

TEMATİK KARTOGRAFYADA KÜMELEME ANALİZİ

Nurbanu Afacan, İbrahim Öztuğ Bildirici

Selçuk Üniversitesi Mühendislik Fakültesi Harita Mühendisliği Bölümü Selçuklu Konya, bildirici@selcuk.edu.tr

Özet

Kümeleme analizi ile seçilen birden çok konuda ortak özelliğe sahip bireyler belirlenir. Tematik kartografyada iller, ilçeler gibi veri toplama birimleri bireyler olarak düşünülüp kümeleme analizi ile seçilen konularda ortak özelliğe sahip bölgeler belirlenebilir. Sonuçlar koropleth haritalar biçiminde sunulabilir. Bu çalışmada bu yaklaşımın uygulanabilirliği ele alınmış, iller bazında suç ve işsizlik açılarından ortak özelliğe sahip bölgeler belirlenmiştir.

Anahtar Sözcükler

Kümeleme Analizi, Tematik Kartografiya, Ward Yöntemi, Koropleth Harita

Abstract

In clustering analysis individuals with common properties can be identified in terms of two or more topics. In thematic cartography enumeration units such as administrative units can be assumed individuals, and clustering analysis can be performed with selected statistical data. The results can be presented in form of choropleth maps. In this study the applicability of this approach is focused on. An application is given on the base of provinces with two variables: unemployment and crime.

Keywords

Clustering Analysis, Thematic Cartography, Ward Method, Choropleth Map

1. Giriş

Kümeleme analizi, birey veya nesneleri benzerliklerine göre kümeler ayırmak için kullanılan çok değişkenli bir istatistik analiz tekniğidir (Tatlıdil, 1996). Kümeleme analizi sonucu oluşturulan kümeler içinde, aynı küme içinde yer alan birimler birbirlerine diğer kümenin içinde yer alan birimlerden daha çok benzerler (Atbaş, 2008). Kümeleme analizi idari birimlerde toplanmış istatistiksel verilere uygulanarak ortak özelliklere sahip bölgeler belirlenip sonuçlar koropleth haritalar biçiminde görselleştirilebilir. Bu şekilde elde edilen tematik haritalar ile belirlenen grupların ilgili kullanıcı kitlesi tarafından kolayca anlaşılması sağlanır.

Bu çalışmada kümeleme analizi Türkiye İstatistik Kurumu verileri ile iller bazında seçilen konularda gerçekleştirilmesi hedeflenmiştir. Hiyerarşik kümeleme yöntemlerinden olan Ward yöntemi seçilmiştir. Uygulamada Türkiye illerine ait suçlu kişi sayıları ve işsiz kişi sayıları temin edilerek kümeleme analizi teknikleri uygulanmıştır. Çalışma sonucunda işsizliğin suç üzerindeki etkisinin tespiti hedeflenmiştir. Kümeleme analizinin en hassas noktalarından biri küme sayısının belirlenmesidir. Bu nedenle uygulama üç farklı küme sayısı için ayrı ayrı analiz yapılmıştır. Burada amaç küme sayısının sonuçları ne oranda değiştirebileceğini gözlemlemektir.

Uygulamada TÜİK'ten elde edilen Türkiye illerine ait 2008 yılı nüfus verileri ile bu nüfusun 18 yaş üstünde kalan lise, üniversite ve üstü eğitim durumlarına sahip kişi sayıları belirlenmiştir. Yine bu illerde işlenen suçların aynı yıla ait 18 yaş üstünde kalan lise, üniversite ve üstü eğitim durumuna sahip kişilere gelen karşılığı belirlenmiştir. Böylelikle veriler 18 yaş üstü eğitilmiş ve suçlu kişi sayıları olmuştur. Suçlar 3 grupta incelenmiştir. Birinci grup suç öldürme-yaralama, ikinci grup suç dolandırıcılık-hırsızlık-yağma, üçüncü grup suç ise sahtecilik-zimmet-rüşvet olarak belirlenmiştir. Eğitim durumunun suç üzerindeki etkisinin Türkiye'deki dağılımını kümeleme analizi yöntemiyle göstermek ve bu bilgiyi tematik haritalar halinde sunmak amaçlanmıştır.

2. Materyal ve Metot

Kümeleme analizinin genel amacı, toplanan çok sayıdaki gruplanmamış gözlemlerden oluşan verilerin, birimlerin benzerliklerine göre anlamlı gruplardan oluşan özel alt kümeler bölünerek verinin indirgenmesini sağlamaktır. En az bilgi kaybıyla, daha net ve anlaşılabilir tanımlı gözlemlere sahip olunması amaçlanır (Atbaş, 2008).

Kümeleme analizi ilk kez 1939 yılında Tryon tarafından kullanılmıştır. 1960'lı yıllardan sonra kullanımı yaygınlaşmıştır (Anderberg, 1973). Kümeleme teknikleri, araştırma problemlerinde geniş bir alanda uygulanmaktadır (Hartigan, 1975). Tıpta hastalıkları, hastalık tedavilerini sınıflandırmada kümeleme çok kullanışlı bir yöntem olmuştur. Herhangi bir hastalığın hangi grupta olduğu, o grubun belirtilerinin neler olduğu vb hususların bilinmesi tedavide yararlı olmaktadır.

Birimlerin benzerliklerine göre kümeler ayrılmasında kullanılabilecek çeşitli yaklaşımlar vardır. Bu yaklaşımlardan biri, en çok benzer iki birimi aynı gruba atamakla başlayıp tüm birimlerin aynı gruba atanması ile biten hiyerarşik bir yaklaşımdır. Bir başka yaklaşım ise tüm verilerin ortalama değerlerine en yakın değerlere sahip birimlerin aynı kümeye

atanmasını esas alan yaklaşımdır. (Blashfield ve Aldenderfer, 1978). En çok kullanılan bu iki yaklaşım dışında diğer yaklaşımlar da mevcuttur. Tüm yaklaşımlarda en önemli ölçüt, kümeler arası farklar ile kümeler içi benzerliklerin maksimum olmasının sağlanmasıdır.

Hiyerarşik bir yöntem olan Ward yönteminde, grup bağlantılarından çok grup içi kareler toplamı işlenmektedir (Chatfield ve Collins, 1980). Yönteme her birinin içinde tek bir birim bulunan n tane küme ile başlanır. Yöntemin ilk basamağında her gözlem bir küme olduğundan Hata Kareler Toplamı sıfır olmaktadır (Everitt, 1974). Her aşamada iki alt küme bir sonraki seviyeyi oluşturmak için birleştirilir. Birimler değişik seviyelerde başarılı bir şekilde bir araya geldikleri "dendrogram" adlı şemada gösterilir (Dibb, 1998). Ward yöntemi aykırı noktalara duyarlı bir yöntemdir (Everitt ve ark., 2001). Bu nedenle uygulamada Ward yöntemi kullanılması uygun görülmüştür.

Kümeleme analizinin uygulama aşamaları aşağıdaki gibidir.

- Değişkenlerin seçilmesi ve veri matrisinin belirlenmesi
- Birimlerin birbirleriyle olan benzerlik ya da uzaklıklarını gösterecek uygun bir benzerlik/uzaklık ölçüsü ile benzerlik/uzaklık matrisinin oluşturulması
- Uygun bir kümeleme tekniği ile benzerlik/uzaklık matrisine göre birimlerin uygun sayıda kümelere ayrılması ve oluşturulan bu kümelerin yorumlanması
- Kümelerin yapılarının kurulan hipotezlerle test edilmesi ve gerekli analitik yöntemlerin uygulanması

3. Uygulama

Günümüzde işsizliğin suça teşvik eden bir faktör olabileceği düşünülebilmektedir. Bu genel kanının doğruluğunu irdelemek amacıyla iller düzeyinde veriler seçilmiştir. Burada amaç Türkiye için genel görünümün ne olduğunu belirlemektir.

Kümeleme analizinin ilk adımı veri kümesinin toplanmasıdır. Her il için Türkiye İstatistik Kurumu web sitesinden (<http://www.tuik.gov.tr>) elde edilen veriler ile kümeleme analizi uygulanarak işsizlik durumunun suç üzerindeki etkisi ve bu durumun iller düzeyindeki dağılımı belirlenecektir.

Küme analizinde gerekmesi halinde veri kümesi standartlaştırılır. Sayılabilir olgunun geniş değerlere sahip olması durumunda standardizasyon uygulanması gerekmektedir. Suçlar; öldürme, yaralama, dolandırıcılık, hırsızlık, yağma, sahtecilik, zimmet, rüşvet vs olmak üzere toplam şeklinde yüzde olarak belirlenmiştir. Türkiye verisi için her il ile ait suç işleyen kişi sayıları ve işsiz kişi sayıları her 100 kişiye gelen karşılığı hesaplanarak standartlaştırılmıştır.

Hiyerarşik kümeleme analizinde her gözlem ayrı bir küme olarak düşünülür ve sonra ilk benzerlik katsayılarıyla açıklandığı üzere birbirine en çok benzeyen kümeler birleştirilir. Kümeleme işlemi, Ward metodu kullanılarak Matlab yazılımında uygulanmıştır. Küme analizinde uygun küme sayısının belirlenmesi önemlidir. Amaç, her grubun göreceli olarak homojen olduğu ve diğer gruptan farklı olduğu gözlem grupları oluşturmaktır. Uygun sayıda küme seçmek için yaygın olan bir yaklaşım dendrogramdaki kırılmalara bir diğeri ise kümenin birleştirilmesi için benzerlik katsayısında oldukça büyük bir sıçramanın olduğu yerlere bakılmasıdır. Bu çalışmada küme sayıları 5, 6 ve 7 olarak seçilerek uygulamalar yapılmıştır. 5 küme seçimi ile işsizlik ve suç oranları bakımından ortak özelliklere sahip iller tablo 1'de görülmektedir.

Tablo 1: 5 Küme seçiminde oluşan gruplar

Küme 1	Adana, Mersin, İzmir, Osmaniye
Küme 2	Antalya, Aydın, Balıkesir, Burdur, Çanakkale, Denizli, Edirne, Manisa, Muğla, Nevşehir, Niğde, Uşak, Karaman, Ardahan
Küme 3	Adıyaman, Ankara, Bitlis, Hakkari, Hatay, İstanbul, Kocaeli, Kahramanmaraş, Muş, Sivas, Van, Yalova
Küme 4	Afyonkarahisar, Ağrı, Amasya, Artvin, Bilecik, Bingöl, Bolu, Bursa, Çankırı, Çorum, Elazığ, Erzincan, Erzurum, Eskişehir, Gaziantep, Giresun, Gümüşhane, Isparta, Kars, Kastamonu, Kayseri, Kırklareli, Kırşehir, Konya, Kütahya, Malatya, Ordu, Rize, Sakarya, Samsun, Sinop, Tekirdağ, Tokat, Trabzon, Tunceli, Yozgat, Zonguldak, Aksaray, Bayburt, Kırıkkale, Bartın, Iğdır, Karabük, Kilis, Düzce,
Küme 5	Diyarbakır, Mardin, Siirt, Şanlıurfa, Batman, Şırnak

Bu çalışmada elde edilen kümeler koroplet harita şeklinde görselleştirilmiştir. Koroplet haritalar; mahalle, ülke, eyalet gibi sınırlı bir saha içindeki olayların nerelerde yer aldığını açıklar ve alanlar hakkında özet bilgiler sunar (Harries,1999).

Koroplet harita ilgili idari (ya da istatistik) sınırların çizilmesi, istatistik veriden oranların ya da yüzdelerin hesaplanması, uygun sınıf aralıklarının seçilmesi ve istatistik alanların gölgelendirilmesi ve/veya renklendirilmesi ile oluşturulur (Gökgöz, 2016). Türkiye'nin coğrafi konumu sebebi ile tematik harita uygulanmasında ölçek olarak 1/7000000, datum olarak ED50 ve projeksiyon olarak Lambert konik projeksiyon seçilmiştir. Şekil 1'de 5 küme uygulamasına göre oluşturulan tematik harita örneği görülmektedir.



Şekil 1: 5 Küme seçimi ile oluşan tematik harita

4.Sonuç

Bu çalışmada Türkiye’de istatistik verilerin kümeleme analizi ile gruplanması ve seçilen verilere göre ortak özellikleri olan illerin belirlenmesi ele alınmış, işsizlik suç ilişkisi açısından benzer bölgeleri belirlemek üzere bir uygulama yapılmıştır. Burada verilen sonuçlar yürütülen bir yüksek lisans tez çalışmasının ön bulgularıdır. Çalışma devam etmekte olup tezin tamamlanması ile konu daha kapsamlı olarak incelenmiş olacaktır.

Kaynaklar

- Anderberg, M. R. (1973) *Cluster analysis for applications*, Academic Press, New York,
 Blashfield, R.K., Aldenderfer, M.S. (1978) *The literature on cluster analysis*, Multivariate Behavioral Research,13, 271-295
 Chatfield, C., Collins, A.J. (1980) *Introduction to multivariate analysis*, Chapman and Hall, London.
 Dibb, S., (1998) *Market Segmentation: Strategies for success*, Marketing Intelligence&Planning, 16/7, 394–406.
 Everitt, B., (1974) *Cluster analysis* Heinmann.London, 122.
 Everitt, B., Landau, S., Leese, M. (2001) *Cluster analysis*, Oxford University Pres,London.
 Gökgöz, T., (2016) *Kartografik Harita Üretimi Ders Notu*, <http://www.yarbis.yildiz.edu.tr/gokgoz> [Ziyaret Tarihi: 28.04.2016]
 Harries, K. (1999) *Mapping Crime: Principle and practice*. No. NCJ 178919.
 Hartigan, J.A. (1975) *Clustering Algorithms*, Wiley, New York, <http://statsoft.com/textbook/esc.html> [Ziyaret Tarihi: 28.04.2016].
 Tatlıdil, H. (1996) *Uygulamalı Çok Değişkenli Analiz*, Ankara, 329-343.