

İkili Kümeleme Algoritmalarının Parametre Seçimi için Çok Ölçütlü Karar Verme Yöntemi

Ahmet Kocatürk^{*}, Bülent Altunkaynak[†]

Gazi Üniversitesi, Fen Fakültesi, İstatistik Bölümü, 06500, Ankara, Türkiye

Öne Çıkanlar

- İkili kümeleme algoritmalarında kullanılan parametrelerin belirlenmesi.
- İkili kümelerin değerlendirilmesinde kullanılan uygunluk indeksi, benzerlik skoru, varyans ve maksimum standart alan ölçüsü.
- Parametrelerin belirlenmesinde NSGA-II çok amaçlı optimizasyon yönteminin kullanılması.
- Çok ölçütlü karar verme yöntemlerinden TOPSIS yöntemi ile çözüm kümesinin sıralanması.

Makale Bilgileri

Geliş: 31/12/2021
Kabul: 07/07/2022

Anahtar Kelimeler

İkili kümeleme,
Değerlendirme ölçüleri,
NSGA-II,
TOPSIS,
Algoritma ayarlanabilir
parametreleri

Öz

İkili kümeleme yöntemlerinde, veri matrisinde benzer satır ve sütunlar, alt kümelerine göre eş zamanlı olarak gruplandırılır. İkili kümeleme algoritmalarında kullanılan parametreler, elde edilecek ikili kümelerin belirlenmesinde oldukça önemlidir. Çünkü ikili kümeleme algoritmaları parametre değerlerine göre farklı ikili kümeler elde eder. Literatürde anlamlı ve etkili ikili kümeler elde etmek için birçok değerlendirme ölçütü bulunmaktadır. İkili kümelerin birden fazla ölçü ile değerlendirilmesi çok amaçlı bir optimizasyon problemini ortaya çıkarmaktadır. Çok amaçlı problemlerde, bir problemi optimal yapan çözüm diğer problemler için optimal değildir. Bu yüzden ideal tek bir çözüm yerine alternatif çözümler (Pareto optimal çözüm) elde edilir. Bu çalışmada, ikili kümeleme algoritmasının ayarlanabilir parametrelerini belirlemek için Pareto optimal çözüm elde edilmiştir. Pareto optimal çözüm elde etmek için en etkili yöntem Baskın Sıralı Genetik Algoritma-II (NSGA-II)'dir. NSGA-II algoritması ile elde edilen çözüm kümesinden tek bir uzlaşık çözüm seçmek için sistematik ve basit hesaplama sürecine sahip çok ölçütlü karar verme yöntemlerinden biri olan İdeal Çözüme Benzerlik Bakımından Sıralama Performansı Tekniği (TOPSIS) algoritması kullanılmıştır. Bu çalışmada ikili kümeleme algoritmasının ayarlanabilir parametrelerini belirlemek için yapay ve gerçek veri matrisleri kullanılmış ve değerlendirme ölçüleri için R fonksiyonları oluşturulmuştur. Her bir değerlendirme ölçüsü ayrı ayrı dikkate alınarak ikili kümeleme algoritmasının ayarlanabilir parametreleri belirlenmiştir. Ayrıca değerlendirme ölçüleri eşit önem derecesine göre çok ölçütlü karar yöntemi ile elde edilerek ikili kümeleme algoritmalarının ayarlanabilir parametreleri karşılaştırılmıştır. Her bir değerlendirme ölçüsüne göre farklı sonuçlar elde edildiği için çok ölçütlü karar verilmesi daha anlamlı ikili kümeler elde edilmesini sağlamıştır.

Multi-Criteria Decision Making Method for Parameter Selection of Biclustering Algorithms

Highlights

- Determining the parameters used in biclustering algorithms.
- The relevance index, similarity score, variance and maximum standard area measure used in the evaluation of biclusters.
- Use of NSGA-II multi-objective optimization method in the determination of parameters.
- Sorting the solution set with the TOPSIS method, one of the multi-criteria decision making methods.

Abstract

In biclustering methods, similar rows and columns in the data matrix are grouped simultaneously according to their subsets. The parameters used in biclustering algorithms are very important in determining the biclusters to be obtained. Because biclustering algorithms obtain different biclusters according to parameter values. There are many evaluation measures in the literature to obtain meaningful and effective biclusters. Evaluating biclusters with more than one measure reveals a multi-objective optimization problem. In multi-objective problems, the solution that makes one problem optimal is not optimal for other problems. Therefore, alternative solutions (Pareto optimal solution) are obtained instead of a single ideal solution. In this study, a Pareto optimal solution was obtained to determine the tuning parameters of the biclustering algorithm. The most effective method to obtain a Pareto optimal solution is the Non-dominated Sorting Genetic Algorithm-II (NSGA-II). Technique for Order Preference by Similarity to Ideal Solution (TOPSIS algorithm), which is one of the multi-criteria decision making methods with a systematic and simple calculation process, was used to select a single compromise solution from the solution set obtained by the NSGA-II algorithm. To determine the tuning parameters of the biclustering algorithm, artificial and real data matrices were used and R functions were created for the evaluation criteria. The tuning parameters of the biclustering algorithm were determined by considering each evaluation measure separately. In addition, the evaluation measures were obtained with the multi-criteria decision method according to the equal importance level and the tuning parameters of the biclustering algorithms were compared. Since different results were obtained according to each evaluation measure, making a multi-criteria decision led to more meaningful biclusters.

Article Info

Received: 31/12/2021
Accepted: 07/07/2022

Keywords

Biclustering,
Evaluation measures,
NSGA-II,
TOPSIS,
Algorithm tuning
parameters

Makale, Creative Commons 4.0 (CC BY NC SA) uluslararası lisansı altında açık erişim olarak yayımlanmaktadır.

* Sorumlu Yazar/Corresponding Author: Ahmet Kocatürk, ahmetkocaturk@gazi.edu.tr, [†] Makale basım sırasında vefat eden hocamız için eklenmiştir.

1. GİRİŞ

Çok değişkenli analiz tekniklerinden biri olan kümeleme analizinin öncelikli amacı, birey ya da nesnelerin temel özelliklerini dikkate alarak onları gruplandırmaktır. Kümeleme analizi yöntemlerinde çözülecek probleme göre farklı algoritmalar mevcuttur. Gruplandırılmak istenen veri yapısı matris formunda olduğunda geleneksel kümeleme analizi yöntemleri sadece satır kümelerini veya sütun kümelerini elde etmeyi amaçlar. Ancak veri matrisinde hem satır hem de sütun kümelerinin eş zamanlı kümelenebilmesi için son zamanlarda yaygın bir şekilde kullanılan ikili kümeleme yöntemleri geliştirilmiştir [1].

İkili kümeleme yöntemlerinin amacı veri matrisinin bir sütun alt kümesine göre benzer olan satırları veya bir satır alt kümesine göre benzer olan sütunları gruplamaktır. İkili kümeleme yöntemleri pazarlama bölümlenmesinde [2], metin madenciliğinde [3] veya veri madenciliği problemlerinde [4] kullanıldığı gibi aslında daha çok biyolojik veri analizlerinde kullanılmıştır. İlk ikili kümeleme yöntemlerinin orijinal çalışma alanları gen ifade verileridir. Gen ifadesini belirlemek, protein etkileşimini incelemek ve mikrodizi veri analizi alanlarında kullanımı yaygınlaştıkça ikili kümeleme yöntemleri önem kazanmıştır.

Literatürde ikili kümeleme için birçok farklı sezgisel algoritmalar geliştirilmiştir. İlk olarak Cheng ve Church (2000) tarafından Ortalama Karesel Artık Değeri (Mean Squared Residue, MSR)'nin minimum değer almasına dayalı Cheng ve Church (CC) algoritması önerilmiştir [5]. Lazzeroni ve Owen (2002) tarafından önerilen Plaid algoritması, iki yönlü ANOVA modellerini içeren bir yapıya sahip olmasının yanında ikili kümelerin örtüşmesine de izin verir [6]. İkili kümeleri yinelemeli bir şekilde keşfeden deterministik olmayan yöntem ISA algoritması Bergmann ve diğerleri (2003) tarafından geliştirilmiştir [7]. OPSM (Order-Preserving Submatrix) algoritması ise ikili kümeleri belirlemek için olasılıksal bir modeli kullanarak yeniden sıralama yaklaşımına dayalı deterministik bir yöntemdir [8]. Prelic ve diğerleri (2006) tarafından ikili kümeleme yaklaşımlarının çoğunun temel özelliklerini barındıran iki değerli referans model Bimax algoritması önerilmiştir [9]. Veri yapısının nitel veya yarı nicel ölçümlerinin bir kombinasyonu olarak diğer algoritmalara göre daha genel bir biçimde ikili kümeler elde eden QUBIC (Qualitative Biclustering Algorithm) algoritması hesaplama süresinde diğer algoritmalara göre oldukça etkilidir [10]. Veri matrisindeki satır ve sütunlar arasındaki doğrusal bağlantıları dikkate alan ve normal dağılımlı olmayan veri yapılarında da etkili olan FABIA (Factor Analysis for Bicluster Acquisition) algoritması Bayes tekniklerine dayalı bir algoritmadır [11]. Bayes ikili kümeleme modelinin genişletilmiş bir versiyonu olan PPM (Plaid Penalized Method) algoritması genel bir örtüşme yapısını dikkate alır [12]. Büyük boyutlu veri matrisleri üzerinde etkili olan ENS (Ensemble Neighborhood Search) algoritması diğer yöntemlere göre daha esnek ve uyarlanabilirliğine sahiptir [13]. Kayıp değerlerin çok olduğu veri matrislerine uygun olan NCBI (Novel Correlation Based Imputing Technique) algoritması, korelasyon tabanlı kayıp değer atama tekniğine dayalıdır [14]. Bu algoritmaların bazıları sistematik arama yaklaşımına, bazıları ise stokastik arama veya metasezgisel yaklaşımına göre ikili kümeleri elde etmektedir. Metasezgisel yaklaşımına göre ikili kümeleri elde eden algoritmalar genetik algoritmalar veya benzetimli tavlama gibi sezgisel yöntemleri kullanan evrimsel ve hibrit algoritmalar [15].

İkili kümeler elde edilirken algoritmalarda kullanılacak parametrelerin belirlenmesi oldukça önemlidir. Çünkü ikili kümeleme algoritmaları parametrelerine göre farklı ikili kümeler elde etmektedir. Örneğin, Prelic ve diğerleri (2006), Liu ve Wang (2006), Al-Akwaa ve diğerleri (2009), Chia ve Karuturi (2010), Padilha ve Campello (2017) ve Karim ve diğerleri (2019) tarafından yapılan çalışmalarda algoritmaların parametreleri farklı seçildiğinden "Saccharomyces cerevisiae" verisinden elde edilen ikili kümeler farklılık göstermiştir [9, 16-20]. İkili kümeleme algoritmalarının performanslarının karşılaştırılmasında sonucu doğrudan etkileyen faktör, parametrelerin belirlenmesidir.

Bu çalışmada, ikili kümeleme algoritmasının ayarlanabilir parametreleri belirlenirken sayısal niteliklerin olduğu veri yapıları dikkate alınmıştır. Bununla birlikte veri yapıları satırlar ve sütunlardan oluşan matris formundadır. Bu veri yapısına en uygun gerçek hayat problemleri gen açıklama verileridir. Gen açıklama verilerinde on binlerce genin çeşitli koşullar altındaki ifade seviyeleri bulunmaktadır. Bu biçimde tanımlı büyük boyutlu veriden anlamlı bilgiler çıkarılması gen çalışmalarında önemli bir etkiye sahiptir [21]. Gen açıklama verilerinde benzer ifade yapılarına göre gen gruplarını belirlemek için ikili kümeleme algoritmaları kullanılır.

Anlamli ve etkili ikili kümeler elde etmek için küme büyüklüğü, satır varyansı ve ortalama karesel artık skoru gibi değerlendirme ölçüleri kullanılmaktadır. İkili kümelerin tek bir ölçü yerine birden fazla ölçü ile değerlendirilmesi daha anlamlı olacaktır. Böylece ikili kümelerin değerlendirilmesi çok amaçlı bir optimizasyon problemini ortaya çıkarmaktadır. Literatürde çok amaçlı optimizasyon problemlerin çözümünde yaygın olarak kullanılan yapay sinir ağları, genetik algoritmalar veya tavlama benzetimi gibi sezgisel yöntemler bulunmaktadır [22, 23]. Sezgisel yöntemler her bir çözümü bulmak için birçok kez tekrarlandığından hesaplama karmaşıklığına neden olmaktadır. Bu yüzden çok amaçlı sezgisel yöntemler geliştirilmiştir. Çok amaçlı optimizasyon problemlerini çözmek için geliştirilen en etkili çok amaçlı sezgisel yöntemlerden biri Baskın Sıralı Genetik Algoritma-II (Non-dominated Sorting Genetic Algorithm-II, NSGA-II)'dir [24-28]. Bu çalışmada, ikili kümeleme algoritmasının ayarlanabilir parametreleri için birden fazla değerlendirme ölçütü dikkate alındığından ideal tek bir çözüm yerine alternatif çözümler kümesi (Pareto optimal çözüm) elde edilmiştir. Pareto optimal çözüm için en etkili yöntem NSGA-II'dir [29].

Karar verme yöntemlerinden literatürde yaygın olarak kullanılan Hwang ve Yoon (1981) tarafından önerilen TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) yöntemi [30, 31], Pamucar ve Cirovic (2015) tarafından önerilen MABAC (Multi-Attributive Border approximation Area Comparison) yöntemi [32], Liu ve diğerleri (2016) tarafından önerilen ELECTRE (ELimination Et Choice Translating Reality) yöntemi [33] vb. bulunmaktadır. NSGA-II algoritması ile elde edilen çözüm kümesi içerisinde uzlaşık tek bir çözüm seçmek için çok ölçütlü karar verme yöntemlerinden yaygın olarak kullanılan, hesaplama süresi ve sistematik süreci basit olan TOPSIS algoritması kullanılmıştır.

Çalışmanın ikinci bölümünde ikili kümeleme yöntemlerinden CC algoritması tanıtılmaktadır. Üçüncü bölümde ise ikili kümelerin değerlendirilmesinde kullanılabilecek varyans ölçüsü (VAR), uygunluk indeksi (RI), maksimum standart alan (MSA) ve benzerlik skoru (SS) ölçüleri açıklanmıştır. Dördüncü bölümde NSGA-II, TOPSIS ve önerilen yöntemlerden bahsedilmiştir. Önerilen yöntem için NSGA-II ve TOPSIS yöntemleri kullanılarak R program kodları ile fonksiyonlar oluşturulmuştur. Beşinci bölümde yapay ve gerçek veri seti kullanılarak yapılan uygulamalar ve sonuçları yer almaktadır. Son bölümde çalışmadan elde edilen sonuçlar verilmiş ve önerilerde bulunulmuştur.

2. İKİLİ KÜMELEME YÖNTEMLERİ

Veri matrislerinden ikili kümeler elde edilirken yapılan analizler çalışmanın amacına göre farklılık gösterebilmektedir. Örneğin araştırmacı, veri matrisinde satırlar (gen verileri için genler, karar matrisi için alternatifler, vb.) veya sütunların (gen verileri için koşullar, karar matrisi için kriterler, vb.) kümelenebilir (Çizelge 1). Bunun için klasik kümeleme algoritmaları kullanılabilir. Ancak her iki durumu aynı anda yapabilmek için (hem satırları hem de sütunları eşzamanlı olarak kümelemek) ikili kümeleme algoritmaları geliştirilmiştir. İkili küme yapısı (2.1)'deki gibidir.

Çizelge 1. Çalışmada kullanılan sembol ve notasyon listesi

Sembol	Tanım	Sembol	Tanım
b_{ij}	İkili kümenin j . sütun ortalaması	σ_i	İkili kümenin i . satır standart sapması
b_{iJ}	İkili kümenin i . satır ortalaması	d_{ij}	i . ve j . satırlar arasındaki uzaklık
b_{IJ}	İkili kümenin genel ortalaması	d_{avg}	İkili küme elemanları arasındaki ortalama uzaklık
σ_{ij}^2	İkili kümenin j . sütun yerel varyansı	s_{ij}	İkili küme benzerlik matrisinin i . satır ve j . sütun elemanı
σ_j^2	İkili kümenin j . sütun genel varyansı	$m_j(\hat{B})$	Tahmin edilen ikili kümenin j . sütunu için alt sınır değeri
μ_i	İkili kümenin i . satır ortalaması	$M_j(\hat{B})$	Tahmin edilen ikili kümenin j . sütunu için üst sınır değeri

$$B = \begin{pmatrix} b_{11} & b_{12} & \cdot & \cdot & \cdot & b_{1|J|} \\ b_{21} & b_{22} & \cdot & \cdot & \cdot & b_{2|J|} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ b_{|I|1} & b_{|I|2} & \cdot & \cdot & \cdot & b_{|I||J|} \end{pmatrix}. \quad (2.1)$$

Burada $b_{1|J|}$ birinci satır (gen) ve $|J|$ inci sütun (koşul veya örneklem) elemanını, $b_{|I|1}$ $|I|$ inci satır ve birinci sütun elemanını ve $b_{|I||J|}$ $|I|$ inci satır ve $|J|$ inci sütun elemanını göstermektedir. $|I|$ ve $|J|$ sırası ile satır ve sütun toplam sayılarını belirtmektedir. İkili kümeleme algoritmalarında yaygın olarak kullanılan bazı formüller (2.2)'deki gibi verilebilir

$$b_{ij} = \frac{1}{|I|} \sum_{i=1}^{|I|} b_{ij}, \quad b_{iJ} = \frac{1}{|J|} \sum_{j=1}^{|J|} b_{ij}, \quad b_{IJ} = \frac{1}{|I||J|} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} b_{ij}. \quad (2.2)$$

2.1. CC Algoritması

CC algoritması Ortalama Karesel Artık Değeri (Mean Squared Residue, MSR)'nin minimum değer almasına dayalıdır. Bir ikili kümenin anlamlılığı, küme içerisindeki elemanların birbirine olan benzerliğini hesaplayan MSR ölçüsü ile belirlenir. MSR ölçüsü (2.3)'teki gibi hesaplanır [5]

$$MSR(I, J) = \frac{1}{|I||J|} \sum_{i \in I, j \in J} (b_{ij} - b_{iJ} - b_{iI} + b_{IJ})^2. \quad (2.3)$$

MSR değerinin küçük olması küme elemanlarının benzerliğinin yüksek olduğunu gösterir. Bununla birlikte MSR değerinin büyük olması ise küme elemanların benzerliğinin düşük olduğunu ifade eder.

CC algoritmasında kullanılan iki parametre değeri vardır. Bunlardan birincisi olan δ (delta) parametresi maksimum kabul edilebilecek MSR değerini gösterir. İkinci olarak kullanılan α (alfa) parametresi ise çoklu satır veya sütun silme işlemindeki eşik değerini belirler. CC algoritmasında δ parametresi önceden belirlenerek MSR değeri bu parametre değeri ile karşılaştırılır. MSR değeri δ parametresinden küçük olduğunda elde edilen ikili küme δ -ikili küme (δ -bicluster) olarak ifade edilir. Algoritmanın temel amacı minimum değerli MSR değerini maksimum büyüklükteki ikili kümeler ile bulmaktır.

CC algoritması başlangıç olarak veri matrisini bir ikili küme olarak alır. Daha sonra veri matrisinde satır ve sütun silme işlemleriyle devam eder. Bu işlemde sonra MSR değerinin δ parametresi değerini geçmeyecek şekilde satır veya sütun ekleme işlemiyle ikili kümelerin satır veya sütun sayıları artırılmaya çalışılır. MSR değeri δ parametresi değerine yaklaşıncaya kadar satır ve sütun ekleme işlemine devam edilir. CC algoritması ilk ikili kümeyi elde ettikten sonra veri matrisinden bu ikili kümeyi çıkararak ikinci iterasyona geçer. Veri matrisindeki tüm ikili kümeler elde edilinceye kadar iterasyon devam eder.

3. İKİLİ KÜME DEĞERLENDİRME ÖLÇÜLERİ

3.1. Varyans Ölçüsü

İkili kümelerde değerlendirme ölçüsü olarak kullanılan varyans ölçüsü (*VAR*), sabit değerli ikili kümelerde etkili bir ölçü türüdür [34]. İkili küme elemanlarının varyans toplamlarına dayanan bu ölçü (3.1)'deki gibi hesaplanır

$$VAR(B) = \frac{1}{|I||J|} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} (b_{ij} - b_{Ij})^2 . \quad (3.1)$$

İkili kümelerin anlamlı ve etkili olabilmesi için *VAR* ölçüsünün minimum olması gerekir. İkili kümelerin karşılaştırılmasında kullanılan *VAR* ölçüsü küçük olan ikili küme diğerine göre daha anlamlı ve etkili olduğu söylenebilir.

3.2. Uygunluk İndeksi

İkili kümelerin sütun elemanlarının uygunluklarının toplamına dayalı olan uygunluk indeksi ölçüsü (*RI*), Yip ve diğerleri (2004) tarafından önerilmiştir [35]. Bu ölçü özellikle sabit değerli ikili kümelerde iyi sonuçlar vermektedir. *RI* ölçüsü her bir sütun değeri için ayrı ayrı hesaplanmaktadır.

$$RI_{ij} = 1 - \frac{\sigma_{ij}^2}{\sigma_j^2} . \quad (3.2)$$

Burada σ_{ij}^2 (yerel varyans) ikili kümedeki sütun değerlerinin varyansı ve σ_j^2 (genel varyans) tüm veri setindeki sütun değerlerinin varyansıdır. Yerel varyans küçüldükçe indeks değeri büyür. Olabilecek en yüksek uygunluk indeksi yerel varyansın sıfır olması durumunda gerçekleşir. Her bir sütun için hesaplanan *RI* ölçülerinin ortalaması ikili kümenin *RI* ölçüsünü temsil etmektedir. Buna göre en yüksek *RI* ölçüsüne sahip ikili küme daha anlamlı olacaktır.

3.3. Maksimum Standart Alan Ölçüsü

Standartlaştırma tabanlı ölçümlerden biri olan maksimum standart alan (*MSA*) ölçüsü veri matrisindeki elemanların ortak bir aralığa ölçeklendirmesini sağlayıp sayısal değerlerden ziyade eğilimi karakterize etmektedir [34]. Bir ikili kümenin elemanlarının standartlaştırılması (3.3)'teki gibidir

$$\hat{b}_{ij} = \frac{b_{ij} - \mu_i}{\sigma_i} , 1 \leq i \leq |I| , 1 \leq j \leq |J| . \quad (3.3)$$

Burada μ_i ve σ_i , sırasıyla *i* inci satırın ortalamasını ve standart sapmasını temsil etmektedir. Her bir satır için standartlaştırılmış değerler hesaplandıktan sonra veri matrisindeki maksimum ve minimum değerler arasındaki ölçüm *MSA* ölçüsünde kullanılır. Buna göre ikili kümedeki her sütun için hesaplanan alt sınır ve üst sınır sırasıyla $m_j(\hat{B})$ ve $M_j(\hat{B})$ hesaplanır

$$m_j(\hat{B}) = \min_i \hat{b}_{ij} , M_j(\hat{B}) = \max_i \hat{b}_{ij} . \quad (3.4)$$

Bu sınırlar kullanılarak standartlaştırılmış ikili küme elemanlarının her sütun değeri için hesaplanan alanların toplamı *MSA* ölçüsünü oluşturur

$$MSA(B) = \sum_{j=1}^{|J|-1} \left| \frac{M_j(\hat{B}) - m_j(\hat{B}) + M_{j+1}(\hat{B}) - m_{j+1}(\hat{B})}{2} \right|. \quad (3.5)$$

Eğer ikili küme içerisindeki satırlar mükemmel bir uyum içerisinde ise MSA ölçüsü sıfıra eşit olacaktır. Aksine MSA ölçüsü değeri arttıkça ikili kümenin anlamlılığı ve etkinliği azalacaktır. Bu yüzden ikili kümenin anlamlı olması için $m_j(\hat{B})$ ve $M_j(\hat{B})$ değerleri birbirine yakın olmalıdır.

3.4. Benzerlik Skoru Ölçüsü

Liu ve Wang (2007) tarafından önerilen benzerlik skoru (SS) ölçüsü ikili kümenin iki satırı arasındaki benzerliğini ölçmeye dayalıdır. İlk olarak referans satır elemanı belirlenir ve iki satır arasındaki uzaklık (3.6)'daki gibi hesaplanır [19]

$$s_{ij} = \begin{cases} 0 & , \text{eğer } d_{ij} > \alpha \times d_{avg} \\ 1 - \frac{d_{ij}}{\alpha \times d_{avg}} + \beta & , \text{diğer durumlarda} \end{cases} \quad (3.6)$$

Burada d_{avg} ikili küme elemanlarının ortalama uzaklık ölçüsünü gösterir ve (3.7)'deki gibi hesaplanır

$$d_{avg} = \frac{\sum_{i \in I, j \in J} d_{ij}}{|I||J|}, \quad d_{ij} = |b_{ij} - b_{i^*j}|. \quad (3.7)$$

Uzaklık ölçüsü d_{ij} i inci satır ve referans satır elemanı arasındaki mutlak fark ile hesaplanır. İkili kümenin benzerlik skoru hesaplanırken ilk olarak her satır ve her sütun için sırasıyla $s(i, J)$ ve $s(I, j)$ değerleri (3.8)'deki gibi bulunur

$$s(i, J) = \sum_{j \in J} s_{ij}, \quad s(I, j) = \sum_{i \in I} s_{ij}. \quad (3.8)$$

Bu eşitlikler kullanılarak ikili küme için benzerlik skoru ölçüsü (3.9)'daki gibi hesaplanır

$$s(B) = s(I, J) = \min \left\{ \min_{i \in I} s(i, J), \min_{j \in J} s(I, j) \right\}. \quad (3.9)$$

Benzerlik skoru ölçüsü yüksek olan ikili küme daha anlamlı ve etkili olacaktır.

4. ÖNERİLEN YÖNTEM

4.1. Çok Amaçlı Optimizasyon

Çok amaçlı optimizasyon problemleri, belirli kısıtlar altında iki veya daha fazla amacın bulunduğu durumlardır. Ayrıca çözümlerin farklı amaçlar için değerlendirilmesini ve hangi çözümün seçilmesi gerektiğini de inceler. Araştırmacının amacına göre tek bir çözümün bulunabileceği gibi bazı durumlarda farklı amaçlar için birden fazla çözümün bulunduğu çözüm kümesini de elde eder. Optimize edilmesi gereken birden fazla çelişen amaç fonksiyonları olduğunda tek bir optimal çözüm yoktur. Ancak tüm amaç fonksiyonları için elde edilen çözüm kümesi vardır. Bu çözüm kümesine pareto optimal küme denir [15]. Çok amaçlı optimizasyon problemleri aşağıdaki gibi formüle edilebilir

$$\min_x f(x_1, x_2, \dots, x_n) = [f_1(x_1, x_2, \dots, x_n) \quad f_2(x_1, x_2, \dots, x_n) \quad \dots \quad f_m(x_1, x_2, \dots, x_n)].$$

Burada n tane x karar değişkeni bulunmaktadır. $f(x_1, x_2, \dots, x_n)$ ise m tane amaç fonksiyonunu ifade etmektedir.

4.1.1. Çok amaçlı ikili kümeleme

İkili kümeleme algoritmalarından anlamlı ve etkili ikili kümeler elde etmek için ikili kümelerin homojen yapıda ve anlamlı büyüklükte olması gibi farklı değerlendirme ölçülerine bakılır. İkili kümelerde benzerlik ölçütü olarak kullanılan *MSR* ölçüsü optimize edilirken yalnızca bir satır ve bir sütuna sahip ikili küme çözümü optimaldir. Ancak ikili küme boyutu arttığında *MSR* ölçüsü değeri de artacaktır. Böylece ikili kümelerde birden fazla değerlendirme ölçüsü dikkate alındığında çok amaçlı bir optimizasyon problemi ortaya çıkar. Literatürde bu problemlerin çözümü için çok amaçlı ikili kümeleme algoritma modelleri önerilmiştir [36-38].

Önerilen modelde birbiriyle çelişen dört farklı değerlendirme ölçüsü (*VAR*, *RI*, *MSA* ve *SS*), amaç fonksiyonu olarak optimize edilmek için kullanılmıştır

$$f_1 = \frac{1}{|I||J|} \sum_{i=1}^{|I|} \sum_{j=1}^{|J|} (b_{ij} - b_{IJ})^2, \quad (4.1)$$

$$f_2 = 1 - \frac{\sigma_{Ij}^2}{\sigma_j^2}, \quad (4.2)$$

$$f_3 = \sum_{j=1}^{|J|-1} \left| \frac{M_j(\hat{B}) - m_j(\hat{B}) + M_{j+1}(\hat{B}) - m_{j+1}(\hat{B})}{2} \right|, \quad (4.3)$$

$$f_4 = s(I, J) = \min \left\{ \min_{i \in I} s(i, J), \min_{j \in J} s(I, j) \right\}. \quad (4.4)$$

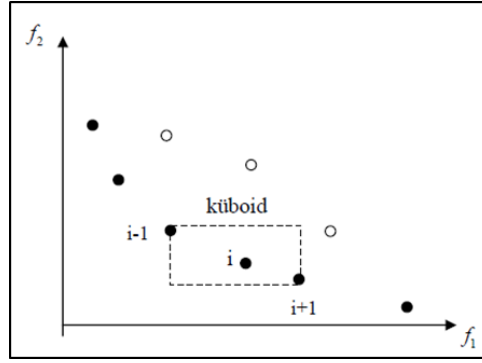
Burada f_1 amaç fonksiyonu için varyans toplamlarına dayalı *VAR* ölçüsü, f_2 amaç fonksiyonu için sütun elemanları uygunluklarına dayalı *RI* ölçüsü, f_3 amaç fonksiyonu için standartlaştırma tabanlı *MSA* ölçüsü ve f_4 amaç fonksiyonu için ikili kümenin iki satırı arasındaki benzerliği ölçen *SS* ölçüsü tanımlanmıştır. Her bir amaç fonksiyonu içerisinde ikili kümeyle ait ortak karar değişkenleri bulunmaktadır. *CC* algoritmasının δ (delta) ve α (alfa) parametresi başlangıç kromozomları olarak ele alınmıştır. Ortak karar değişkenleri δ ve α parametresi için alt sınır ve üst sınır değerleri arasında *NSGA-II* yöntemi için arama uzayı belirlenir. Optimal çözümler ile elde edilen parametre değerlerine göre ikili kümeler belirlenir. Belirlenen ikili kümeleri değerlendirmede ise f_1 , f_2 , f_3 ve f_4 amaç fonksiyonları kullanılır. Amaç fonksiyon değerlerine göre *NSGA-II* algoritması δ ve α parametresi için tekrar optimal çözüm için arama yapar. Elde edilen yeni δ ve α parametresi kullanılarak *CC* algoritması ile yeni ikili kümeler elde edilir. Elde edilen yeni ikili kümeleri değerlendirmek için tekrardan amaç fonksiyon değerleri hesaplanır. Anlamlı ikili kümeler elde etmek için f_1 ve f_3 fonksiyonları minimum olması gerekirken f_2 ve f_4 fonksiyonları maksimum olması gerekir. *NSGA-II* algoritması genel olarak bilgisayar programlama dillerinde optimizasyon süreci olarak varsayılan olarak bütün amaç fonksiyonlarını minimize edilmesi biçiminde oluşturulmuştur. Bu yüzden f_2 ve f_4 fonksiyonları maksimizasyon probleminden minimizasyon problemine dönüştürülerek *NSGA-II* algoritmasına dahil edilmiştir.

NSGA-II yöntemi kullanılarak *CC* algoritmasının ayarlanabilir parametreleri için Pareto optimal çözüm belirlenir. İkili küme değerlendirmesi için *VAR*, *RI*, *MSA* ve *SS* ölçütleri belirlendikten sonra *TOPSIS* yöntemi ile Pareto optimal çözüm içerisinden uzlaşık tek bir çözüm seçilir. Böylece en anlamlı ve etkili ikili küme için *CC* algoritmasının ayarlanabilir parametre değerleri bulunur.

4.1.2. NSGA-II yöntemi

Çok amaçlı optimizasyon problemlerinde sezgisel bir yöntem olan NSGA-II yönteminde, genetik algoritmaya göre ek olarak baskınlık sıralaması ve kalabalık uzaklığı hesaplanır [29]. İlk olarak başlangıç popülasyonundan elde edilen çözümler Pareto optimal çözümdeki üstünlüklere göre sıralanır. Sıralama için kullanılan hızlı baskınlık sıralama algoritmasında her bir çözüm diğer çözümlerle karşılaştırılıp popülasyondaki bireylerin baskınlıklarına bakılır. İlk baskın yüzeydeki çözümler popülasyondan çıkarılarak ikinci sıradaki baskın yüzey çözümleri için işlem tekrarlanarak devam eder. Böylece popülasyondaki bireyler farklı baskınlık kümelerine göre sınıflandırılır. Popülasyondaki her bir eleman için baskınlık sayısı ve baskın olduğu çözüm kümesi belirlenir. Baskınlık sayısı ilk baskın yüzeydeki elemanlar için 0'dır. Baskınlık sayısı 0 olan elemanların kümesinde karşılaştırma yapılır ve hızlı baskınlık sıralama algoritması süreci devam eder.

NSGA-II yönteminde kullanılan bir diğer hesaplama kalabalık uzaklığı yaklaşımıdır. Çözüm kümesi içerisindeki yayılımı ve çeşitliliği sağlamak için kullanılan kalabalık uzaklığında baskın yüzeylerdeki çözümlerin fonksiyon değerleri arasındaki uzaklık belirlenir. Uzaklık ölçüsü olarak genellikle Öklid uzaklık ölçüsü kullanılmaktadır. Kalabalık uzaklığının hesaplanması iki amaç fonksiyonu için Şekil 1'de gösterilmiştir.



Şekil 1. Kalabalık uzaklığının hesaplanması (iki amaç fonksiyonu için) [30]

Baskın yüzeydeki çözümler içerisindeki bir çözümün kalabalık uzaklığı, aynı baskın yüzeydeki komşu çözümler ile arasındaki uzaklık hesaplanılarak bulunur. Komşu çözümlerin koordinat düzlemindeki köşeleri kullanılarak elde edilen kenarların uzunluğu tahmin edilir. Tahmin edilen kenar uzunluklarına kalabalık uzaklığı denir.

NSGA-II algoritmasında kullanılan bir diğer hesaplama ise kalabalıklaştırılmış turnuva seçim operatörüdür. Kalabalıklaştırılmış turnuva seçiminde eşleşme havuzu oluşturulur. İlk olarak iki birey rastgele seçilir ve baskınlık sıraları karşılaştırılır. Karşılaştırılan bireylerden birinin baskınlık sırası diğerine göre küçük veya eşitse çözüm için o birey tercih edilir. Ayrıca turnuva seçiminde kalabalık uzaklığına da bakılır. Eğer bireylerden birinin kalabalık uzaklık değeri diğerine göre büyükse çözüm için o birey tercih edilir. Turnuva seçiminin tamamlanması için başlangıç popülasyonu boyutuna ulaşılması gerekir.

NSGA-II algoritmasının genetik operatörlerinden biri olan çaprazlama operatörü, iki yeni kromozom üretmek için ebeveynler arasında gen alışverişini gerçekleştirir. Genellikle tek nokta, çift nokta, sıralı ve tekdüze çaprazlama yöntemleri kullanılır. Genetik operatörlerden bir diğeri ise mutasyon operatörüdür.

Bir kromozomda gen değişimi dışında başka nedenlerden dolayı oluşan kalıtsal değişimler mutasyon ile meydana gelir. Optimizasyon probleminin yerel çözümde kalmasını engellemek için kullanılır. Başka bir ifadeyle popülasyonun yerel optimumda durgunlaşmasını önlemede yardımcı olur. Mutasyona uğrayacak gen dizileri rastgele seçilir. İki değerli diziler için bit değerinin tersi alınarak gerçekleşir. Gerçek değerli diziler için bit değerinden küçük bir sayı çıkarılarak veya eklenerek gerçekleşir [29].

NSGA-II algoritmasının genel yapısı Türkşen (2011) (Şekil 2) tarafından yapılan çalışmada özetlenmiştir [30]. Buna göre algoritmanın adımları aşağıdaki gibidir:

Adım 1

Her bir $p \in P$ için;

- $S_p = \emptyset$, p çözümünün baskın olduğu çözüm kümesi ve $n_p = 0$, p çözüme baskın çözümler sayısı tanımlanır.

Her bir $q \in P$ için;

- p çözümü q çözüme baskın ise $S_p = S_p \cup \{q\}$.

$n_p = 0$ ise p çözüme baskın çözüm yoktur ve p çözümü birinci yüzeye aittir. $F_1 = F_1 \cup \{p\}$.

Adım 2

Her bir $p \in P$ için Adım 1 tekrarlanır.

Adım 3

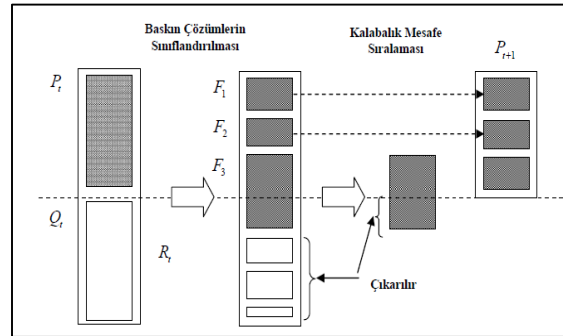
Yüzey sayacı $i=1$ olarak alınır.

- $F_i \neq \emptyset$ ise $(i+1)$. yüzey için elemanların toplandığı küme $Q = \emptyset$ alınır.
- F_i yüzeyindeki her p ve S_p kümesindeki her bir q için baskınlık sayacı azaltılır. $n_q = n_q - 1$.

Adım 4

$n_q = 0$ ise ardışık yüzeydeki hiçbir çözüm q çözüme baskın değildir. $Q = Q \cup \{q\}$.

- Yüzey sayacı bir artırılır. $i = i + 1$.
- Yeni yüzey Q kümesine atanır. $F_i = Q$.



Şekil 2. NSGA-II'de yeni popülasyonun oluşturulması [30]

4.2. Çok Ölçütlü Karar Verme

Çok ölçütlü karar verme yöntemi, birden fazla ölçütün bulunduğu durumlarda farklı özellikteki seçeneklerden bir veya daha fazlasını seçmek için kullanılan bir yöntemdir. Ölçütler eşit önem derecesine veya farklı ağırlıklandırma yapılarak seçenekler/alternatifler içerisinde sıralama veya sınıflandırma yapılabilir. Çok ölçütlü karar verme yöntemlerinden bazıları ağırlıklı toplam veya çarpım yöntemi, TOPSIS, MABAC, AHP ve ELECTRE yöntemleridir [31-33]. Genel olarak çok ölçütlü karar verme yöntemlerinde ilk olarak ölçütler ve seçenekler/alternatifler belirlenir. Daha sonra ölçütler için ağırlık katsayıları araştırmanın amacına göre belirlenir. Seçilecek olan alternatifler bu ölçütlere göre değerlendirilip her bir alternatif için sıralama yapılır. Bu çalışmada çok ölçütlü karar verme yöntemlerinden yaygın olarak kullanılan TOPSIS yöntemi seçilmiştir.

4.2.1. TOPSIS yöntemi

Hwang ve Yoon (1981) tarafından önerilen TOPSIS (Technique for Order Preference by Similarity to an Ideal Solution) yöntemi, pozitif ideal çözüme en yakın ve negatif ideal çözüme en uzak olan alternatifleri seçmeye dayalıdır [31]. Maliyet ölçütünü minimum ve fayda ölçütünü maksimum yapan çözüm pozitif ideal çözümdür. Negatif ideal çözüm ise maliyet ölçütünü maksimum ve fayda ölçütünü minimum yapan çözümdür. TOPSIS yönteminde pozitif ideal çözüme en yakın mesafedeki alternatif ile negatif ideal çözüme en uzak mesafedeki alternatif aynı anda değerlendirilir.

TOPSIS yönteminin uygulanmasında en az iki karar değişkeninin bulunması gerekmektedir. Karar değişkenleri dikkate alınarak alternatiflerin seçilmesinde Öklid, City-Blok ve Manhattan uzaklıkları gibi çeşitli uzaklık yöntemlerinden yararlanılabilir.

Algoritma, n ölçütün ve m alternatifin bulunduğu bir karar matrisi üzerinden uygulanır. Öncelikle karar matrisinde bulunan ölçütlerin ağırlıklandırılması gerekir. Her bir ölçüt farklı birimlere sahip olabileceğinden karar matrisi normalleştirilir. Ağırlıklandırılmış normalize karar matrisi hesaplanır. Fayda ölçütünü en üst düzeye çıkaran ve maliyet ölçütünü en aza indiren pozitif ideal çözüm ile maliyet ölçütünü en üst düzeye çıkaran ve fayda ölçütünü en aza indiren negatif ideal çözüm bulunur. Daha sonra alternatifler dizisi hesaplanan pozitif ve negatif ideal çözüme göre sıralanır.

TOPSIS algoritmasının adımları aşağıdaki gibidir:

Adım 1

$n \times m$ boyutlu karar matrisi oluşturulur

$$D = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1m} \\ d_{21} & d_{22} & \dots & d_{2m} \\ \cdot & \cdot & & \\ \cdot & \cdot & & \\ d_{n1} & d_{n2} & \dots & d_{nm} \end{bmatrix}.$$

Burada tanımlanan D karar matrisi, alternatiflerin ve değerlendirme ölçütleri belirlendikten sonra oluşturulan $n \times m$ boyutlu bir matristir. D matrisinde alternatiflerin sayısı n , değerlendirme ölçütlerinin sayısı ise m 'dir. D matrisi elemanlarından d_{ij} , i . alternatifin j . ölçütüne göre mevcut performansını gösterir.

Adım 2

Karar matrisi normalleştirilir

$$R = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1m} \\ r_{21} & r_{22} & \dots & r_{2m} \\ \cdot & \cdot & & \\ \cdot & & & \\ r_{n1} & r_{n2} & \dots & r_{nm} \end{bmatrix}.$$

Burada, tanımlanan R standart karar matrisi her bir değerlendirme ölçütüne ait değerlerin karelerinin toplamının karekökü alınarak, ilgili elemanın bu çıkan değere bölünmesiyle (4.5)'teki gibi elde edilir.

$$r_{ij} = \frac{d_{ij}}{\sqrt{\sum_{k=1}^n d_{kj}}} \quad , \quad i = 1, 2, \dots, n \quad , \quad j = 1, 2, \dots, m ; \quad (4.5)$$

$0 \quad , \quad \text{diğer durumlarda.}$

Adım 3

Ağırlıklı standart karar matrisi oluşturulur

$$V = \begin{bmatrix} w_1 r_{11} & w_2 r_{12} & \dots & w_m r_{1m} \\ w_1 r_{21} & w_2 r_{22} & \dots & w_m r_{2m} \\ \cdot & \cdot & & \\ \cdot & & & \\ w_1 r_{n1} & w_2 r_{n2} & \dots & w_m r_{nm} \end{bmatrix} = \begin{bmatrix} v_{11} & v_{12} & \dots & v_{1m} \\ v_{21} & v_{22} & \dots & v_{2m} \\ \cdot & \cdot & & \\ \cdot & & & \\ v_{n1} & v_{n2} & \dots & v_{nm} \end{bmatrix}.$$

Değerlendirme ölçütlerine ilişkin ağırlıklar w_i , $i = 1, 2, \dots, m$ olarak belirlenir. w_i değerleri, değerlendirme ölçütlerinin önem derecesine göre araştırmacı tarafından belirlenir. Burada, belirlenen ağırlık değerlerinin toplamının 1 olması gerekir ($\sum_{i=1}^m w_i = 1$).

Adım 4

Ağırlıklı standart karar matrisi V kullanılarak, değerlendirme ölçütünün amacına göre her bir ölçüt için pozitif ve negatif ideal çözüm kümeleri elde edilir. Fayda ölçütü için pozitif ideal çözüm kümesi, V matrisinin sütunlarının en büyük değerlerinden oluşur. Negatif ideal çözüm kümesi ise V matrisinin sütunlarının en küçük değerlerinden oluşur. Maliyet ölçütünde ise bu durumun tam tersidir. Pozitif ve negatif ideal çözüm kümesinin matematiksel gösterimi sırasıyla $V^* = \{v_1^*, v_2^*, \dots, v_m^*\}$ ve $V^- = \{v_1^-, v_2^-, \dots, v_m^-\}$ biçiminde tanımlanır.

Adım 5

Pozitif ve negatif ideal çözümlere olan uzaklık değerleri sırasıyla (4.6) ve (4.7)'deki gibi elde edilir

$$S_i^* = \sqrt{\sum_{j=1}^m (v_{ij} - v_j^*)^2} \quad , \quad i = 1, 2, \dots, n \quad (4.6)$$

$$S_i^- = \sqrt{\sum_{j=1}^m (v_{ij} - v_j^-)^2} \quad , \quad i = 1, 2, \dots, n. \quad (4.7)$$

Burada karar seçeneği sayısı kadar uzaklık değerleri hesaplanır. Her bir karar seçeneğine ilişkin pozitif ve negatif ideal çözüm değerlerinden sapmaları bulabilmek için uzaklık yöntemi olarak genellikle Öklid uzaklık ölçüsü kullanılır.

Adım 6

İdeal çözüme göreli yakınlık katsayıları (4.8)'deki gibi hesaplanır

$$C_i^* = \frac{S_i^-}{S_i^* + S_i^-}, \quad i = 1, 2, \dots, n. \quad (4.8)$$

Pozitif ve negatif ideal çözüme uzaklıkları hesaplanan her bir karar değişkeninin ideal çözüme göreli yakınlık katsayıları, negatif ideal çözüme uzaklığın toplam uzaklık içerisindeki payıdır. Burada, hesaplanan C_i^* değerleri 0 ile 1 arasında değer alır. Buna göre 1'e en yakın değere sahip karar seçenekleri öncelikli olarak tercih edilir.

5. DENEYSEL SONUÇLAR

Değerlendirme ölçülerinin CC algoritması ile elde edilen ikili küme yapılarında uygulanması yapay ve gerçek veri seti üzerinden gösterilmiştir. İkili küme elde etmek ve bu ikili kümenin değerlendirme ölçülerini hesaplamak için R programında f_1 , f_2 , f_3 ve f_4 fonksiyonları oluşturulmuştur. Her bir değerlendirme ölçüsü fonksiyonu içerisinde CC algoritmasının δ ve α parametresi başlangıç kromozomları olarak ele alınmıştır. Böylece her bir değerlendirme ölçüsü CC algoritmasının farklı parametre değerlerine göre hesaplanacaktır.

NSGA-II yönteminde kullanılan parametrelerden sırasıyla girdi boyutu 2 (CC algoritmasının δ ve α parametreleri) ve çıktı boyutu 4 (değerlendirme ölçüsü fonksiyonları) olarak seçilmiştir. Bununla birlikte nesil sayısı literatürde yaygın olarak 100 ile 200 olarak seçilmektedir [15, 25, 28]. Bu çalışmada problemin yapısına göre 200 olması yeterlidir. Nesil sayısının düşük değerde olması çözümlerin global sonuçlara olan uzaklığını artıracak, fazla olması ise hesaplama süresini geciktirecektir. Çözümdeki yeterli çeşitlilik sağlanması ve optimum hesaplanma süresi açısından popülasyon büyüklüğü 20 seçilmiştir.

NSGA-II algoritmasında kullanılan genetik operatörlerden çaprazlama işlemi tek noktalı çaprazlama yöntemi ile gerçekleştirilir. Eğer çaprazlama oranı gerekenden büyük seçilirse popülasyonun hızlı değişmesine sebep olur. Eğer küçük seçilirse yakınsamanın yavaş gerçekleşmesine sebep olur. Bu yüzden çaprazlama oranı genellikle 0.45 ve 0.95 aralığında seçilir. Bir diğer genetik operatörlerden mutasyon oranı ise yüksek belirlendiğinde rastgele aramaya sebep olacağı için %0.1 ve %1 aralığında seçilir [39]. Literatürde yapılan çalışmalarda çaprazlama ve mutasyon operatörleri belirli deneysel çalışmalar sonucunda bu varsayımları sağlayan değerler olarak belirlenmiştir. Örneğin Mitra ve Banka (2006) tarafından yapılan çalışmada çaprazlama ve mutasyon oranları sırasıyla 0.75 ve 0.03 olarak seçilmiştir [38].

Bir başka çalışmada Seridi ve diğerleri (2012) tarafından araştırmanın amacına göre önceden belirlenen çaprazlama ve mutasyon oranları 0.50 ve 0.40 olarak seçilmiştir [15]. Türkşen ve Akgün (2018) tarafından yapılan çalışmada ise genetik algoritmanın parametre değerleri Taguchi deney tasarımı kullanılarak belirlenmiştir [40]. Dale ve diğerleri (2019) tarafından yapılan çalışmada ise genetik operatörler ikili kümeleme algoritmalarından elde edilen ikili kümelerin yapısını etkilemediği için varsayılan değer olarak seçilmiştir [41]. Bu bilgiler doğrultusunda bu çalışmada çaprazlama oranı 0.70 olarak mutasyon oranı ise 0.10 olarak belirlenmiştir.

Girdi parametrelerinden δ için alt sınır ve üst sınır değerleri veri matrisi elemanlarının aldığı değerler aralığında seçilmiştir. Çünkü δ parametresi maksimum kabul edilebilecek MSR değerini gösterir. Buna göre yapay veri matrisi için 1 ve 100 değerleri arasında, gerçek veri matrisi için 1 ve 600 değerleri arasında

seçilmiştir. Bir diğer girdi parametrelerinden α parametresi ölçek faktörü olduğundan ve hesaplama süresini optimal seviyede tutmak için yapay ve gerçek veri matrislerinin her ikisinde de alt sınır ve üst sınır değerleri sırasıyla 1 ve 10 değerleri arasında seçilmiştir. NSGA-II algoritması için kullanılan parametreler *Çizelge 2*'de gösterilmiştir.

Çizelge 2. NSGA-II yönteminde kullanılan parametre değerleri

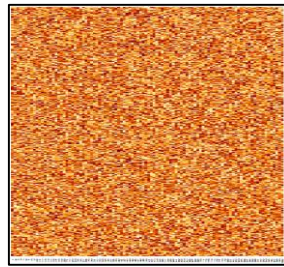
Parametreler	Yapay Veri Seti için Değerler	Gerçek Veri Seti için Değerler
Girdi Boyutu	2 (δ ve α)	2 (δ ve α)
Çıktı Boyutu	4 (VAR, RI, MSA, SS)	4 (VAR, RI, MSA, SS)
Nesil Sayısı	200	200
Popülasyon Büyüklüğü	20	20
Çaprazlama Olasılığı	0.7	0.7
Mutasyon Olasılığı	0.1	0.1
Girdiler için Alt Sınır ve Üst Sınır	$1 < \delta < 10$, $1 < \alpha < 10$	$1 < \delta < 600$, $1 < \alpha < 10$

TOPSIS yönteminde kullanılacak değerlendirme ölçütleri sırasıyla varyans ölçüsü (VAR), uygunluk indeksi (RI), maksimum standart alan (MSA) ve benzerlik skoru (SS)'dur. Bununla birlikte ölçütlerin her biri için ağırlık katsayısı 0.25 olarak alınmıştır. Böylece dört farklı ölçüt için ağırlık katsayıları hem eşit hem de toplamları 1 olacak şekilde ayarlanmıştır. Veri matrisinin büyüklüğüne göre NSGA-II algoritmasının parametreleri değiştirilebilir. Veri matrisinin ve ikili kümenin boyutuna göre 200 nesil tekrarı ve 20 başlangıç popülasyonu en iyi çözüm kümesini elde etmek için yeterli olmuştur.

NSGA-II yöntemi kullanılarak elde edilen alternatif çözümlerin dört farklı ölçüte (değerlendirme ölçüsüne) göre sıralaması TOPSIS yöntemi ile yapılmıştır. Her bir ölçüte eşit ağırlıklandırma yapılmış ve önem dereceleri de eşit olarak seçilmiştir. Böylece çözüm kümesi içerisindeki en iyi δ ve α parametreleri her bir değerlendirme ölçüsünün en iyi değerine göre ve TOPSIS yöntemi ile tüm değerlendirme ölçüleri eşit önem derecesine göre belirlenmiştir.

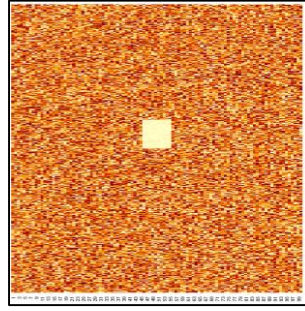
5.1. Yapay Veri

Veri seti 500×100 boyutlu matris şeklinde, 0 ile 100 değerleri arasında tekdüze dağılımdan rastgele üretilmiştir. Veri matrisinin ısı grafiği *Şekil 3*'te gösterilmiştir.



Şekil 3. Veri matrisinin ısı grafiği

Veri matrisi içerisine 50×10 boyutlu referans ikili küme eklenmiştir. Böylece algoritmalarından elde edilen sonuçlarla karşılaştırıldığında referans ikili küme yakın sonuçlar vermesi beklenmektedir. İkili küme elemanları ortalaması 50 ve standart sapması 2 olan normal dağılımdan rastgele üretilmiştir. Buradaki amaç ikili küme elemanlarının hem ortalama etrafında olması hem de sabit değerli ikili küme yapısında olmasını sağlamaktır. Ayrıca sabit değerli ikili küme elemanlarındaki hata payı için normal dağılımın standart sapması 2 olarak seçilmiştir. İçerisinde ikili kümenin bulunduğu veri matrisinde satır ve sütun elemanlarının yerleri rastgele değiştirilmiştir. Böylece ikili küme veri matrisi içerisine gizlenmiştir. İçerisinde referans ikili kümenin bulunduğu veri matrisinin ısı grafiği *Şekil 4*'te gösterilmiştir.



Şekil 4. İçerisinde ikili kümenin bulunduğu veri matrisinin ısı grafiği

Değerlendirme ölçülerine göre en anlamlı ve etkili ikili kümeler elde etmek için CC algoritmasının parametre değerlerinin çözüm kümesi elde edilmiştir. NSGA-II algoritması stokastik aramalara dayalı bir sezgisel optimizasyon algoritması olduğu için algoritma her yinelemede farklı sonuçlar verecektir. Elde edilen çözüm kümesinin belirli bir sonuca yakınsadığını göstermek amacıyla NSGA-II algoritmasının sonuçları 100 yinelemede incelenmiştir.

İçerisinde 50×10 boyutlu ikili kümenin bulunduğu veri matrisi kullanarak NSGA-II yönteminde 100 yinelemeli tekrar ile ikili kümeleme algoritmasının ayarlanabilir parametreleri (δ ve α) elde edilmiştir. NSGA-II algoritmasının popülasyon büyüklüğü 20 seçildiği için çözüm kümesinde δ ve α parametreleri için farklı 20 çözüm bulunmuştur. Her bir çözüm için ortalama, standart sapma, minimum ve maksimum değerleri Çizelge 3'de gösterilmiştir.

Çizelge 3. Yapay veri seti için NSGA-II yöntemi ile elde edilen çözüm kümesi

	Delta Parametresi (δ) $n = 100$				Alfa Parametresi (α) $n = 100$			
	Ort.	Std. Sp.	Min.	Max.	Ort.	Std. Sp.	Min.	Max.
Çözüm 1	1.0824	0.0105	1.0142	1.1610	6.7345	0.0115	6.7155	6.7559
Çözüm 2	3.8468	0.0085	3.8298	3.8705	1.0167	0.0044	1.0008	1.0346
Çözüm 3	1.4938	0.0124	1.4582	1.5238	9.4088	0.0098	9.3898	9.4208
Çözüm 4	2.3041	0.0087	2.2935	2.3197	1.5334	0.0117	1.5293	1.5544
Çözüm 5	1.2912	0.0142	1.2661	1.3124	8.2255	0.0177	8.2089	8.2487
Çözüm 6	1.2298	0.0054	1.2115	1.2374	4.7784	0.0180	4.7583	4.7911
Çözüm 7	3.1557	0.0107	3.1299	3.1719	2.9736	0.0055	2.9657	2.9917
Çözüm 8	2.8416	0.0073	2.8358	2.8516	5.8188	0.0087	5.8047	5.8314
Çözüm 9	2.4688	0.0155	2.4519	2.4912	1.2316	0.0115	1.2211	1.2533
Çözüm 10	3.8366	0.0189	3.8217	3.8575	1.0157	0.0136	1.0007	1.0375
Çözüm 11	3.8398	0.0155	3.8244	3.8611	1.0160	0.0075	1.0089	1.0389
Çözüm 12	3.1295	0.0077	3.1041	3.1519	1.0775	0.0105	1.0599	1.0974
Çözüm 13	1.2844	0.0118	1.2654	1.3024	5.7986	0.0068	5.7911	5.8101
Çözüm 14	3.8412	0.0057	3.8249	3.8656	1.0210	0.0179	1.0173	1.0444
Çözüm 15	2.3518	0.0015	2.3395	2.3679	7.6618	0.0063	7.6422	7.6882
Çözüm 16	3.2525	0.0116	3.2318	3.2797	7.5913	0.0083	7.5741	7.6111
Çözüm 17	2.3452	0.0022	2.3318	2.3519	8.8547	0.0045	8.8488	8.8772
Çözüm 18	1.2953	0.0191	1.2757	1.3193	6.7415	0.0167	6.7311	6.7575
Çözüm 19	2.8205	0.0069	2.8045	2.8404	4.7912	0.0177	4.7854	4.8109
Çözüm 20	2.2318	0.0088	2.2122	2.2543	9.3866	0.0023	9.3618	9.3955

Çözüm kümesindeki parametre değerlerine bakıldığında δ parametresinin ortalama değeri yaklaşık olarak 1.08 ile 3.84 değerleri arasında, α parametresinin ortalama değeri yaklaşık olarak 1.01 ile 9.40 değerleri arasındadır. Her iki parametre değeri için 100 yineleme sonucunda elde edilen çözümlerin standart sapma değeri ortalama 0.01'dir. Standart sapma değerinin küçük olması çözümlerin ortalamaya yakınsadığını göstermektedir. Böylece ikili kümeleme algoritmasının ayarlanabilir parametre değerlerinin yineleme sonucundaki ortalaması alınarak ikili kümeler elde edilir ve ikili kümelerin değerlendirme ölçüleri hesaplanabilir. İkili küme değerlendirme ölçülerine göre tek amaçlı ve çok amaçlı olarak CC algoritmasının ayarlanabilir parametre değerleri Çizelge 4'te gösterilmiştir.

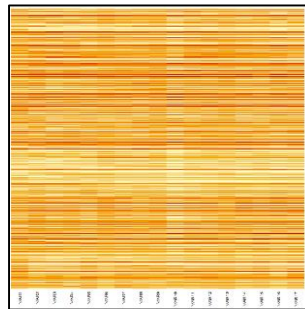
Çizelge 4. Yapay veri seti için farklı değerlendirme ölçülerine göre en iyi parametreler

Amaç Fonksiyonu	min f_1	max f_2	min f_3	max f_4	min $[(f_1), (-f_2), (f_3), (-f_4)]$
VAR	2.1388	2.1814	2.1418	4.1280	2.2966
RI	0.9972	0.9976	0.9973	0.9945	0.9973
MSA	3.4622	3.6204	3.4593	6.5644	3.7675
SS	2.2816	2.4117	2.2782	4.0383	3.1938
δ	1.0824	1.2844	1.2912	3.8412	1.2912
α	6.7345	5.7986	8.2255	1.0210	8.2255
Çözüm	1	13	5	14	5
TOPSIS Sıra	6	4	7	17	1

CC algoritmasının δ ve α parametrelerini belirlemek için oluşturulan deneysel çalışmada sonuçlara bakıldığında, tek amaçlı optimizasyon problemi olarak elde edilen parametre değerleri ve çözüm kümesi içerisindeki sıralama değerleri farklılık göstermektedir. Örneğin VAR ölçüsünü minimum yapan parametreler δ için 1.0824 ve α için 6.7345 iken, RI ölçüsünü maksimum yapan parametreler δ için 1.2844 ve α için 5.7986'dır. Bununla birlikte sadece birinci problem için TOPSIS yöntemi ile uzlaşık tek bir çözüm seçilirse çözüm kümesinde birinci sırada bulunan çözüm, ikinci problem için 13.sırada bulunan çözüm optimaldir. Ancak dört amaç eşit ağırlıklı olacak şekilde dikkate alınırsa δ için 1.2912 ve α için 8.2255 değerleri belirlenir. Çok amaçlı optimizasyon problemine göre çözüm kümesinde 5.sırada bulunan çözüm TOPSIS yöntemi ile belirlenen uzlaşık tek bir çözümdür.

5.2. Gerçek Veri

İkili kümeleme çalışmalarında iyi bilinen gerçek veri setlerinden biri olan "Saccharomyces cerevisiae" ailesinden maya (yeast) verisi kullanılmıştır. Bu veri seti matris şeklinde olup 2884 gen ve 17 koşuldan oluşmaktadır. Veri matrisi elemanları 0 ile 600 arasında değerler almaktadır. Kayıp değerler veri setinden çıkarılmış ve çalışmada 2265×17 boyutlu veri matrisi kullanılmıştır. Veri setine "http://arep.med.harvard.edu/biclustering" adresinden ulaşılabilir. Veri matrisinin ısı grafiği Şekil 5'te gösterilmiştir.



Şekil 5. Maya verisinin ısı grafiği

Maya veri matrisi içerisinde en anlamlı ve etkili ikili kümeler elde etmek için CC algoritmasının parametre değerlerinin çözüm kümesi elde edilmiştir. Yapay veri matrisinde olduğu gibi maya verisinde de NSGA-II algoritması ile 100 yinelemeli tekrar sonucunda ikili kümeleme algoritmasının ayarlanabilir parametreleri (δ ve α) için çözüm kümesi elde edilmiştir. CC algoritmasının ayarlanabilir parametre değerlerinin her biri için ortalama, standart sapma, minimum ve maksimum değerleri hesaplanmış ve Çizelge 5’de gösterilmiştir.

Çizelge 5. Maya veri seti için NSGA-II yöntemi ile elde edilen çözüm kümesi

	Delta Parametresi (δ) $n = 100$				Alfa Parametresi (α) $n = 100$			
	Ort.	Std. Sp.	Min.	Max.	Ort.	Std. Sp.	Min.	Max.
Çözüm 1	8.5169	0.0078	8.5097	8.5214	1.1182	0.0067	1.0972	1.1258
Çözüm 2	1.1347	0.0103	1.1197	1.1498	1.7720	0.0082	1.7654	1.7915
Çözüm 3	2.0916	0.0012	2.0877	2.0983	1.7792	0.0092	1.7629	1.7922
Çözüm 4	2.9914	0.0069	2.9866	3.0144	1.1425	0.0751	1.1255	1.1599
Çözüm 5	2.8942	0.0088	2.8843	2.9116	1.0667	0.0105	1.0581	1.0706
Çözüm 6	3.8723	0.0144	3.8679	3.8868	1.0961	0.0078	1.0875	1.1131
Çözüm 7	1.0597	0.0154	1.0458	1.0716	1.0993	0.0658	1.0897	1.1103
Çözüm 8	8.5046	0.0097	8.4948	8.5113	1.3451	0.0092	1.3369	1.3588
Çözüm 9	10.0977	0.0099	10.0864	10.1048	1.1127	0.0126	1.0987	1.1213
Çözüm 10	10.8345	0.0035	10.8247	10.8431	1.0993	0.0625	1.0866	1.1089
Çözüm 11	43.9225	0.0106	43.9141	43.9377	1.1193	0.0038	1.1092	1.1263
Çözüm 12	18.0156	0.0083	18.0087	18.0248	1.1062	0.0097	1.0973	1.1122
Çözüm 13	13.6945	0.0044	13.6847	13.7044	1.0978	0.0105	1.0836	1.1010
Çözüm 14	35.9566	0.0079	35.9497	35.9711	1.1042	0.0094	1.0949	1.1176
Çözüm 15	32.0863	0.0169	32.0770	32.0955	1.0594	0.0074	1.0491	1.0719
Çözüm 16	32.0788	0.0029	32.0657	32.0874	1.0628	0.0316	1.0545	1.0770
Çözüm 17	13.7264	0.0021	13.7145	13.7374	1.1148	0.0110	1.0981	1.1237
Çözüm 18	9.2146	0.0196	9.2064	9.2248	1.0598	0.0068	1.0455	1.0711
Çözüm 19	18.0158	0.0067	18.0086	18.0251	1.1022	0.0239	1.0903	1.1126
Çözüm 20	4.3059	0.0499	4.2949	4.3136	1.1293	0.0072	1.1187	1.1491

NSGA-II yöntemi ile elde edilen çözüm kümesinde CC algoritmasının ayarlanabilir parametreleri δ için yaklaşık 1.06 ile 43.92 değerleri arasında, α için yaklaşık 1.05 ile 1.78 değerleri arasında bulunmuştur. δ ve α parametreleri için 100 yineleme sonucunda elde edilen çözümlerin standart sapma değeri ortalaması yaklaşık 0.01’dir. Böylece sapmanın çok küçük değerde olmasıyla yineleme ile elde edilen çözümler için ortalama değer kullanılabilir. CC algoritmasının parametre değerlerinin ortalaması kullanılarak ikili kümeler elde edilir. Elde edilen ikili kümelerin değerlendirme ölçüleri Çizelge 6’deki gibi hesaplanmıştır.

Çizelge 6. Maya verisi için farklı değerlendirme ölçülerine göre en iyi parametreler

Amaç Fonksiyonu	$\min f_1$	$\max f_2$	$\min f_3$	$\max f_4$	$\min [(f_1), (-f_2), (f_3), (-f_4)]$
VAR	1149.1577	1153.7813	1166.1734	3284.9496	3166.1216
RI	0.8249	0.8395	0.8211	0.7594	0.7861
MSA	54.4699	55.8978	52.1779	96.4496	93.1271
SS	0.0349	0.0288	0.0468	2.1442	3.6544
δ	1.1347	1.0597	2.0916	8.5046	9.2146
α	1.7720	1.0993	1.7792	1.3451	1.0598
Çözüm	2	7	3	8	18
TOPSIS Sıra	7	9	8	3	1

VAR, RI, MSA ve SS ölçütlerine göre δ ve α parametreleri incelendiğinde, tek amaçlı optimizasyon problemlerinin her biri için ayrı ayrı değerler elde edilmiştir. Sadece VAR ölçütü minimize edildiğinde, yani f_1 fonksiyonu optimalliğe ulaştığında δ için 1.1347 ve α için 1.7720 değeri elde edilir. Sadece RI ölçütü maksimize edildiğinde, yani f_2 fonksiyonu optimalliğe ulaştığında δ için 1.0597 ve α için 1.0993 değeri elde edilir. Sadece MSA ölçütü minimize edildiğinde, yani f_3 fonksiyonu optimalliğe ulaştığında δ için 2.0916 ve α için 1.7792 değeri elde edilir. Sadece SS ölçütü maksimize edildiğinde, yani f_4 fonksiyonu optimalliğe ulaştığında δ için 8.5046 ve α için 1.3451 değeri elde edilir. Çok amaçlı optimizasyon problemi olarak değerlendirildiğinde ise VAR, RI, MSA ve SS ölçütleri eşit önem derecesine sahip olduğunda δ için 9.2146 ve α için 1.0598 değeri elde edilir. TOPSIS yöntemine göre çözüm kümesi içerisinde 18.sırada bulunan çözüm, belirlenen uzlaşık tek bir çözümdür.

6. SONUÇ VE ÖNERİLER

Gen açıklama verilerinde benzer gen yapılarının belirli koşullara göre kümelenmesi son yıllarda önem kazanmıştır. Bu yapıları belirlemek için geleneksel kümeleme yöntemleri yerine ikili kümeleme yöntemi kullanılır.

Bu çalışmada literatürde yaygın olarak kullanılan CC algoritması dikkate alınmıştır. İkili kümeler elde etmek için CC algoritmasının δ ve α parametreleri belirlenmesi gerekir. Parametre değerleri değiştikçe veri matrisinden farklı ikili kümeler elde edilmektedir. Elde edilen farklı ikili kümelerden anlamlı ve etkili olanları belirlemek için kullanılan birçok değerlendirme ölçüsü bulunmaktadır. Birden fazla değerlendirme ölçüsüne dikkate alınarak anlamlı ve etkili ikili kümeler belirlemek çok amaçlı optimizasyon problemini ortaya çıkarır. Böylece δ ve α parametrelerinin belirlenmesi için çok amaçlı sezgisel yöntemlerden biri olan NSGA-II algoritması kullanılmıştır.

CC algoritmasının NSGA-II yöntemiyle elde edilen parametre değerlerine göre bulunan ikili kümelerin değerlendirilmesi varyans ölçüsü (VAR), uygunluk indeksi (RI), maksimum standart alan (MSA) ve benzerlik skoru (SS) ölçüleri ile yapılmıştır. Bu değerlendirme ölçüleri çözüm kümesi içerisinde ölçütler olarak belirlenmiş ve TOPSIS yöntemi uygulanarak çözümlerin ölçütlere göre sıralaması yapılmıştır. NSGA-II yöntemi kullanılarak bulunan çözüm kümesi içerisinde TOPSIS yöntemiyle sıralama yapılmış ve yapay veri matrisinde referans ikili kümeye benzer en iyi ikili küme bulunmuştur. Bununla birlikte literatürde yaygın bir şekilde kullanılan gen verilerinden maya verisi için en iyi ikili kümelerin belirlenmesinde parametre değerleri elde edilmiştir. Böylece en etkili ve anlamlı ikili kümelerin elde edilmesinde CC algoritmasının hangi parametre değerlerinin alacağı belirlenmiştir.

İkili küme algoritmasını uygulamak için 0 ile 100 değerleri arasında tekdüze dağılımdan rastgele üretilen 500×100 boyutlu yapay veri matrisi ve 2265×17 boyutlu maya verisi kullanılmıştır. Yapay veri matrisi içerisinde 50×10 boyutlu referans ikili küme bulunmaktadır. CC algoritmasının δ ve α parametreleri için başlangıç çözüm kümesi yapay veri matrisi için 1 ile 10 değerleri arasında rastgele üretilen 20 sayıdan oluşmaktadır. Maya verisi için parametresi 1 ile 600 arasında, parametresi ise 1 ile 10 değerleri arasında rastgele üretilen 20 sayıdan oluşmaktadır. Her bir parametre değeri için VAR, RI, MSA ve SS ölçüleri amaç fonksiyonu olarak belirlenmiştir. Amaç fonksiyonlarının minimizasyon problemi olması için RI ve SS ölçüsü fonksiyonlarına dönüşüm uygulanmıştır. NSGA-II yönteminde çözüm kümesinin 100 yinelenmeli tekrarlar sonuçları incelenmiştir. Bununla birlikte çözüm kümesi sonuçlarının ortalaması, standart sapması, minimum ve maksimum değerleri elde edilmiştir. Bu sonuçlara göre optimal amaç fonksiyonu değerlerinden elde edilen CC algoritmasının ayarlanabilir parametre değerleri bulunmuştur. Bulunan parametre değerlerine göre elde edilen ikili kümelerin değerlendirilmesi tek amaçlı ve çok amaçlı optimizasyon yöntemi ile yapılmıştır. Tek amaçlı optimizasyon problemindeki dezavantaj, hangi ölçüte göre optimal yapılmasına karar vermektir. Çünkü her bir değerlendirme ölçütüne göre farklı sonuçlar ortaya çıkmaktadır. Bu dezavantajı ortadan kaldırmak için çok amaçlı optimizasyon problemi olarak dört farklı ölçüt eşit ağırlıklandırılarak optimal çözüm elde edilmiştir.

Bu çalışmada sadece CC algoritması ve ikili küme türlerinden sabit ikili küme yapısı ele alınmıştır. İkili kümenin değerleri sabitlerden oluşabileceği gibi satırlar veya sütunlar bazında toplamsal veya çarpımsal artışların bulunduğu değerler de olabilir. Ayrıca hem toplamsal hem de çarpımsal değerlerden oluşan ikili küme yapıları da mümkündür. Bundan sonraki çalışmalarda diğer algoritmalar ve farklı yapıdaki ikili küme türlerinin bulunduğu deneysel çalışma, ikili kümeler için kullanılacak farklı değerlendirme ölçüleri veya farklı gerçek veri matrisleri kullanılarak ikili kümeleme algoritmalarının parametre değerleri çok amaçlı sezgisel yöntemler kullanılarak belirlenebilir.

ÇIKAR ÇATIŞMASI/ÇAKIŞMASI BİLDİRİMİ

Yazarlar arasında çıkar çatışması/çakışması bulunmamaktadır.

KAYNAKLAR

- [1] Mirkin, B. (1998). Mathematical classification and clustering: From how to what and why. *Classification, Data analysis, and Data Highways*, 172-181.
- [2] Wang, B., Miao, Y., Zhao, H., Jin, J., & Chen, Y. (2016). A biclustering-based method for market segmentation using customer pain points. *Engineering Applications of Artificial Intelligence*, 47, 101-109.
- [3] Dhillon, I. S. (2001). Co-clustering documents and words using bipartite spectral graph partitioning. *In Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, 269-274.
- [4] Busygin, S., Prokopyev, O., & Pardalos, P. M. (2008). Biclustering in data mining. *Computers & Operations Research*, 35(9), 2964-2987.
- [5] Cheng, Y., & Church, G. M. (2000). Biclustering of expression data. "International Conference on Intelligent Systems for Molecular Biology" kongresinde sunulan bildiri, UC San Diego, California, USA.
- [6] Lazzeroni, L., & Owen, A. (2002). Plaid models for gene expression data. *Statistica Sinica*, 61-86.
- [7] Bergmann, S., Ihmels, J., & Barkai, N. (2003). Iterative signature algorithm for the analysis of large-scale gene expression data. *Physical Review E*, 67(3), 031902.
- [8] Ben-Dor, A., Chor, B., Karp, R., & Yakhini, Z. (2003). Discovering local structure in gene expression data: the order-preserving submatrix problem. *Journal of Computational Biology*, 10(3-4), 373-384.
- [9] Prelic, A., Bleuler, S., Zimmermann, P., Wille, A., Bühlmann, P., Gruissem, W., Zitzler, E. (2006). A systematic comparison and evaluation of biclustering methods for gene expression data. *Bioinformatics*, 22(9), 1122-1129.
- [10] Li, G., Ma, Q., Tang, H., Paterson, A. H., & Xu, Y. (2009). QUBIC: a qualitative biclustering algorithm for analyses of gene expression data. *Nucleic Acids Research*, 37(15), 101.
- [11] Hochreiter, S., Bodenhofer, U., Heusel, M., Mayr, A., Mitterecker, A., Kasim, A., Talloen, W. (2010). FABIA: factor analysis for bicluster acquisition. *Bioinformatics*, 26(12), 1520-1527.
- [12] Chekouo, T., & Murua, A. (2015). The penalized biclustering model and related algorithms. *Journal of Applied Statistics*, 42(6), 1255-1277.
- [13] Biswal, B. S., Mohapatra, A., & Vipsita, S. (2019). Ensemble Neighborhood Search (ENS) for biclustering of gene expression microarray data and single cell RNA sequencing data. *Journal of King Saud University-Computer and Information Sciences*, 5(2), 105-112.
- [14] Chowdhury, H. A., Ahmed, H. A., Bhattacharyya, D. K., & Kalita, J. K. (2020). NCBI: A Novel Correlation Based Imputing Technique Using Biclustering. *Computational Intelligence in Pattern Recognition*, 1, 509-519.
- [15] Seridi, K., Jourdan, L., & Talbi, E.-G. (2012). Hybrid metaheuristic for multi-objective biclustering in microarray data. "2012 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)" kongresinde sunulan bildiri, San Diego, California, USA.
- [16] Al-Akwaa, F. M., Ali, M. H., & Kadah, Y. M. (2009). Bicat_plus: An automatic comparative tool for bi/clustering of gene expression data obtained using microarrays. "2009 National Radio Science Conference" kongresinde sunulan bildiri.
- [17] Chia, B. K. H., & Karuturi, R. K. M. (2010). Differential co-expression framework to quantify goodness of biclusters and compare biclustering algorithms. *Algorithms for Molecular Biology*, 5(1), 23.

- [18] Karim, M. B., Kanaya, S., & Altaf-UI-Amin, M. (2019). Implementation of BiClusO and its comparison with other biclustering algorithms. *Applied Network Science*, 4(1), 1-15.
- [19] Liu, X., & Wang, L. (2006). Computing the maximum similarity bi-clusters of gene expression data. *Bioinformatics*, 23(1), 50-56.
- [20] Padilha, V. A., & Campello, R. J. (2017). A systematic comparative evaluation of biclustering techniques. *BioMed Central Bioinformatics*, 18(1), 55.
- [21] Seridi, K., Jourdan, L., & Talbi, E.-G. (2015). Using multiobjective optimization for biclustering microarray data. *Applied Soft Computing*, 33, 239-249.
- [22] Goldberg, D. E., & Holland, J. H. (1988). Genetic algorithms and machine learning. *Machine Learning*, 3(2), 95-99.
- [23] Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671-680.
- [24] Alikar, N., Mousavi, S. M., Ghazilla, R. A. R., Tavana, M., & Olugu, E. U. (2017). Application of the NSGA-II algorithm to a multi-period inventory-redundancy allocation problem in a series-parallel system. *Reliability Engineering & System Safety*, 160, 1-10.
- [25] Vo-Duy, T., Duong-Gia, D., Ho-Huu, V., Vu-Do, H., & Nguyen-Thoi, T. (2017). Multi-objective optimization of laminated composite beam structures using NSGA-II algorithm. *Composite Structures*, 168, 498-509.
- [26] Wang, B., Liang, Y., Zheng, T., Yuan, M., & Zhang, H. (2018). Multi-objective site selection optimization of the gas-gathering station using NSGA-II. *Process Safety and Environmental Protection*, 119, 350-359.
- [27] Wang, S., Ma, S., & Duan, W. (2018). Seakeeping optimization of trimaran outrigger layout based on NSGA-II. *Applied Ocean Research*, 78, 110-122.
- [28] Yang, Y., Cao, L., Wang, C., Zhou, Q., & Jiang, P. (2018). Multi-objective process parameters optimization of hot-wire laser welding using ensemble of metamodels and NSGA-II. *Robotics and Computer-Integrated Manufacturing*, 53, 141-152.
- [29] Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multiobjective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182-197.
- [30] Türkşen, Ö. (2011). *Çok Yanıtlı Yüzey Problemlerinin Çözümüne Bulanık ve Sezgisel Yaklaşım*. Doktora Tezi. Ankara Üniversitesi Fen bilimleri Enstitüsü. Ankara. 126.
- [31] Hwang, C., & Yoon, K. (1981). *Multiple decision attribute making: Methods and applications*. New York: Springer-Verlag, 58-191.
- [32] Pamucar D. & Cirovic G. (2015). The selection of transport and handling resources in logistics centers using multi-attributive border approximation area comparison (mabac). *Expert Systems with Applications*, 42(6), 3016-3028.
- [33] Liu P., Li H., Wang P., & Liu J. (2016). Electre method and its application in multiple attribute decision making based on ins. *Journal of Shandong University of Finance and Economics*, 28(2), 80-87.
- [34] Pontes, B., Girdlez, R., & Aguilar-Ruiz, J. S. (2015). Quality measures for gene expression biclusters. *Plos One*, 10(3), 0115497.
- [35] Yip, K. Y., Cheung, D. W., & Ng, M. K. (2004). Harp: A practical projected clustering algorithm. *IEEE Transactions on Knowledge and Data Engineering*, 16(11), 1387-1397.
- [36] Lashkargir, M., Monadjemi, S. A., & Dastjerdi, A. B. (2009). A new biclustering method for gene expression data based on adaptive multi-objective particle swarm optimization. *In 2009 Second International Conference on Computer and Electrical Engineering*, 1, 559-563.
- [37] Liu, J., Li, Z., Hu, X., & Chen, Y. (2009, April). Biclustering of microarray data with MOSPO based on crowding distance. *In BMC bioinformatics*, 10(4), 1-10.
- [38] Mitra, S., & Banka, H. (2006). Multi-objective evolutionary biclustering of gene expression data. *Pattern Recognition*, 39(12), 2464-2477.
- [39] Talbi, E.-G. (2009). *Metaheuristics: from design to implementation* (74). New Jersey: John Wiley & Sons, 34-48.
- [40] Türkşen, Ö., & Akgün, F. (2018). Genetik-Simpleks hibrit algoritması ile doğrusal olmayan regresyon model parametrelerinin nokta tahmini. *İstatistikçiler Dergisi: İstatistik ve Aktüerya*, 11(2), 81-92.
- [41] Dale, J., Zhao, J., & Obafemi-Ajayi, T. (2019). Multi-objective optimization approach to find biclusters in gene expression data. *In 2019 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*, 1-8.