ANADOLU ÜNİVERSİTESİ BİLİM VE TEKNOLOJİ DERGİSİ
ANADOLU UNIVERSITY JOURNAL OF SCIENCE AND TECHNOLOGY
Cilt/Vol.:7 Sayı/No: 1 : 33-39 (2006)

DERLEME/**REVIEW**

SAĞLAM (ROBUST) KÜMELEME YÖNTEMLERİ

Kamile ŞANLI, Ayşen APAYDIN[1],

ÖZ

Bulanık *c* ortalamaya dayalı kümeleme algoritmaları sıfır kırılma noktasına sahiptir. Başka bir ifadeyle tek bir aykırı değer modeli tümüyle değiştirebilir. Veri kümesinde aykırı değer olduğunda bulanık *c* ortalamaya dayalı kümeleme tipi algoritmalar pratik değere sahip değildir. Bu nedenle aykırı değerden etkilenmeyen sağlam kümeleme yöntemleri önerilmiştir.

Bu çalışmada kümeleme analizlerinde yaygın olarak kullanılan katı *c* ortalamaya dayalı kümeleme ve bulanık *c* ortalamaya dayalı kümeleme yöntemleri ele alınmıştır. Veri kümesinde aykırı değer olması durumunda aykırı değerden etkilenmeyen gürültü kümeleme, olanaklı *c* ortalamaya dayalı kümeleme ve karma *c* ortalamaya dayalı kümeleme yöntemlerinin tanımları verilmiş ve karşılaştırılmıştır.

Anahtar kelimeler: Kümeleme, Bulanık kümeleme, Sağlam kümeleme

ROBUST CLUSTERING METHODS

ABSTRACT

Fuzzy *c* means algorithms have zero break point. In other words a single outlier can completely change the model. When data set has outlier, fuzzy *c* means type algorithms haven't practical value. Therefore, robust clustering methods which are not affected by outlier have been suggested.

In this paper, hard *c* means clustering and fuzzy *c*-means clustering methods which are commonly used in clustering analysis are examined. Noise clustering, possibilistic *c* means clustering and mixed *c* means clustering methods which aren't affected by outlier are defined and compared, when data set has outlier.

Key words: Clustering, Fuzzy Clustering, Robust clustering

[1] Ankara Üniversitesi Fen Fakültesi, İstatistik Bölümü, 06100 Tandoğan/ANKARA
E-posta: apaydin@science.ankara.edu.tr

Geliş: 15 Eylül 2004; **Düzeltme:** 09 Mart 2005; **Kabul:** 26 Temmuz 2005

1. GİRİŞ

Katı *c* ortalamaya (hard *c*-means) dayalı kümelemede birimler ya bir kümeye aittir ya da değildir. Fakat birimler kümelere tam üye olmamakla birlikte belirli üyeliklerle birden fazla kümeye ait olabilirler. Bu durumda bulanık *c* ortalamaya (fuzzy *c*-means (FCM)) dayalı kümeleme yöntemi daha uygundur. Veri kümesinde aykırı değer ya da gürültü olması durumunda ise katı ve bulanık *c* ortalamaya dayalı kümeleme yöntemleri uygun değildir. Çünkü bulanık *c* ortalamaya dayalı kümeleme algoritmaları sıfır kırılma noktasına sahiptir. Başka bir ifadeyle tek bir aykırı değer modeli tümüyle değiştirebilir (Dave 1993, Dave ve Krishnapuram 1997). Bu durumda sağlam (robust) kümeleme yöntemleri kullanılır.

Sağlam bulanık kümeleme yöntemleri ile ilgili yapılan çalışmalar ele alındığında aşağıdaki gibi özetlenebilir:

Gürültüye karşı daha fazla dirençli parametre tahminleri yapmak için bulanık *c* ortalamaya dayalı kümelemenin (FCM) amacını yeniden düzenlemeye dayanan ilk yöntem Ohashi (1984) tarafından önerilmiştir (Dave ve Krishnapuram 1997, Nasraoui 2004).

Daha sonra Dave (1991) gürültü kümeleme (noise clustering (NC)) yöntemini ileri sürmüştür. NC yaklaşımının ana avantajı FCM algoritmasının sağlam şekli olmasıdır (Dave 1993, Dave ve Sen 1998).

FCM'in göreli üyelik probleminin üstesinden gelmek için Krishnapuram ve Keller (1993) olanaklı *c* ortalamaya (possibilistic *c*-means (PCM)) dayalı kümeleme yöntemini önermişlerdir.

Dave ve Krisnapuram (1996) sağlam kümeleme yöntemlerini karşılaştırmışlardır. Sağlam istatistikler ve bulanık küme teorisi arasındaki benzerlikleri saptayıp sağlam kümeleme yöntemleri ve M-tahmin edicileri arasındaki benzerlikleri göstermişlerdir.

Pal ve Bezdek (1997) hem bulanık kümelemedeki üyeliklerin hem de bir gözlemin bir kümeye ne kadar uygun olduğunu gösteren küme özelliğini (typicality (t_{ij})) içeren karma c ortalamaya (mixed c-means (FPCM)) dayalı kümeleme modelini tanımlamışlardır.

Dave ve Sen (1998), Dave (1991) tarafından önerilen gürültü kümeleme algoritmasını genelleştirmişlerdir.

Dave ve Krishnapuram (1997) sağlam kümeleme yöntemlerinin karşılaştırıldığı bir çalışma yapmışlardır.

Bu çalışmanın amacı, literatüre yer alan ve sıklıkla kullanılan katı, bulanık c ortalamaya dayalı kümeleme yöntemleri ve sağlam kümeleme yöntemlerinin tanımlarını vermek ve bu yöntemleri karşılaştırmaktır. Bu amaçla, İkinci Bölümde katı ve bulanık *c* ortalamaya dayalı kümeleme yöntemlerinin tanımları ve Üçüncü Bölümde ise sağlam kümeleme yöntemleri için tanımlar verilmiştir.

2. KATI VE BULANIK KÜMELEME

$X=\{x_1,x_2,...,x_n\}$ veri kümesi ele alınsın. Burada x_j herhangi bir eleman ve $x_i \in R^p$'dir. X'in kuvvet kümesi $P(X)$ ile gösterilsin. $\bigcup_{i=1}^{c} A_i = X$ ve $A_i \cap A_j = \varnothing, 1 \le i \ne j \le c$ için X'in katı *c* parçalanması $\{A_i \in P(X) : 1 \le i \le c\}$ ailesidir. Her bir A_i küme olarak düşünülebilir, böylece $\{A_1, A_2, ..., A_c\}$ c kümedeki X parçalanmalarıdır.

Katı *c* parçalanma,'deki x_j elemanlarının üyelik fonksiyonu ile

$$u_{ij} = \begin{cases} 1 & ; \ x_j \in A_i \\ 0 & ; \ x_j \notin A_i \end{cases} \qquad \textbf{(1)}$$

olarak tanımlanır. Burada
$x_j \in X, A_i \in P(X), i = 1,2,...,c$ 'dır. Eşitlik (1)'deki u_{ij} aşağıda ifade edilen üç koşulu sağlamaktadır.

$$u_{ij} \in [0,1] \qquad 1 \le i \le c, \ 1 \le j \le n \qquad \textbf{(2)}$$

$$\sum_{i=1}^{c} u_{ij} = 1, \ \ \forall j \in \{1,2,...,n\} \qquad \textbf{(3)}$$

$$0 < \sum_{j=1}^{n} u_{ij} < n, \ \forall i \in \{1,2,...,n\} \qquad \textbf{(4)}$$

Koşul (2) ve (3) her bir $x_j \in X$'in yalnız ve yalnız bir kümeye ait olmasını, Koşul (4) ise her bir A_i kümesinin en az bir en çok $n-1$ veri noktasını içermesini sağlar. c x n boyutlu U matrisi u_{ij}'lerden oluşur ($1 \le i \le c$ ve $1 \le k \le n$) (Wang 1997).

Tanım 1. $X=\{x_1,x_2,...,x_n\}$ herhangi bir küme ve V_{cn}, c x n boyutlu $U=[u_j]$ matrisinin kümesi olsun. X'in katı c parçalanma uzayı,

$M_c = \{U \in V_{cn} \setminus$ Koşul (2) ve Koşul (3) doğru $\}$

kümesidir. M_c

$$|M_c| = \frac{1}{c!}\left[\sum_{j=1}^{c}\binom{c}{j}(-1)^{c-j} j^n\right]$$

olarak tanımlanır (Wang 1997).

Tanım 2. $X=\{x_1,x_2,...,x_n\}$ herhangi bir küme ve V_{cn}, c x n boyutlu $U=[u_j]$ matrisinin kümesi olsun. X 'in bulanık c parçalanma uzayı

$M_{fc}=\{U\in V_{cn} / u_{ik}\in[0,1], 1\le i\le c, 1\le k\le n$; Koşul (3) doğru $\}$

şeklinde tanımlanır. u_{ij}, A_i kümesine ait olan x_j 'nin üyeliğidir (Wang 1997).

2.1 Katı *c* Ortalamaya Dayalı Kümeleme

Katı c ortalamaya dayalı kümelemede amaç fonksiyonu,

$$J(B,U;X)=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}\left(d_{ij}\right)^2 \qquad \textbf{(5)}$$

olarak tanımlanır. Burada $\left(d_{ij}\right)^2=\left\|x_j-v_i\right\|^2$ 'dir. Eşitlik (5)'ü minimum yapacak küme merkezi ve üyelikler

$$v_i=\frac{\sum_{j=1}^{n}x_j u_{ij}}{\sum_{j=1}^{n}u_{ij}} \qquad \textbf{(6)}$$

$$u_{ij}=\begin{cases}1 \quad ; \quad \left\|x_j-v_i\right\|=\min_{1\le k\le c}\left(\left\|x_j-v_k\right\|\right)\\ 0 \quad ; \qquad\qquad d.d.\end{cases} \qquad \textbf{(7)}$$

olarak tanımlanır (Nasraoui 2004, Wang 1997).

2.2 Bulanık c Ortalamaya Dayalı Kümeleme

Katı c ortalamaya dayalı kümeleme ile FCM arasındaki temel fark birimlerin bir kümeye ait olma üyeliğinin belirlenmesidir. $U=[u_j]\in M_{fc}$ ve $v_i\in R^p$ olan $V=(v_1,v_2,...,v_c)$'yi bulmak için FCM'de amaç,

$$J(B,U;X)=\sum_{i=1}^{c}\sum_{j=1}^{n}\left(u_{ij}\right)^m\left(d_{ij}\right)^2 \qquad \textbf{(8)}$$

biçimindeki fonksiyonunun minimum yapılmasıdır. Burada $m\in[1,\infty)$ bulanıklık olarak adlandırılan ağırlıklandırma sabitidir. m'ye küçük değerler verilirse bulanıklık azalır, büyük değerler verilirse bulanıklık artar. Genellikle m'ye "2" değeri verilir. Eşitlik (8)'de verilen amacı minimum yapan küme merkezleri ve üyelikler

$$v_i=\frac{\sum_{j=1}^{n}x_j\left(u_{ij}\right)^m}{\sum_{j=1}^{n}\left(u_{ij}\right)^m} \qquad \textbf{(9)}$$

$$u_{ij}=\frac{1}{\sum_{k=1}^{c}\left(\frac{\left(d_{ij}\right)^2}{\left(d_{kj}\right)^2}\right)^{\frac{1}{(m-1)}}} \qquad 1\le i\le c, 1\le j\le n \qquad \textbf{(10)}$$

şeklinde tanımlanır (Dave ve Krishnapuram 1997, Pal ve Bezdek 1997, Wang 1997).

FCM algoritmaları sıfır kırılma noktasına sahiptir. Başka bir ifadeyle tek bir aykırı değer modeli tümüyle değiştirebilir. En küçük hata kare minimizasyonu ya da pek çok kümeleme algoritması gürültüye karşı güçlü değildir. Veri kümesinde gürültü olduğunda FCM tipi algoritmalar pratik değere sahip değildir. Bu nedenle gürültüden etkilenmeyen sağlam kümeleme yöntemleri önerilmiştir (Dave 1993, Dave ve Krishnapuram 1997).

3. SAĞLAM KÜMELEME YÖNTEMLERİ

Sağlam kümeleme yöntemleri iki sınıfta toplanabilir. Bunlardan birincisi doğrudan sağlam istatistiklere dayanan yöntemleri içerir. İkincisi ise gürültüye daha fazla dirençli parametre tahminleri yapmak için FCM'in amacını yeniden düzenlemeye dayanır.

Sağlamlık ile varsayılan modelden küçük sapmaların algoritmanın performansını önemli bir şekilde etkilemediği, aykırı değerler ve gürültüden dolayı çok ciddi şekilde bozulmanın olmadığı ifade edilir.

Verideki aykırı değer ve gürültünün etkisini azaltma özelliğine sahip yöntemler çok önemli ve yaygındır. Sağlam istatistikler ve bulanık küme teorisi son yirmi yılda bağımsız olarak yavaş yavaş gelişen iki disiplindir. Bununla birlikte olabilirlik (possibility) teorisindeki, olabilirlik dağılımları ya da bulanık küme teorisindeki üyelik fonksiyonları kavramı sağlam istatistiklerdeki ağırlık fonksiyonu kavramıyla benzer ve pek çok ortak noktaya sahiptirler.

Sağlam istatistik bakış açısından, sağlamlık kavramını araştırmak yararlıdır. Huber (1981)'e göre bir sağlam yöntem,

1. Varsayılan model altında iyi bir etkinliğe sahiptir,

2. Model varsayımlarından küçük sapmalar yalnızca küçük miktarda performansı bozar,

3. Model varsayımlarından büyük sapmalar çok büyük bozulmaya neden olmaz

şeklinde tanımlanan üç özelliği sağlamalıdır.

Bulanık küme teorisi ve sağlam istatistikler otuz yıl önce ileri sürülmesine rağmen kümeleme ile bağlantısı son yıllarda yapılmıştır (Dave ve Krishnapuram 1997, Nasraoui 2004).

3.1 Gürültü Kümeleme Yöntemi

Gürültüye karşı daha fazla dirençli parametre tahminleri yapmak için FCM'in amacını yeniden düzenlemeye dayanan ilk yöntem Ohashi (1984) tarafından önerilmiştir (Dave ve Krishnapuram 1997, Nasraoui 2004). Ohashi (1984) Eşitlik (8)'de verilen amaç fonksiyonunu düzenleyerek aykırı değer sınıfı fikrini ileri sürmüştür. Bu durumda amaç fonksiyonu,

$$J(B,U;X)=\alpha\sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ij})^m(d_{ij})^2+(1-\alpha)\sum_{j=1}^{n}(u_{*j})^m \quad \textbf{(11)}$$

olarak tanımlanır. Burada u_{*j} aykırı değer sınıfındaki x_j noktasının üyeliği ve α parametresi önceden belirlenen bir sabittir (Dave ve Krishnapuram 1997, Nasraoui 2004). Daha sonra Dave (1991) gürültü kümeleme (NC) yöntemini ileri sürmüştür (Dave 1993, Dave ve Sen 1998). NC yönteminde, gürültü tüm veri noktasından δ sabit uzaklığına sahip olarak tanımlanır. NC'de x_j 'nin üyeliği,

$$u_{*j}=1-\sum_{i=1}^{c}u_{ij} \quad \textbf{(12)}$$

biçiminde tanımlanır. Gürültü sınıfındaki u_{*j} üyelik tanımlamasında Eşitlik (12) kullanıldığında FCM'in alışılmış üyelik kısıtı gerekmez. Böylece iyi sınıf için üyelik kısıtı,

$$0<\sum_{i=1}^{c}(u_{ij})\le 1 \quad \textbf{(13)}$$

olarak yumuşatılır. Eşitlik (13)'de verilen kısıt iyi kümelerde gürültü noktalarının küçük üyelik değerleri almasına izin verir. NC yönteminde amaç fonksiyonu,

$$J(B,U;X)=\sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ij})^m(d_{ij})^2+\sum_{j=1}^{n}\delta^2\left(1-\sum_{i=1}^{c}u_{ij}\right)^m \quad \textbf{(14)}$$

şeklinde tanımlanır. Eşitlik (14)'de verilen amaç fonksiyonun minimum yapılmasıyla üyelikler,

$$u_{ij}=\frac{1}{\sum_{k=1}^{c}\left(\frac{(d_{ij})}{(d_{kj})}\right)^{\frac{2}{(m-1)}}+\left(\frac{(d_{ij})}{\delta}\right)^{\frac{2}{(m-1)}}}$$

olarak elde edilir. Eğer $\alpha=\frac{1}{1+\delta^2}$ olarak alınırsa Eşitlik (11) ve Eşitlik (14) aynıdır. NC yaklaşımının ana avantajı FCM algoritmasının sağlam şekli olması ve δ için uygun bir değer bulunarak FCM algoritmasının yerine kullanılabilir olmasıdır (Dave 1993, Dave ve Krishnapuram 1996, Dave ve Sen 1998).

3.2 Olanaklı *c* Ortalamaya Dayalı Kümeleme Yöntemi

FCM'de kullanılan üyelik üzerine konan kısıt yumuşatılarak gürültü probleminin üstesinden gelmek için Krishnapuram ve Keller (1993) olanaklı c ortalamaya dayalı kümeleme (PCM) yöntemini ileri sürmüşlerdir.

PCM'de kümeyi temsil etmeyen gözlemlerin küçük üyelik değerine sahip olması istenirken kümeyi temsil eden gözlemlerin mümkün olduğunca yüksek üyelik değerine sahip olması istenir. Bu durumda amaç fonksiyonu

$$J(B,U;X)=\sum_{i=1}^{c}\sum_{j=1}^{n}(u_{ij})^m(d_{ij})^2+\sum_{i=1}^{c}\eta_i\sum_{j=1}^{n}(1-u_{ij})^m \quad \textbf{(15)}$$

olarak tanımlanır. Eşitlik (15)'de ikinci terimin mümkün olduğunca büyük üyeliğe (u_{ij}) sahip olması, ilk terim için uzaklığın mümkün olduğunca küçük olması istenir.

Uygulamada

$$\eta_i=K\frac{\sum_{j=1}^{n}(u_{ij})^m(d_{ij})^2}{\sum_{j=1}^{n}(u_{ij})^m} \qquad K>0 \quad \textbf{(16)}$$

iyi sonuçlar vermektedir. Burada K genellikle "1" alınır. Ayrıca Eşitlik (16) yerine

$$\eta_i=\frac{\sum_{x_j\in(\pi_i)_\alpha}(d_{ij})^2}{|(\pi_i)_\alpha|} \quad \textbf{(17)}$$

eşitliği kullanılabilir. Burada $(\pi_i)_\alpha, \pi_i$'nin uygun α kesmesidir.

η_i değeri tüm iterasyonlar için sabit olarak alınabilir ya da her bir iterasyonda değişebilir. En iyi yaklaşım Eşitlik (16) kullanılarak bir başlangıç bulanık parçalanmaya dayanan η_i hesaplamak ve algoritma yakınsadıktan sonra Eşitlik (17) kullanılarak η_i için daha kesin değerler hesaplamaktır. Ayrıca kümelerin yapısı bilindiğinde η_i saptanmış bir ön bilgi (priori) olabilir.

Eşitlik (15)'ün minimum yapılmasıyla elde edilen üyelik derecesi

$$u_{ij}=\frac{1}{1+\left(\frac{d_{ij}^2}{\eta_i}\right)^{\frac{1}{m-1}}}$$

biçiminde tanımlanır (Barni vd. 1996, Dave ve Krishnapuram 1997, Krishnapuram ve Keller 1993, Krishnapuram ve Keller 1996).

NC yaklaşımındaki δ^2 gürültü uzaklığı ile PCM yaklaşımındaki η_i ağırlığı benzerdir. Fakat PCM her bir küme için ağırlığın farklı olması avantajına sahiptir. NC yaklaşımında bir gürültü sınıfı varken PCM yaklaşımında *c* gürültü sınıfı vardır. c=1 olduğunda NC ve PCM yaklaşımları $\delta^2 = \eta$ için aynıdır.

3.3 Karma c Ortalamaya Dayalı Kümeleme

Pal ve Bezdek (1997) hem bulanık kümelemedeki üyeliklerin hem de bir gözlemin bir kümeye ne kadar uygun olduğunu gösteren küme özelliğini (t_{ij}) içeren karma *c* ortalamaya dayalı kümeleme (FPCM) modelini tanımlamışlardır. FPCM'de amaç fonksiyonu,

$$J(B,U;X) = \sum_{i=1}^{c}\sum_{j=1}^{n}\left((u_{ij})^m + (t_{ij})^\eta\right)(d_{ij})^2 \qquad \textbf{(18)}$$

olarak tanımlanır. (18) eşitliği aşağıda verilen kısıtlar altında minimum yapılmalıdır.

$$u_{ij} = \left(\sum_{k=1}^{c}\left(\frac{d_{ij}}{d_{kj}}\right)^{\frac{2}{(m-1)}}\right)^{-1}, t_{ij} = \left(\sum_{k=1}^{n}\left(\frac{d_{ij}}{d_{ik}}\right)^{\frac{2}{(\eta-1)}}\right)^{-1},$$

$$v_i = \frac{\sum_{j=1}^{n}\left((u_{ij})^m + (t_{ij})^\eta\right)x_j}{\sum_{j=1}^{n}\left((u_{ij})^m (t_{ij})^\eta\right)} \quad \sum_{i=1}^{c} u_{ij} = 1, \qquad \sum_{j=1}^{n} t_{ij} = 1$$

$m > 1, \eta > 1, u_{ij} \geq 0, t_{ij} \leq 1$ i=1,2,…,c; j=1,2,…,n

$T = \lfloor t_{ij} \rfloor$ olmak üzere, $\sum_{j=1}^{N} t_{ij} = 1$ kısıtı T'nin her bir sütun toplamının "1" olmasını gerektirir. Böylece T'lerin küme üyelikleri

$$\sum_{j=1}^{n} t_{ij} = 1, \qquad \sum_{i=1}^{c} t_{ij} > 0$$

olmalıdır. Daha önceki yapılan çalışmalardan FPCM'de η'nün [3, 5] aralığında olmasının iyi olacağı önerilmiştir (Pal ve Bezdek 1997).

4. UYGULAMA

Bu Bölümde FCM ve Sağlam kümeleme yöntemleri için uygulama amaçlandı. Bu amaçla veri kümesinde aykırı değer olması durumuna uygun yapay veri türetildi. Bu örnek verinin türetilmesi ve çözümü için Matlab paket programında yazılan programdan yararlanılmıştır. Ele alınan verilere ilişkin saçılım grafiği Şekil 1'de verilmiştir. FCM, PCM ve FPCM yöntemleri için elde edilen küme merkezleri Tablo 1'de ve üyelikler Tablo 2'de verilmiştir.

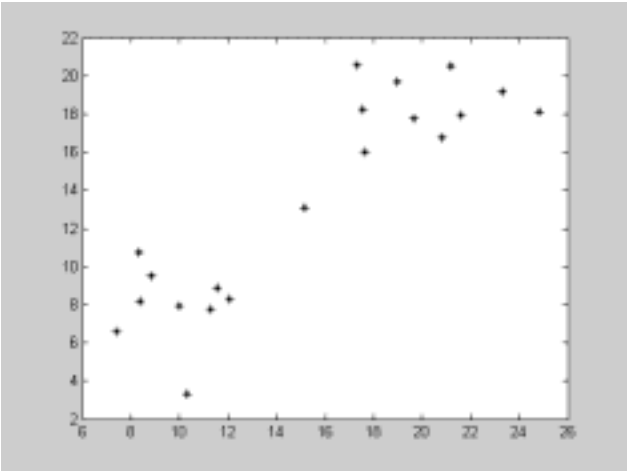
Tablo 1'de FCM, PCM ve FPCM için küme merkezleri verilmiştir. FCM için birinci kümenin merkezi (9.9692; 8.1248) iken ikinci kümenin merkezi (20.1687;18.3340)'dir. PCM için küme merkezleri (10.2503; 8.4043) ve (19.8559;18.1899)'dir. FPCM için birinci küme merkezi (9.9694;8.1076), ikinci küme merkezi ise (19.8559;18.1899)'dir.

Tablo 1. FCM, PCM ve FPCM için Küme Merkezleri

FCM		**PCM**		**FPCM**	
1	**2**	**1**	**2**	**1**	**2**
$\begin{pmatrix}9.9692\\8.1248\end{pmatrix}$	$\begin{pmatrix}20.1687\\18.3340\end{pmatrix}$	$\begin{pmatrix}10.2503\\8.4043\end{pmatrix}$	$\begin{pmatrix}19.8559\\18.1899\end{pmatrix}$	$\begin{pmatrix}9.9694\\8.1076\end{pmatrix}$	$\begin{pmatrix}20.1559\\18.3173\end{pmatrix}$

Tablo 2. FCM, PCM ve FPCM için Üyelikler

Sıra No	FCM Üyelik		PCM Üyelik		FPCM Üyelik		t_{ij}		Sıralanmış t_{ij}	
1	0.9898	0.0102	0.6906	0.0375	0.9898	0.0102	0.0081	0.0012	2	20
2	0.9999	0.0001	0.9669	0.0426	0.9999	0.0001	0.9566	0.0014	3	11
3	0.9908	0.0092	0.8463	0.0470	0.9908	0.0092	0.0114	0.0015	1	17
4	0.9330	0.0670	0.2317	0.0280	0.9333	0.0667	0.0009	0.0009	8	19
5	0.5029	0.4971	0.1440	0.1583	0.5006	0.4994	0.0004	0.0056	10	18
6	0.9710	0.0290	0.4041	0.0301	0.9710	0.0290	0.0023	0.0010	7	15
7	0.9746	0.0254	0.7058	0.0535	0.9745	0.0255	0.0046	0.0017	6	13
8	0.9817	0.0183	0.8027	0.0546	0.9815	0.0185	0.0065	0.0018	9	12
9	0.9536	0.0464	0.4577	0.0457	0.9530	0.0470	0.0021	0.0015	4	16
10	0.9843	0.0157	0.7025	0.0438	0.9841	0.0159	0.0061	0.0014	5	14
11	0.0092	0.9908	0.0342	0.7463	0.0093	0.9907	0.0001	0.1324	12	5
12	0.0889	0.9111	0.0647	0.4822	0.0877	0.9123	0.0002	0.0248	20	8
13	0.0343	0.9657	0.0264	0.4088	0.0346	0.9654	0.0001	0.0268	11	7
14	0.0636	0.9364	0.0248	0.2671	0.0638	0.9362	0.0001	0.0132	17	3
15	0.0412	0.9588	0.0493	0.6283	0.0408	0.9592	0.0001	0.0425	19	9
16	0.0589	0.9411	0.0379	0.4280	0.0588	0.9412	0.0001	0.0222	18	2
17	0.0149	0.9851	0.0411	0.7509	0.0147	0.9853	0.0001	0.1006	13	10
18	0.0199	0.9801	0.0285	0.5611	0.0202	0.9798	0.0001	0.0502	15	1
19	0.0149	0.9851	0.0368	0.7489	0.0150	0.9850	0.0001	0.0885	16	6
20	0.0034	0.9966	0.0425	0.9727	0.0032	0.9968	0.0001	0.4808	14	4



Şekil 1. Verilerin Saçılım Grafiği

Tablo 2 incelendiğinde FCM için üyelikler toplamı bir'e eşittir. PCM'de ise üyelikler toplamının bir olması zorunluluğu yoktur. FPCM üyelikler ile birlikte gözlemlerin küme özelliği de (t_{ij}) verilmiştir. Sıralanmış t_{ij} değerleri incelendiğinde (2, 3, 1, 8, 10, 7, 6, 9, 4, 5) gözlemlerinin birinci kümeye ait, (20, 11, 17, 19, 18, 15, 13, 12, 16, 14, 5) gözlemlerinin ikinci kümeye ait olduğu görülmektedir. Beşinci gözlemin üyeliği her iki küme içinde çok yakındır. Birinci küme için üyelik 0.5006 iken ikinci küme için üyelik 0.4994 dir. Beşinci gözlem her iki küme içinde sıralanmış t_{ij}'de en son sırada yer almaktadır, böylece Beşinci gözlemin her iki kümenin özelliğini en az taşıdığı söylenebilir. Birinci kümenin özelliğini en fazla İkinci gözlem taşımaktadır, çünkü sıralanmış t_{ij} lerde ikinci gözlem ilk sırada yer almaktadır. İkinci kümenin özelliğini en fazla taşıyan ise yirminci gözlemdir.

5. SONUÇ ve TARTIŞMA

Katı *c* ortalamaya dayalı kümelemede birimler ya bir kümeye aittir ya da değildir. Fakat birimler kümelere tam üye olmamakla birlikte belirli üyeliklerle birden fazla kümeye ait olabilirler. Bu durumda FCM yöntemi daha uygundur. Veri kümesinde aykırı değer ya da gürültü olması durumunda ise FCM yöntemleri uygun değildir. Çünkü FCM algoritmaları sıfır kırılma noktasına sahiptir. Başka bir ifadeyle tek bir aykırı değer modeli tümüyle değiştirebilir. En küçük hata kare minimizasyonu ya da pek çok kümeleme algoritması gürültüye karşı sağlam değildir. Veri kümesinde gürültü olduğunda FCM tipi algoritmalar pratik değere sahip değildir. Bu nedenle gürültüden etkilenmeyen sağlam kümeleme yöntemleri önerilmiştir.

NC yaklaşımının ana avantajı FCM algoritmasının sağlam şekli olmasıdır. PCM yaklaşımı ise FCM'in göreli üyelik probleminin üstesinden gelmek için önerilmiştir. Bu yaklaşım klasik yöntemlerdeki göreli üyelik probleminin üstesinden gelmek için geliştirilmesine rağmen verideki gürültüyü de tolere edebilmektedir. Böylece NC ve PCM yöntemleri sağlamlıkla ilgilenir. Her iki yaklaşım farklı amaçlarla geliştirilmesine rağmen

klasik bulanık *c* ortalamaya dayalı kümeleme yönteminin sahip olduğu başlangıç noktasına sahiptir. c=1 olduğunda NC ve PCM yaklaşımları $\delta^2 = \eta$ için aynıdır. NC yaklaşımında bir gürültü sınıfı varken PCM yaklaşımında c gürültü sınıfı vardır. Karma *c* ortalamaya dayalı kümelemede ise hem bulanık kümelemedeki üyeliklerin hem de bir gözlemin bir kümeye ne kadar uygun olduğunu gösteren küme özelliği (t_{ij}) birlikte ele alınmaktadır.

6. KAYNAKÇA

Barni, M., Cappellini V., and Mecocci A. (1996). Comments on "A Possibilistic Approach to Clustering". *IEEE Transactions on Fuzzy Systems* 4(3), 393-396

Dave R.N. (1993). Robust Fuzzy Clustering Algorithms. *IEEE Transactions on Fuzzy Systems* 1281-1286

Dave R.N., and Krishnapuram R. (1996). M-estimators and Robust Fuzzy Clustering. *IEEE Transactions on Fuzzy Systems* 400-404

Dave R.N., and Krishnapuram R. (1997). Robust Clustering Methods: A Unified View. *IEEE Transactions on Fuzzy Systems*. 5(2) 270-293

Dave R.N., and Sen S. (1998). Generalized Noise Clustering as a Robust Fuzzy c-M-Estimators Model. *IEEE Transactions on Fuzzy Systems* 256-260

Huber, P.J. 1981. *Robust Statistics*. John Willey & Son, 1-20, 153-243s., New York.

Krishnapuram R., and Keller M.J. (1993). A Possibilistic Approach to Clustering. *IEEE Transactions on Fuzzy Systems*. 1(2) 98-110

Krishnapuram R. and Keller M.J. (1996). The Possibilistic c-Means Algorithm: Insights and Recommendations. *IEEE Transactions on Fuzzy Systems* 4(3) 385-393

Nasraoui O. (2004). A Brief Overview of Robust Clustering Tekniques. http://archer.ee.memphis.edu./www.eee.memphis.edu./people/faculty/nasraoui/MY_TUTORIALS/RobustClustering/RobustClustering.

Pal N.R., and Bezdek C.J. (1997). A Mixed c-Means Clustering Model. *Fuzzy IEEE* 11-21

Wang L.X. (1997). *A course in fuzzy systems and control*. Prentice- Hall, 20-30s., USA.

Kamile Şanlı, lisans öğrenimini Osmangazi Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü'nden 1997 yılında, yüksek lisans öğrenimini 1999 yılında, doktora öğrenimini 2005 yılında Ankara Üniversitesi Fen Bilimleri Enstitüsü İstatistik Anabilim Dalında tamamladı. 1997-1999 yılları arasında Osmangazi Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü'nde, 1999-2005 yılları arası Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü'nde Araştırma Görevlisi olarak görev yaptı. Temmuz 2005'den bu yana Osmangazi Üniversitesi Fen-Edebiyat Fakültesi İstatistik Bölümü'nde Araştırma Görevlisi Dr. olarak görev yapmaktadır.

Ayşen Apaydın, lisans öğrenimini Hacettepe Üniversitesi İstatistik Bölümü'nde 1979 yılında, yüksek lisans öğrenimini 1981, doktorasını 1987 yılında Hacettepe Üniversitesi İstatistik Bölümü'nde tamamladı. 1996 yılında Doçent, 2002 yılında Yöneylem Araştırması Anabilim Dalında Profesör oldu. 1988 yılından bu yana Ankara Üniversitesi Fen Fakültesi İstatistik Bölümü'nde öğretim üyesi olarak görev yapmaktadır. Ayşen Apaydın evli ve iki çocuk annesidir.