7] Karnik, N.N., ve Mendel, J.M., "Type-2 fuzzy logic systems: type-reduction", *IEEE International Conference on Systems, Man, and Cybernetics*, 1998.