



Karma veri içeren çok yanıtlı problemlerde NSGA-II'ye dayalı bir optimizasyon yaklaşımı

Gözde Karakoç¹

Ankara Üniversitesi, Fen Bilimleri Enstitüsü,
İstatistik Anabilim Dalı
Ankara, Türkiye
gozcirpan@ankara.edu.tr
ORCID: 0000-0001-9334-765X

Özlem Türkşen

Ankara Üniversitesi, Fen Fakültesi,
İstatistik Bölümü
Ankara, Türkiye
turksen@ankara.edu.tr
ORCID: 0000-0002-5592-1830

Öz

Bir optimizasyon probleminin matematiksel modeli, kesikli ve sürekli değer alan girdi ve/veya yanıt değişkenlerini içermesi durumunda problem, karma veri içeren optimizasyon problemi olarak adlandırılır. Bu çalışmada, girdi değişkenleri bakımından karma veri içeren çok yanıtlı problemlerin modelleme ve optimizasyon aşamaları ele alınmıştır. Modelleme aşamasında Genelleştirilmiş Lineer Modeller (GLM) kullanılarak tahmini yanıt fonksiyonları elde edilmiştir. Optimizasyon aşamasında ise elde edilen tahmini yanıt fonksiyonları bir amaç fonksiyonu olarak dikkate alınıp problem, eş anlı optimizasyonu istenilen çok amaçlı optimizasyon (ÇAO) problemi biçiminde ifade edilmiştir. Çalışmada, ÇAO'da sıklıkla kullanılan yapay zeka optimizasyon algoritmalarından biri olan NSGA-II (Non-dominated Sorting Genetic Algorithm-II)'ye dayalı yeni bir çözüm algoritması önerilmiştir. NSGA-II'de, değişken gösterimi, başlangıç popülasyonu oluşturma ve genetik operatörlerin uygulanması aşamalarında çeşitli uyarlamalar yapılarak hazırlanan bu algoritma, çalışma kapsamında MDNSGA-II (Mixed Data NSGA-II) olarak adlandırılmıştır. MDNSGA-II'de, her bir kesikli değişken değerine bir pozitif tam sayı değeri atanarak, kesikli değişken değerleri için bir tam sayı indekslemesi yapılmıştır. Yapılan indeksleme işlemiyle kesikli değişkenin tanım kümesinden değerler alması sağlanmıştır. Çalışmanın uygulama kısmında, UCI Repository veri tabanından enerji verimliliği konulu karma veri seti ve gıda alanında literatürde mevcut olan deneysel karma veri seti kullanılarak önerilen MDNSGA-II ile Pareto çözümlerin elde edilebilir olduğu gösterilmiştir.

Anahtar sözcükler: Çok amaçlı optimizasyon, Çok yanıtlı problem, Karma veri, MDNSGA-II, NSGA-II

¹ Bu çalışma, birinci yazarın, ikinci yazarın danışmanlığında hazırladığı doktora tezinden üretilmiştir.

Abstract

An optimization approach based on NSGA-II for multi-response problems with mixed data

If the mathematical model of an optimization problem contains input and/or response variables that take discrete and continuous values, the problem is called as a mixed data optimization problem. In this study, the modeling and the optimization phases of multi-response problems with mixed data in terms of input variables are discussed. In the modeling phase, estimated response functions are obtained using Generalized Linear Models (GLM). In the optimization phase, the estimated response functions are considered as an objective function and the problem is expressed as a multi-objective optimization (MOO) problem with simultaneous optimization. In this study, a new solution algorithm based on the Non-dominated Sorting Genetic Algorithm-II (NSGA-II), one of the most frequently used artificial intelligence optimization algorithms in MOO, is proposed. This algorithm, which was prepared by making various adaptations in the variable representation, initial population generation and application of genetic operators stages of the NSGA-II, is called MDNSGA-II (Mixed Data NSGA-II) in this study. In MDNSGA-II, each discrete variable value is assigned a positive integer value and an integer indexing is performed for the discrete variable values. It is ensured that the discrete variable takes values in support set of variables by using the indexing approach. In the application part of the study, it is shown that Pareto solutions can be obtained with the proposed MDNSGA-II by using the mixed data set on energy efficiency from the UCI Repository database and the experimental mixed data set available in the literature in the field of food.

Keywords: Multi-objective optimization, Multi-response problem, Mixed data, MDNSGA-II, NSGA-II

1. Giriş

Fen, mühendislik, sağlık, sosyal, finans vb. alanlarda yapılan araştırmalarda kullanılacak verilerin toplanması veri-bilgi keşfi sürecinde büyük önem taşımaktadır. Günümüzde gelişen teknolojiye bağlı olarak veri toplama yöntemleri hızlı ve pratik bir hal alsada anketler, görüşmeler, gözlemler ve deneysel çalışmalar veri toplama yöntemleri başlığında yer alan dört temel yöntem olarak değerlendirilebilir. Toplanan verilerin analizinden önce veri yapısının araştırmacı tarafından iyi anlaşılması gerekir. Gözlem ve deney yolu ile elde edilmiş bir veri seti birden fazla yanıt (bağımlı, açıklanan) ve girdi (bağımsız, açıklayıcı) değişkeni içerebilir.

Değişkenler sürekli, kesikli ve/veya tam sayı değerli olabilir. Veri setindeki değişkenlerin (yanıt ve/veya girdi değişkenlerinin) bazılarının sürekli bazılarının kesikli değerli olması durumunda veri, karma veri olarak tanımlanır. Karma veri ile ilgili modelleme ve optimizasyon çalışmaları disiplinler arası çalışmalarda oldukça önemli yer tutar. Optimizasyon sonuçlarının güvenilirliği, veriye ilişkin oluşturulan istatistiksel modelin güvenilirliği ile doğru orantılıdır.

Karma veri setindeki yanıt değişkeninin kesikli ya da sürekli değerli olmasına göre oluşturulacak model farklılık gösterir. Sürekli değerli yanıt değişkeni ile girdi değişkenleri arasındaki fonksiyonel ilişkinin tanımlanmasında çoklu doğrusal regresyon analizi uygulanırken, yanıt değişkeninin kesikli değerli olması durumunda dağılım yapısına uygun olarak lojistik, binom veya poisson regresyon analizleri uygulanabilir. Uygulanan doğrusal regresyon modelleri, Genelleştirilmiş Lineer Modeller (Generalized Linear Models - GLM)'in özel halleridir. GLM, kitle ortalamasının bir bağ (link) fonksiyonu ile doğrusal tahmin ediciye bağlı olmasına izin veren lineer modellerin genelleştirilmiş halidir [1]. GLM, genellikle biyolojik analizlerde, çeşitli uygulamalı biyomedikal alanlarda, güvenilirlik ve hayatta kalma analizlerinde elde edilen veri setlerinden istatistiksel sonuçlar çıkarmak için kullanılır.

Literatürde, GLM kullanılarak farklı alanlarda karma veri setleri için yapılan çalışmalar mevcuttur. GLM, HIV virüsü riski ile partnerlerle temas sayısı arasındaki ilişkinin incelenmesinde [2], entomolojide (böcek bilimi) böcek davranışındaki değişiklikleri bir bitki özütünün kimyasal bileşimindeki değişikliklerle

ilişkilendirmek amacıyla [3], klinik çalışmalarda hastanelerde hastalara uygulanan tedavilerin etkilerine ilişkin tahminlerin elde edilmesinde [4], önemli bir ağaç türünün mekansal desenini incelemede [5], klimatolojide (iklim bilimi) belirli bölgelerdeki temel klimatolojik modeli ve günlük maksimum rüzgar hızındaki eğilimlerin belirlenmesinde [6] kullanılmıştır.

GLM ile tahmini yanıt fonksiyonları oluşturulduktan sonra her bir yanıt fonksiyonu bir amaç fonksiyonu olarak ele alınıp çok amaçlı optimizasyon (ÇAO) aşamasına geçilir. ÇAO'da çözümler üstel hesaplama gerektirebilir. Klasik optimizasyon yöntemleri, bu tür zor optimizasyon problemlerini çözmeye yetersiz kalmaktadır. Bu nedenle, zor optimizasyon problemlerinin çözümünde makul zamanda optimal sonuca yakın çözümler üreten yapay zeka optimizasyon algoritmaları önemli yer tutar. Karma veri içeren çok yanıtlı optimizasyon problemlerini çözmek için uygun yöntemin seçimi kesikli değer alan değişkenlerin tipine ve amaç fonksiyonunun yapısına bağlıdır. Literatürde, karma veri içeren ÇAO problemlerini çözmek için farklı alanlarda yapılan çalışmalar mevcuttur. Rajeev ve Krishnamoorthy [7], Lin ve Hajela [8] karma veri içeren optimizasyon problemlerinin çözümü için Genetik Algoritma (GA)'nın uygun olduğunu gösteren çalışmalar yapmış ve çalışmalarında ikili (0-1) kodlama kullanmışlardır. Wang ve ark. [9], bir bina tasarımında, tasarımcılara yardımcı olabilecek ÇAO problemini hem ekonomik hem de çevresel kriterler bakımından değerlendirerek Pareto optimal çözümün belirlenmesi için çok amaçlı GA kullanmışlardır. Rao ve Xiong [10], karma kesikli bulanık çok amaçlı programlama problemlerini çözmek için bulanık λ formülasyonu ve oyun teorisi tekniklerinin karma kesikli hibrit GA ile birleştirildiği yeni bir yöntem sunmuşlardır. Sundukları yöntemin kesin olmayan bir ortamda daha gerçekçi ve tatmin edici sonuçlar elde etmek için çeşitli mühendislik tasarım problemlerine esnek ve etkili bir biçimde uygulanabildiğini göstermişlerdir. Ahmadi ve ark. [11], havza ölçeğinde tarımsal koruma uygulamalarının hedeflenen şekilde uygulanması için karma karar değişkenlerine sahip NSGA-II kullanmışlardır. El-Kribi ve ark. [12], sürekli ve kesikli değişkenlere sahip dört çubuklu bir sistemin mekatronik tasarımını eş anlı optimize etmek için NSGA-II'yi kullanarak farklı tasarım koşulları için analiz etmişlerdir. Tong ve ark. [13], karma kesikli problemler için daha önce geliştirilen ve erken parçacık kümelenmesini önlemek için özel bir çeşitlilik koruma tekniği içeren karma-kesikli Parçacık Sürü Optimizasyonu algoritmasını çok amaçlı problemleri çözmek için kullanmış ve NSGA-II sonuçları ile karşılaştırmışlardır. Holzman ve Smith [14], kesikli ÇAO problemine tam verimli bir çözüm kümesi üretmek için modifiye edilmiş artırılmış ağırlıklı Tchebychev normunu sunmuşlardır. Kullandıkları algoritmanın çalışma sürelerini literatürde önerilen algoritmaların çalışma süreleriyle karşılaştırmışlardır. Guangyong ve ark. [15], belirsizlikler içeren mühendislik yapılarının tasarımı için yeni çok amaçlı kesikli robust optimizasyon algoritması önermişlerdir. Çalışmalarında çok kriterli karar verme tekniğini Taguchi yöntemine dahil ederek çok amaçlı kesikli sağlam tasarımı ele almayı amaçlamışlardır. Önerdikleri algoritmanın NSGA-II ile elde edilen Pareto sınırlarına yakın olduğunu belirtmişlerdir. Roy ve ark. [16], iki dallı bir genetik algoritmayı küresel bir arama aracı olarak yerel arama için gradyan tabanlı bir yaklaşımla birleştiren, kısıtlı çok amaçlı karma kesikli doğrusal olmayan programlama problemine çözüm bulmak için yeni bir hibrit yaklaşım önermişlerdir.

Çoğu zaman karma veri içeren çok yanıtlı problemlerin optimizasyonu için uygun bir yöntemin seçimi, optimizasyon problemindeki amaç ve kısıt fonksiyonlarının yapısı ile kesikli değişkenlerin aldığı değerlere bağlıdır. Literatürde mevcut olan ÇAO yaklaşımlarında kesikli değişken gösterimi aşamasında ikili kodlamanın kullanıldığı görülmüştür. Bu kodlama tipi kolaylıkla uygulanabilir olsa da kesikli değişken sayısının ve kesikli değişkenlerin aldığı değerlerin fazla olması durumunda hesaplama süresi arttığından farklı kodlama yaklaşımları geliştirilmiştir. Yapılan bu çalışmayla, kesikli değişken gösterimi için ikili kodlama yerine değer kodlama yaklaşımı kullanılarak hesaplama süresinin kısaltılması hedeflenmiştir. NSGA-II'nin değişken gösterimi adımı yapılan bu değer kodlaması uyarlaması ile kesikli değişkenlerin aldığı değerlerin optimizasyon sürecine dahil edilmesi sağlanmıştır.

Bu çalışmada, kesikli ve sürekli değer alan değişkenlere sahip karma verilerin modellenmesi ve bu modellerin tanımlanan amaca yönelik ÇAO kapsamında Pareto çözüm sonuçlarının belirlenmesi amaçlanmıştır. Çok yanıtlı problemlerin optimizasyonunda etkin bir yapay zeka optimizasyon algoritması olan NSGA-II yönteminde, değişken gösterimi, başlangıç popülasyonu oluşturulması ve genetik operatörlerin belirlenmesi aşamalarında uyarlamalar yapılarak karma veri içeren yanıt fonksiyonlarının eş

anlı optimizasyonu sağlanmıştır. Uyarlanan NSGA-II, MDNSGA-II olarak adlandırılmıştır. Çözümlerin hesaplama süresini kısaltması nedeniyle çalışmada değer kodlama kullanılmıştır. Değer kodlaması ile ele alınan ÇAO problemde kesikli değişken değerleri için pozitif tam sayı indekslemesi yapılmıştır. Yapılan indeksleme sonucunda kesikli değişkenlerin indeks değerleri dikkate alınarak karma veri içeren çok yanıtlı problemler için Pareto çözüm kümesi elde edilmiştir. Çalışmanın ikinci bölümünde, çok yanıtlı karma verilerin GLM ile modellenmesi hakkında kısa bilgi verilerek, çalışmada kullanılan GLM modelleri sunulmuştur. Üçüncü bölümde NSGA-II yönteminin bazı adımlarında uyarlamalar yapılarak oluşturulan MDNSGA-II detaylı olarak açıklanmıştır. Dördüncü bölümde, UCI Repository veri tabanından enerji verimliliği konulu karma veri seti ve literatürde mevcut olan gıda alanından deneysel karma veri seti uygulamalarına yer verilmiştir. Çalışmanın sonuç bölümünde, önerilen MDNSGA-II ile elde edilen Pareto çözüm kümesi değerlendirilerek, modelleme ve optimizasyon aşamalarında yapılması planlanan sonraki çalışmadan söz edilmiştir.

2. Çok yanıtlı karma verilerin modellenmesi

Birçok disiplinde, yapılan çalışmalardan elde edilen verilerin modellenmesi oldukça önem taşımaktadır. Güvenilir bir model kurma, veriyi iyi anlayıp doğru analizlerin yapılması ile mümkündür. Bu nedenle modelleme aşamasına geçmeden önce veri yapısı incelenmelidir. Veriler, girdi değişkenleri ve yanıt değişken değerlerinden oluşur. Yanıt değişkenlerinin birden fazla olması durumunda çok yanıtlı verilerin analizinde çok değişkenli yapının göz önünde bulundurulması gerekir. Yapılan modelleme çalışmalarında girdi değişkenlerinin ($X_i, i = 1, 2, \dots, p$) ilgilenilen yanıt değişkenleri ($Y_j, j = 1, 2, \dots, r$) üzerindeki etkisi incelenir. Yanıt değişkeni ile girdi değişkenleri arasındaki fonksiyonel ilişki, GLM ile elde edilir. GLM, kesikli ve sürekli yanıt değişkenleri için birleştirilmiş regresyon modelleri sınıfıdır [17]. Çizelge 1’de, p tane girdi değişkeni, r tane yanıt değişkenine sahip n gözlemlili çok yanıtlı karma veri seti yer almaktadır. Burada, $\mathbf{X}, n \times p$ boyutlu girdi değişkenlerin tasarım matrisi ve $\mathbf{Y} = [Y_1 \ Y_2 \dots Y_r]^T$, her bir bileşeni $Y_i, i = 1, 2, \dots, r, n \times 1$ boyutlu olan yanıt değişkenleri vektörüdür.

Çizelge 1. Çok yanıtlı karma veri seti

No	Girdi Değişkenleri				Yanıt Değişkenleri			
	X_1	X_2	...	X_p	Y_1	Y_2	...	Y_r
1	x_{11}	x_{12}	...	x_{1p}	y_{11}	y_{12}	...	y_{1r}
2	x_{21}	x_{22}	...	x_{2p}	y_{21}	y_{22}	...	y_{2r}
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	x_{n1}	x_{n2}	...	x_{np}	y_{n1}	y_{n2}	...	y_{nr}

GLM, yanıt değişkeni dağılımının Normal dağılım varsayımını sağlamadığı durumları da göz önüne alan hem doğrusal hem de doğrusal olmayan regresyon modellerinin bir birleşimidir [18]. Başka bir deyişle GLM, kitle ortalamasının bir bağ fonksiyonu ile doğrusal tahmin ediciye bağlı olmasına izin veren lineer modellerin genelleştirilmiş halidir [1]. Nelder ve Wedderburn [19] tarafından ilk olarak kullanılan GLM, üç temel bileşenden meydana gelmektedir. Bu bileşenler, yanıt değişkeninin dağılımı, lineer tahmin edicilerin bulunduğu sistematik kısım ve bağ fonksiyonudur. Modelin sistematik kısmı, X_1, X_2, \dots, X_p girdi değişkenlerini içerir. GLM’de yanıt değişkeninin dağılımı, Normal, Poisson, Binom, Üstel ve Gamma dağılımlarını içeren üstel ailenin bir üyesi olmalıdır. Yanıt değişkeninin sürekli olduğu durumlarda Gamma, Ters Gauss dağılımı ve Normal dağılım, kesikli olduğu durumlarda Poisson, Bernoulli ve Binom dağılımı üstel aile örnekleridir [20]. GLM’de, Y ’nin dağılımı üstel ailenin genel formunda

$$f(y|\theta, \phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\right\} \quad (1)$$

biçiminde yazılır [21]. Burada, $a(\cdot)$, $b(\cdot)$ ve $c(\cdot)$ bazı özel fonksiyonlardır. Dağılımın ortalamasının (μ) bir fonksiyonu olan θ , kanonik parametre olarak adlandırılır ve dağılımın konumu hakkında bilgi içerir. ϕ , yayılım veya ölçek parametresi ve $c(y, \phi)$, gözlemler ile yayılım parametresinin bir fonksiyonudur. Üstel dağılım ailesinin olabilirlik fonksiyonunun logaritmasının, θ 'ya göre birinci ve ikinci türevleri alınıp sıfıra eşitlendiğinde, Y 'nin ortalama ve varyansı sırasıyla

$$E(Y) = \mu = b'(\theta) \quad (2)$$

ve

$$Var(Y) = b''(\theta)a(\phi) \quad (3)$$

olur. GLM'de temel amaç, yanıt değişkeninin beklenen değerinin uygun bir fonksiyonu için bir model geliştirmektir. Bağ ya da link fonksiyonu olarak adlandırılan bu fonksiyon, yanıt değişkeninin ortalaması ile doğrusal tahmin ediciler arasında ilişki kurulmasını sağlar ve

$$g(\mu) = \eta = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p = \mathbf{X}^T \boldsymbol{\beta} \quad (4)$$

biçiminde ifade edilir. Eşitlik (4) ile ifade edilen bağ fonksiyonu $g(\bullet)$, monoton ve türevlenebilir olmalıdır. Bu fonksiyonun 1-1 olduğu anlamına gelir. Bu nedenle bağ fonksiyonunun ters dönüşümü vardır. Ters fonksiyon $g^{-1}(\bullet)$, ortalama fonksiyon olarak da adlandırılır, $\mu = g^{-1}(\eta) = g^{-1}(\mathbf{X}^T \boldsymbol{\beta})$ 'dir [22]. Çizelge 2'de, GLM ile kullanılan yaygın üstel aile dağılımları ve bu dağılımlarda kullanılan bağ fonksiyonları verilmiştir.

Çizelge 2. GLM ile kullanılan yaygın üstel aile dağılımları, bağ ve varyans fonksiyonları [23]

<i>Dağılım</i>	<i>Bağ fonksiyonu</i>	<i>Varyans fonksiyonu</i>
Normal	$\eta = \mu$ (özdeş bağ)	1
Binom	$\eta = \log\left(\frac{\mu}{1-\mu}\right)$ (logit bağ)	$\mu(1-\mu)$
Poisson	$\eta = \log(\mu)$ (log bağ)	μ
Gamma	$\eta = \mu^{-1}$ (ters bağ)	μ^2
Ters Normal	$\eta = \mu^{-2}$ (ters kare bağ)	μ^3

Bağ fonksiyonunun özdeş olması durumunda dağılım Normal olur ve değişkenler arasındaki ilişki klasik bilinen regresyon analizi ile oluşturulur. Çoklu doğrusal regresyon modelinin en genel hali

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_p X_p + \varepsilon \quad (5)$$

biçiminde tanımlanır. Burada β_j , $j = 1, 2, \dots, p$ parametreleri regresyon katsayıları ve ε , hata terimidir. Polinom regresyon, çoklu doğrusal regresyonun özel bir durumudur. Polinom modelleri etkili ve esnek bir eğri uydurma tekniğidir [24]. k değişkenli ikinci dereceden bir polinom modeli

$$Y = \beta_0 + \sum_{i=1}^k \beta_i X_i + \sum_{i=1}^k \beta_{ii} X_i^2 + \sum_{i=1}^k \sum_{i < j}^k \beta_{ij} X_i X_j + \varepsilon, \quad i = 1, 2, \dots, k, \quad j = 1, 2, \dots, k, \quad i \neq j \quad (6)$$

biçiminde verilen doğrusal regresyon modelidir. Eşitlik (6)'daki model parametreleri En Küçük Kareler (EKK) (Ordinary Least Squares-OLS) yöntemi kullanılarak tahmin edilebilir. Parametreler tahmin edildiğinde elde edilen tahmini yanıt modeli

$$\hat{Y} = \hat{\beta}_0 + \sum_{i=1}^k \hat{\beta}_i X_i + \sum_{i=1}^k \hat{\beta}_{ii} X_i^2 + \sum_{i=1}^k \sum_{i < j}^k \hat{\beta}_{ij} X_i X_j, \quad i=1,2,\dots,k, \quad j=1,2,\dots,k, \quad i \neq j \quad (7)$$

olur.

3. Karma veri modelleri için çok amaçlı optimizasyon

İki ve ikiden fazla sayıda amaç fonksiyonunun eş anlı optimal değerinin elde edilmesini hedefleyen optimizasyon modeli ÇAO modeli olarak adlandırılır. ÇAO'da amaç fonksiyonlarının hepsini aynı anda optimize etmek mümkün değildir. Amaç fonksiyonlarının birbiriyle çelişmesi durumunda etkin, uzlaşık, baskın çözümler olarak adlandırılan Pareto çözümler çok amaçlı modelin optimizasyonunu karakterize eder [25].

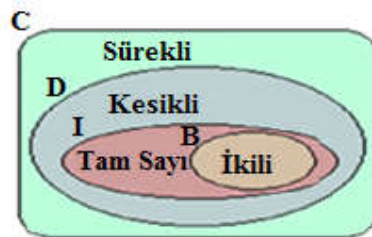
ÇAO problemi genel halde

$$\begin{aligned} \min / \max \quad & \mathbf{f} = [f_1(\mathbf{x}) \ f_2(\mathbf{x}) \ \dots \ f_r(\mathbf{x})] \\ & h_i(\mathbf{x}) = 0, \quad i = 1, 2, \dots, k \\ & g_j(\mathbf{x}) \{ \leq, \geq \} 0, \quad j = 1, 2, \dots, m \\ & \mathbf{x} \in \mathbb{R}^n \end{aligned} \quad (8)$$

biçiminde tanımlanır. Burada, $\mathbf{x} = [x_1 \ x_2 \ \dots \ x_n]$ biçiminde tanımlı girdi değişkenleri vektörü, $\mathbf{f} = [f_1 \ f_2 \ \dots \ f_r]$ amaç fonksiyonu vektörü, $h_i, i = 1, 2, \dots, k$ eşitlik ve $g_j, j = 1, 2, \dots, m$ eşitsizlik biçiminde tanımlı kısıt fonksiyonlarıdır. Optimizasyon problemlerinde değişkenlerin bazılarının sürekli ve bazılarının kesikli değerli olması durumunda problem, karma kesikli optimizasyon problemi olarak adlandırılır. Genel kısıtlı karma kesikli ÇAO problemi

$$\begin{aligned} \min / \max \quad & [f_1(\mathbf{x}) \ f_2(\mathbf{x}) \ \dots \ f_r(\mathbf{x})] \\ & h_i(\mathbf{x}) = 0, \quad i = 1, 2, \dots, k \\ & g_j(\mathbf{x}) \{ \leq, \geq \} 0, \quad j = 1, 2, \dots, m \\ & \mathbf{x} \in \mathbb{R}^B \cup \mathbb{R}^I \cup \mathbb{R}^D \cup \mathbb{R}^C \end{aligned} \quad (9)$$

biçiminde olur. Burada, $\mathbb{R}^B, \mathbb{R}^I, \mathbb{R}^D$ ve \mathbb{R}^C sırasıyla ikili, tam sayılı, kesikli ve sürekli değişkenlerin destek kümeleridir. Hem \mathbb{R}^D hem de \mathbb{R}^C mevcut olduğunda problem karma kesikli optimizasyon problemine dönüşür. Şekil 1'de karma veriler için ilişki şeması yer almaktadır. Şekil 1'den görüldüğü gibi tam sayılı ve ikili değişkenler, kesikli değişkenlerin bir alt kümesi olarak kabul edilir. Bu nedenle, hem tam sayılı hem de kesikli değişkenleri ifade etmek için genel olarak kesikli değişkenler terimi kullanılır.



Şekil 1. Karma veriler için ilişki şeması

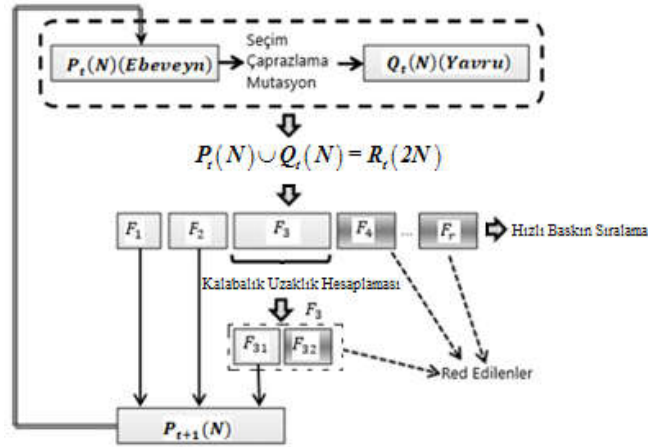
3.1. NSGA-II

ÇAO problemlerinde bütün amaç fonksiyonlarına aynı anda optimal değer veren bir çözüm vektörü bulunamayacağından ÇAO'da optimal çözüm yerine etkin, uzlaşık ve baskın çözümler olarak adlandırılan Pareto çözümler önem kazanır. Pareto çözüm kümesi, amaç fonksiyonları için uzlaşık çözümler içerir. ÇAO problemlerinin çözümü için birçok yapay zeka optimizasyon algoritması geliştirilmiştir. NSGA-II, en çok kullanılan popülasyon tabanlı bir yapay zeka optimizasyon algoritmasıdır. İlk olarak Goldberg [26] tarafından önerilen NSGA-II'nin Srinivas ve Deb'in [27] uygulamaları ile eksiklikleri giderilerek Deb ve ark. [28] tarafından geliştirilip gerçek Pareto yüzey yakınında çok daha iyi çözüm dağılımı veren bir algoritma olarak literatüre kazandırılmıştır. NSGA-II, ayarlanabilir parametre değerleri belirlendikten sonra bir başlangıç çözüm popülasyonu ile aramaya başlar. Uzlaşık çözümleri bulmak için genetik operatörleri kullanarak arama uzayında olasılıksal bir keşif uygular. Ayarlanabilir parametre değerlerinin seçimi ve kullanıcı tarafından seçilen operatörler, GA'nın performansı üzerinde oldukça etkilidir. Bu parametrelerin her kombinasyonu farklı optimum çözümlerle sonuçlanabilir. Ayarlanabilir parametrelerin farklı kombinasyonlarının uygun biçimde belirlenmesi amacıyla Türkşen ve Akgün [29] çalışmalarında Taguchi tasarımı kullanmışlardır. Ardışık nesillerdeki bireysel çözümlerin uygunluğu seçim, çaprazlama ve mutasyon yoluyla artırılır [26]. Çizelge 3'te GA'da kullanılan bazı terimlerin optimizasyon terminolojisindeki karşılıkları verilmiştir.

Çizelge 3. GA'da kullanılan terimlerin optimizasyondaki karşılıkları [30]

Biyoloji / Genetik	Optimizasyon
Popülasyon	Aday çözümler kümesi
Birey / Kromozom	Kodlanmış aday çözüm
Gen	Tasarım / Karar değişkeni
Uygunluk fonksiyon değeri	Amaç fonksiyon değeri
Çevre	Kısıtlar
Kuşak / Generasyon / Nesil	Döngü / İterasyon / Yineleme

NSGA-II'nin diğer çok amaçlı GA'lara göre üstünlüğü, hızlı baskın sıralama ve kalabalık uzaklığı yaklaşımlarıdır. Bu yaklaşımlarla sıralı baskın yüzeyler oluşturularak, baskın çözümler kümesi olarak bilinen Pareto çözüm kümesinde farklı seçenek çözümler elde edilir. NSGA-II'de Pareto çözümlerin sıralanmasında, her bir çözüm için baskınlık sayacı ve ilgili çözümün baskın olduğu çözüm kümesi kullanıldığından çözümlerin sıralanması NSGA'ya göre daha hızlı olmaktadır. Bu nedenle tanımlanan sıralama algoritması hızlı baskın sıralama algoritması olarak adlandırılır. Pareto çözüm kümesi içerisinde çözümlerin dağılımı ve çeşitliliği için kalabalık uzaklığı yaklaşımı kullanılır. Bir çözümün kalabalık uzaklığı, o çözümün bulunduğu yüzeyde ilgili çözüm ile komşu çözümleri arasındaki uzaklıktır [31]. Şekil 2'de, NSGA-II'nin hızlı baskın sıralama ve kalabalık uzaklık işleyişi görülmektedir. Mevcut popülasyon ($P_t(N)$) ile oluşturulan yavru popülasyon ($Q_t(N)$) birleştirilir, $P_t(N) \cup Q_t(N) = R_t(2N)$. Oluşturulan $R_t(2N)$ popülasyonundaki tüm çözümlerin birlikte değerlendirilmesi ile seçkinlik (elitizm) işlemi gerçekleştirilir. Hızlı baskın sıralama algoritması kullanılarak, $R_t(2N)$ 'deki baskın F_1, F_2, \dots, F_r yüzeyleri belirlenir. Tüm yüzeylerde sıralanan çözümlerin kalabalık uzaklığı hesaplanır. İkili turnuva seçim yöntemi kullanılarak $R_t(2N)$ 'deki ilk N birey seçilerek yeni nesil ebeveyn popülasyonu $P_{t+1}(N)$ oluşturulur [32].



Şekil 2. NSGA-II'nin hızlı baskın sıralama ve kalabalık uzaklık hesaplama işleyişi

ÇAO problemlerinin NSGA-II ile çözümü için adımları aşağıdaki gibidir.

Adım 0: NSGA-II'nin ayarlanabilir parametrelerinin belirlenmesi

NSGA-II'nin ayarlanabilir parametre değerleri girdi değişken sayısı (p), popülasyon büyüklüğü (N), çaprazlama olasılığı (Pr_{cr}), çaprazlama indeksi (η_{cr}), mutasyon olasılıkları (Pr_{mut}) ve mutasyon indeksi (η_{mut}), yineleme sayısı ($t = 1, 2, \dots, n_{gen}$) uzman görüşüne göre tanımlanır. Burada, yineleme sayısı NSGA-II için durdurma koşulu olarak kullanılır.

Adım 1: Başlangıç popülasyonunun oluşturulması ve uygunluk fonksiyonu değerlerinin hesaplanması

Değişkenler reel değerli alınarak başlangıç popülasyonu rastgele oluşturulur ve bireylerin uygunluk fonksiyon değerleri hesaplanır.

Adım 2: Uygunluk fonksiyon değerlerine göre bireylerin sıralanması ve kalabalık uzaklık değerlerinin hesaplanması

Bireyler uygunluk fonksiyon değerlerine göre sıralanır ve kalabalık uzaklık değerleri hesaplanır. Bireyler sıralamaya ve kalabalık uzaklığa göre seçildiğinden popülasyondaki tüm bireylere bir kalabalık uzaklık değeri atanır.

Adım 3: Genetik operatörlerin uygulanması

Genetik operatör uygulamaları seçim, çaprazlama ve mutasyon aşamalarından oluşur. Adım 0'da belirtilen ayarlanabilir parametre değerleri kullanılarak genetik operatör işlemleri uygulanır.

Adım 3.1: Baskınlık kriterine göre sıralanan bireylere kalabalık uzaklık değeri atandıktan sonra seçim, bir kalabalık karşılaştırma operatörü (\prec) kullanılarak gerçekleştirilir. Bireyler, ikili turnuva seçimi kullanılarak seçilir. Seçim, iki kritere göre yapılır. Birincisi ve en önemlisi, çözümlerin bulunduğu yüzey veya sıradır. Daha düşük sıralamaya sahip bireyler seçilir. İkinci olarak, iki bireyin sıralaması aynı ise, kalabalık uzaklığı karşılaştırılır. Kalabalık uzaklığı daha büyük olan bireyler seçilir.

Adım 3.2: Seçilen ebeveynlerden Simüle Edilmiş İkili (SBX) çaprazlama operatörü kullanılarak yavrular üretilir.

Adım 3.3: SBX çaprazlama operatörü ile üretilen bazı yavrulara polinomsal mutasyon operatörü uygulanır.

Adım 4: Yeni popülasyonun oluşturulması

N ve $2N$ birey sayısı içeren yeni popülasyonlar oluşturulur.

Adım 4.1: Yeni $Q_t(N)$ (yavru) popülasyonu oluşturulur.

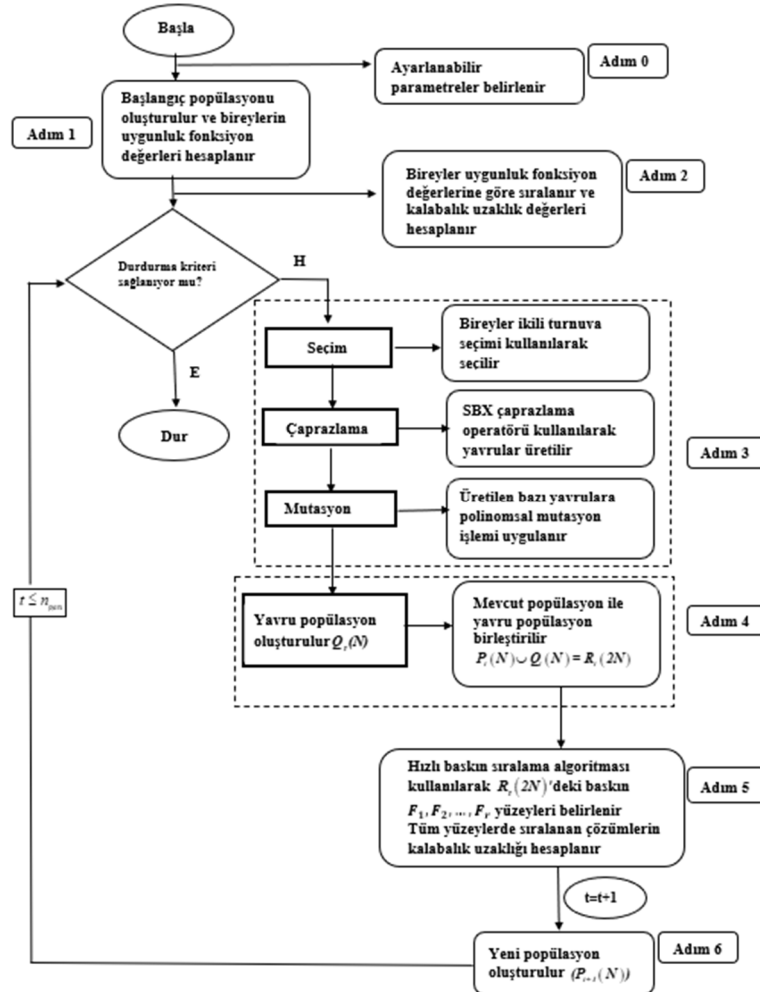
Adım 4.2: Mevcut popülasyon ile oluşturulan yavru popülasyon birleştirilir,

$P_t(N) \cup Q_t(N) = R_t(2N)$. Oluşturulan $R_t(2N)$ popülasyonundaki tüm çözümlerin birlikte değerlendirilmesi ile seçkinlik işlemi gerçekleştirilir.

Adım 5: Baskın yüzeylerin oluşturulması ve oluşturulan yüzeylerde sıralanan çözümlerin kalabalık uzaklıklarının hesaplanması

Hızlı baskın sıralama algoritması kullanılarak, $R_t(2N)$ 'deki baskın F_1, F_2, \dots, F_r yüzeyleri belirlenir. Tüm yüzeylerde sıralanan çözümlerin kalabalık uzaklığı hesaplanır. Kalabalık uzaklık değerlerine göre başlangıçta belirlenen popülasyon sayısı (N) elde edilecek biçimde sıralı yüzeylerden bireyler seçilerek yeni popülasyon oluşturulur.

Adım 5.1: İkili turnuva seçim yöntemi kullanılarak $R_t(2N)$ 'deki ilk N birey seçilerek yeni nesil ebeveyn popülasyon $P_{t+1}(N)$ oluşturulur. Başlangıçta belirlenen yineleme sayısına ulaşılmış ise algoritma sonlandırılır. Aksi halde, Adım 3'e gidilir. Şekil 2'de hızlı baskın sıralama ve kalabalık uzaklık hesaplama işleyişi gösterilmiştir. NSGA-II'nin akış şeması Şekil 3'te verilmiştir.



Şekil 3. NSGA-II akış şeması

3.2. MDNSGA-II

NSGA-II'nin, değişken gösterimi, başlangıç popülasyonu ve genetik operatörlere ilişkin adımlarında uyarlamalar yapılarak karma veri içeren çok yanıtlı problemler için Pareto çözüm kümesi elde edilmeye çalışılmıştır. Yapılan uyarlamalarla algoritma, MDNSGA-II olarak adlandırılmıştır.

3.2.1 Değişkenlerin gösterimi

Değişkenlerin gösteriminde, problemin doğasına ve verilerin yapısına göre kodlama türü seçilir. Başlıca kodlama türleri, ikili, sekizli, onaltılı, gri, permütasyon kodlama, değer kodlaması ve ağaç kodlamasıdır. Literatürde karma veri içeren problemler için ikili kodlama tercih edilmiştir. Çözümlerin hesaplama süresini kısaltması ve karma veri içeren problemler için uyarlanabilmesi sebebiyle bu çalışmada değer kodlaması kullanılmıştır. Değer kodlamasında sürekli ve tam sayılı değişkenlerin kendi değeriyle çalışılır. Fakat, değişkenin kesikli reel değer alması durumunda değişken gösteriminde uyarlama yapılması gerekir. Kesikli değişkenin tanım kümesindeki her bir değer küçükten büyüğe sıralanır. Sıralanan değerler sıra numarası ile indekslenir. Böylece kesikli değişkenin her bir değeri bir indeks değeriyle ifade edilmiş olur. Yapılan bu uyarlama, kesikli reel değer alan değişkenlerin indekslenmesi olarak tanımlanır.

Örneğin, kesikli bir değişkenin tanım kümesi $S \subset \mathbb{R}^D$, kesikli değişkenin aldığı değer kümesi $S = \{s_1, s_2, \dots, s_q\}$ olmak üzere, kesikli değişkenin her bir değerinin I tam sayılar kümesinde pozitif bir tam sayıya eşlenmesi

$$\begin{array}{l} S = \{s_1, s_2, s_3, \dots, s_q\} \\ \quad \downarrow \downarrow \downarrow \dots \downarrow \\ I = \{1, 2, 3, \dots, q\} \end{array} \quad (10)$$

biçiminde gösterilir. Buna göre, s_1 için 1, s_2 için 2, ..., s_q için q indekslemesi yapılır. Böylece kesikli değişkenin aldığı değerler yerine değişkenin indeksleri kullanılır.

3.2.2 Başlangıç popülasyonunun oluşturulması

MDNSGA-II için başlangıç popülasyonu rastgele sayı üretici kullanılarak oluşturulur. α , $[0,1]$ aralığından üretilen bir rastgele sayı olmak üzere, lb , sürekli veya tam sayılı değişkenin alt sınırı ve ub , sürekli veya tam sayılı değişkenin üst sınırı olsun. Sürekli ve tam sayılı değişkenler için belirlenen $[lb, ub]$ tanım aralıklarında sırasıyla

$$x = lb + \alpha(ub - lb) \quad (11)$$

ve

$$x = lb + \text{round}(\alpha(ub - lb)) \quad (12)$$

biçiminde tanımlı eşitliklerle başlangıç popülasyonu bileşenleri elde edilir. İkili (0,1) değerli değişkenler için ise

$$x = \text{round}(\alpha) \quad (13)$$

eşitliği ile başlangıç popülasyonu bileşen değerleri elde edilir. Kesikli değişkenler için başlangıç popülasyonu oluşturulurken değişken gösteriminde yapılan indeksleme işlemine göre uyarlama yapılır.

Kesikli değişken değerlerinin indeksleri için rastgele sayı üretilir. r sıra numarası ve tam sayı eşlemesi yapılan kesikli değişkenin tanım aralığı $[1, q]$ olmak üzere kesikli değişken için üretilecek rastgele sayı

$$x^{(r)} = \text{round}(1 + \alpha(q-1)) \quad (14)$$

biçiminde elde edilir. $S = \{s_1, s_2, \dots, s_q\}$ kesikli değişken değerleri kümesi olmak üzere $\#(S)=q$ 'dur. Buna göre $[1, q]$ aralığından q sayıda rastgele indeks numaraları (I_q) üretilir. Üretilen I_q indeks numaralarına göre, S kümesinde karşılık gelen kesikli değer seçilerek yeni bir kesikli değer kümesi oluşturulur.

3.2.3 Genetik operatörlerin uyarlanması

Genetik operatörler seçim, çaprazlama ve mutasyon aşamalarından oluşur. MDNSGA-II'de seçim operatörü adımı herhangi bir uyarlama yapmaya gerek yoktur. Bu çalışmada, uyarlanan NSGA-II genetik operatörleri olarak Turnuva seçimi, SBX çaprazlama ve Polinomsal mutasyon kullanılmıştır. Genetik operatörlerin çaprazlama ve mutasyon aşamalarında yapılan uyarlamalar aşağıda verilmiştir.

SBX çaprazlama operatörü: Bu operatör reel değerli değişkenler için oluşturulmuştur. t . yinelemede mevcut $x^{(1,t)}$ ve $x^{(2,t)}$ çözümleri kullanılarak yapılan hesaplamalarla iki yeni çözüm elde edilir. $U \in [0,1]$ aralığından rastgele bir sayı oluşturulur. Burada η_{cr} , belirlenen bir aralıkta tanımlı, pozitif reel sayı değerli çaprazlama dağılım indeksi olmak üzere

$$\beta_q = \begin{cases} (2U)^{\frac{1}{\eta_{cr}+1}}, & U \leq 0.5 \\ \left(\frac{1}{2(1-U)}\right)^{\frac{1}{\eta_{cr}+1}}, & d.y. \end{cases} \quad (15)$$

ile tanımlı fonksiyona göre β_q hesaplanır. Yavru çözümler ise

$$\begin{aligned} x^{(1,t+1)} &= 0.5 \left[(1 + \beta_q) x^{(1,t)} + (1 - \beta_q) x^{(2,t)} \right] \\ x^{(2,t+1)} &= 0.5 \left[(1 - \beta_q) x^{(1,t)} + (1 + \beta_q) x^{(2,t)} \right] \end{aligned} \quad (16)$$

biçiminde elde edilir. Sürekli değişkenler için Eşitlik (16) kullanılarak yavru çözümler oluşturulur.

MDNSGA-II'de çaprazlama operatörü kullanılarak oluşturulan yavru çözümler, tam sayı ve kesikli değişkenin tanım kümesi dışında değerler alabilir. Bu nedenle, yavru çözüm oluşturulurken çaprazlama operatöründe yapılacak değişikliklerle değişkenin tanım kümesi içinde değer alması sağlanır. Tam sayılı değişkenler için yavru çözümler

$$\begin{aligned} x^{(1,t+1)} &= \text{round} \left(0.5 \left[(1 + \beta_q) x^{(1,t)} + (1 - \beta_q) x^{(2,t)} \right] \right) \\ x^{(2,t+1)} &= \text{round} \left(0.5 \left[(1 - \beta_q) x^{(1,t)} + (1 + \beta_q) x^{(2,t)} \right] \right) \end{aligned} \quad (17)$$

biçiminde oluşturulur. Burada, tam sayılı değerlerin ardışık olması durumunda Eşitlik (17)'nin kullanımı uygundur. Aksi halde kesikli reel sayı değerli değişkenler için önerilen uyarlamanın yapılması uygun olur. Eşitlik (16)'da SBX çaprazlama operatörünün uygulanmasıyla elde edilen yavru çözümler kesikli değişkenlerin tanım kümesinde yer almayabilir. Bu durumda elde edilen yavru çözüm değerlerinin

$(x^{(1,r+1)}$ ve $x^{(2,r+1)})$, $S = \{s_1, s_2, \dots, s_q\}$ kümesindeki elemanlardan biri olması istenir. Eğer, $x^{(1,r+1)} \in [s_i, s_{i+1}]$, $i = 1, 2, \dots, q-1$ ise $s'_i = \frac{s_i + s_{i+1}}{2}$ hesaplanır. s'_i değeri ile Eşitlik (16) ile elde edilen yavru çözüm değerleri karşılaştırılır. Yapılan karşılaştırmada yavru çözüm değerinin s_i ya da s_{i+1} , $i = 1, 2, \dots, q-1$ değerlerinden hangisine daha yakın olduğu belirlenir ve yavru çözüm değerlerine yakın olan kesikli değişken değeri alınır. Böylece, çaprazlama aşaması sonrasında elde edilen yavru çözüm

$$x^{(1,r+1)} = \begin{cases} s_i & , \quad 0.5 \left[(1 + \beta_q) x^{(1,r)} + (1 - \beta_q) x^{(2,r)} \right] \leq s'_i \\ s_{i+1} & , \quad d.y. \end{cases} \quad (18)$$

biçiminde elde edilir. $x^{(2,r+1)}$ 'de Eşitlik (18)'e benzer olarak hesaplanır.

Polinomsal mutasyon operatörü: SBX ile elde edilen yavru çözümlere mutasyon aşaması uygulanarak çözüm çeşitliliği sağlanır. Mevcut çözüm $x^{(r)}$ 'den, yeni çözüm

$$x^{(r+1)} = x^{(r)} + (ub - lb)\delta \quad (19)$$

biçiminde elde edilir. Burada, lb ve ub , sırasıyla çözümler için belirlenen alt ve üst sınır değerleri, $r \in [0, 1]$ aralığından rastgele bir sayı ve η_{mut} mutasyon dağılım indeksi olmak üzere, bir polinomsal dağılım kullanılarak Eşitlik (19)'da belirtilen değişim

$$\delta = \begin{cases} (2r)^{\frac{1}{\eta_{mut}+1}} - 1, & r \leq 0.5 \\ 1 - [2(1-r)]^{\frac{1}{\eta_{mut}+1}}, & d.y. \end{cases} \quad (20)$$

biçiminde hesaplanır [30]. Sürekli değişkenler için Eşitlik (19) kullanılarak yeni çözüm elde edilir.

MDNSGA-II'de mutasyon aşaması uygulanarak oluşturulan yeni çözümlerde yapılacak değişikliklerle değişkenin tanım kümesi içinde değer alması sağlanır. Çaprazlamada yapılan uyarlamalar benzer şekilde mutasyon aşamasında da yapılarak tam sayılı değişkenler için yeni çözüm

$$x^{(r+1)} = \text{round}(x^{(r)} + (ub - lb)\delta) \quad (21)$$

ve kesikli değişkenler için $x^{(r+1)} \in [s_i, s_{i+1}]$, $i = 1, 2, \dots, q-1$ ise $s'_i = \frac{s_i + s_{i+1}}{2}$ hesaplanarak mutasyon aşaması sonrasında elde edilen yeni çözüm

$$x^{(1,r+1)} = \begin{cases} s_i & , \quad x^{(r)} + (ub - lb)\delta \leq s'_i \\ s_{i+1} & , \quad d.y. \end{cases} \quad (22)$$

biçiminde elde edilir. Şekil 3'te verilen NSGA-II akış şemasında Adım 0 (Değişkenlerin gösterimi), Adım 1 (Başlangıç popülasyonunun oluşturulması) ve Adım 3'te (Genetik operatörler) yapılan uyarlamalarla MDNSGA-II oluşturulur.

4. Uygulama

Karma veri içeren çok yanıtli optimizasyon problemlerinin çözümü için bu çalışmada önerilen MDNSGA-II yöntemi, literatürden çok iyi bilinen iki veri seti (enerji ve gıda alanı) üzerinde uygulanmıştır. İlk veri seti, optimizasyon alanındaki çalışmalarda kabul gören veri tabanından alınan enerji verimliliği ile ilgilidir

[33]. Bu veri setinde, binaların enerji verimliliğine ilişkin ısıtma yükü (Y_1) ve soğutma yükü (Y_2) gereksinimlerinin, bina özniteliklerinin bir fonksiyonu olarak değerlendirilmesi araştırılmıştır. İkinci veri seti ise, gıda alanında literatürde tanımlı çok yanıtlı deneysel çalışma ile edilen bir veri setidir. Veri setleri için GLM ve SUR yöntemleri ile uygun tahmini yanıt modelleri oluşturularak, karma veriler için modelleme aşaması sağlanmıştır. Veri setlerinde yer alan yanıt değişkenlerinin modellenmesinde RStudio ve IBM SPSS Statistics 24 programları kullanılmıştır. Optimizasyon aşamasında, MDNSGA-II yöntemi ile MATLAB R2023a programı kullanılarak veri setleri için Pareto çözüm kümesi elde edilmiştir.

4.1. Enerji verimliliği veri seti

Veri kümesi, iki gerçek değerli yanıt (ısıtma yükü ve soğutma yükü) ilişkin 768 gözlem ve 8 özellik (girdi değişkenleri) içermektedir. Girdi değişkenleri ile yanıtlar arasındaki fonksiyonel ilişkinin belirlenmesi ve yanıt değişkenlerini eş anlamlı minimum yapacak girdi değişken değerlerinin belirlenmesi istenmektedir. Çizelge 4'te, 8 girdi değişkeni ve 2 yanıt değişkeninden oluşan 768 gözlemlilik karma veri seti örneği yer almaktadır.

Çizelge 4. Enerji verimliliği veri seti

No	Girdi Değişkenler								Yanıt Değişkenleri	
	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	Y_1	Y_2
1	0.98	514.5	294	110.25	7	2	0	0	15.55	21.33
2	0.98	514.5	294	110.25	7	3	0	0	15.55	21.33
3	0.98	514.5	294	110.25	7	4	0	0	15.55	21.33
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
766	0.62	808.5	367.5	220.5	3.5	3	0.4	5	16.44	17.11
767	0.62	808.5	367.5	220.5	3.5	4	0.4	5	16.48	16.61
768	0.62	808.5	367.5	220.5	3.5	5	0.4	5	16.64	16.03

Çizelge 5'te girdi ve yanıt değişkenlerinin değişken türleri ve tanım aralığı görülmektedir. Girdi değişkenlerinden yapı yüksekliği ve yönelimi ile cam alanı ve cam alan dağılımı açısından bina özniteliklerinin farklılık gösterdiği görülmektedir. Yapı yüksekliği (X_5), sürekli bir değişken olmasına rağmen bu veri setinde tanım kümesinin $\{3.5, 7\}$ değerlerinden oluştuğu görülmektedir. Bu durum yapı yüksekliğinin bu veri seti için kesikli değişken olarak ele alındığını göstermektedir. Bina tasarımında yapı yüksekliği sadece iki değerden birini alabilmektedir. X_6 değişkeni yapının hangi cephede (kuzey, güney, doğu, batı) yer aldığını belirtmektedir. X_5, X_6, X_7 ve X_8 değişkenlerinin tanım kümelerine bakıldığında kesikli ve/veya tam sayılı değişkenlerden oluştuğu görülmektedir.

Çizelge 5. Enerji verimliliği veri setinin girdi değişkenleri, değişken türü ve tanım kümesi

	Değişkenler	Değişken Türü	Tanım Kümesi
Girdi	Bağıl kompaktlık (X_1)	Sürekli	[0.62, 0.98]
	Yüzey alanı (X_2)	Sürekli	[514.5, 808.5]
	Duvar alanı (X_3)	Sürekli	[245, 416.5]
	Çatı alanı (X_4)	Sürekli	[110.25, 220.50]
	Yapı yüksekliği (X_5)	Kesikli	{3.5, 7}
	Yapı yönelimi (X_6)	Kesikli, Tam sayı	{2, 3, 4, 5}
	Cam alanı (X_7)	Kesikli	{0, 0.10, 0.25, 0.40}
	Cam alan dağılımı (X_8)	Kesikli, Tam sayı	{0, 1, 2, 3, 4, 5}
Yanıt	Isıtma yükü (Y_1)	Sürekli	$Y_1 \geq 0$
	Soğutma yükü (Y_2)	Sürekli	$Y_2 \geq 0$

Binaların enerji tüketimi göz önünde bulundurulduğunda enerjinin kullanımı için verimli bina tasarımı önemlidir. İhtiyaç duyulan ısıtma ve soğutma ekipmanının özelliklerini belirlemek için ısıtma ve soğutma yüklerinin hesaplanması gerekir. Isıtma ve soğutma yüklerinin enerji verimliliği bakımından minimum olması istenir. Bu çalışmada, Tsanas ve Xifara'nın [34] çalışmasında uygulanan IRLS (Iteratively Reweighted Least Squares) yöntemi ile elde edilen regresyon modeli kullanılmıştır. Varyans homojenliği varsayımının sağlanamadığı durumlarda IRLS'nin kullanımı daha uygundur. IRLS ile regresyon katsayılarındaki ağırlıklar ayarlanarak regresyon eğrisi oluşturulurken aykırı değerlerin etkisi azaltılır. Böylece daha geliştirilmiş bir en küçük kareler tahmini sağlanır.

Çizelge 6'da, Y_1 ve Y_2 yanıtlarının normallik varsayımı için Kolmogrov-Smirnov testi sonuçları görülmektedir. %95 güven düzeyinde (p -değeri= $0.001 < 0.05$) Y_1 ve Y_2 yanıtlarına ilişkin gözlem değerlerinin Normal dağılım göstermediği söylenir.

Çizelge 6. Yanıtlar için Kolmogrov-Smirnov testi sonuçları

<i>Yanıt Değişkenleri</i>	<i>p-değeri</i>
Y_1	.001
Y_2	.001

Y_1 ve Y_2 yanıtları için elde edilen tahmini yanıt modelleri sırasıyla

$$\hat{Y}_1 = -4.75X_1 - 0.03X_2 + 0.07X_3 - 3.44X_5 - 0.01X_6 + 18.13X_7 + 0.09X_8 \quad (23)$$

ve

$$\hat{Y}_2 = -9.02X_1 - 0.01X_2 + 0.04X_3 - 4.30X_5 - 0.12X_6 + 14.49X_7 + 0.03X_8 \quad (24)$$

dır. \hat{Y}_1 ve \hat{Y}_2 yanıtlarını eş anlı minimize eden optimal girdi değerlerini elde edebilmek için optimizasyon aşamasında MDNSGA-II uygulanmıştır. Her bir yanıt fonksiyonu bir amaç fonksiyonu olarak değerlendirildiğinde çok yanıtlı optimizasyon problemi

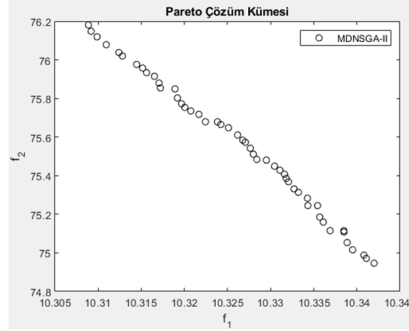
$$\begin{aligned} \min f_1 &= \hat{Y}_1(\mathbf{x}) \\ \min f_2 &= \hat{Y}_2(\mathbf{x}) \\ \mathbf{x} &\in S \end{aligned} \quad (25)$$

biçiminde tanımlı CAO problemine dönüşecektir. Burada, $\{X_1, X_2, \dots, X_8\}$ değişkenlerinin aldığı değerler vektörü \mathbf{x} olarak belirtilmiştir. f_1 ve f_2 amaç fonksiyonları, sırasıyla \hat{Y}_1 ve \hat{Y}_2 tahmini yanıt fonksiyonlarına karşılık gelmektedir. Çizelge 7'de, bu çalışmada uygulanan MDNSGA-II'nin ayarlanabilir parametreleri ve parametrelerin aldığı değerler yer almaktadır.

Çizelge 7. MDNSGA-II'nin ayarlanabilir parametreleri ve parametre değerleri

<i>Ayarlanabilir parametreler</i>	<i>Parametre değerleri</i>
Girdi değişken sayısı (p)	8
Popülasyon büyüklüğü (N)	50
Yineleme sayısı (n_{gen})	100
Çaprazlama operatörü	SBX
Mutasyon operatörü	Polinomsal
Seçim operatörü	Turnuva
Çaprazlama olasılığı (Pr_{cr})	0.90
Çaprazlama indeksi (η_{cr})	20
Mutasyon olasılığı (Pr_{mut})	1/8
Mutasyon indeksi (η_{mut})	20

Şekil 4’te enerji veri seti için elde edilen Pareto çözüm kümesi görülmektedir. Karma veri içeren çok yanıtlı enerji veri seti için MDNSGA-II ile çözüm çeşitliliğinin sağlandığı görülmektedir. Enerji veri seti için f_1 ve f_2 amaç fonksiyonlarına ilişkin Pareto çözüm değerleri Ek-1’de verilmiştir. Ek-1’de görüldüğü gibi, MDNSGA-II ile $\{X_5, X_6, X_7, X_8\}$ değişkenlerinin tanım kümesinden kesikli değerler olarak Pareto çözümlerin elde edilmesi sağlanmıştır.



Şekil 4. Enerji veri seti için f_1 ve f_2 amaç fonksiyonlarına ilişkin Pareto çözüm kümesi

4.2. Gıda veri seti

Bu uygulamada veri seti, Schmidt ve ark. [35] çalışmalarında kullandığı gıda alanından seçilmiştir. Karma veri modelleri için deneyin optimal değeri elde edilmeye çalışılmıştır. Çok yanıtlı deneysel karma veri seti 2 girdi değişkeni ve 4 yanıt değişkeni içermektedir. Sistein jel dokusu (X_1) ve bir tuz çeşidi olan kalsiyum klorür ($CaCl_2$) (X_2) maddelerinin peynir altı suyu konsantresinin yapısal özellikleri ve su tutma özelliği üzerine etkisini incelemek amacıyla bir deney düzenlenmiştir. Burada, sistein ve kalsiyum klorür girdi değişkenleri; sertlik, yapışkanlık, esneklik yapısal özellikleri ile peynir altı suyunda tutulan sıkıştırılabilir su yanıt değişkenleridir. Peynir altı suyu proteini konsantresi sistemlerinin yanıt değişkenleri üzerindeki etkilerini ölçmek için çoklu doğrusal regresyon analizi kullanılmıştır. Çok yanıtlı deneysel karma veri seti için yanıtlar arasında doğrusal ilişkili olması durumunda yanıtlar, SUR (Seemingly Unrelated Regression) yöntemi ile modellenebilir [36]. Çizelge 8’de deneysel çalışma sonucunda elde edilen kodlanmış girdi değişken değerleri ile gözlenen yanıt değişkenlerine ait değerler yer almaktadır.

Çizelge 8. Deneysel çalışma sonucu elde edilen kodlanmış girdi değişken değerleri ve gözlenen yanıt değişken değerleri

No	X_1	X_2	Y_1	Y_2	Y_3	Y_4
1	-1	-1	2.48	0.55	1.95	0.22
2	1	-1	0.91	0.52	1.37	0.67
3	-1	1	0.71	0.67	1.74	0.57
4	1	1	0.41	0.36	1.20	0.69
5	-1.414	0	2.28	0.59	1.75	0.33
6	1.414	0	0.35	0.31	1.13	0.67
7	0	-1.414	2.14	0.54	1.68	0.42
8	0	1.414	0.78	0.51	1.51	0.57
9	0	0	1.50	0.66	1.80	0.44
10	0	0	1.66	0.66	1.79	0.50
11	0	0	1.48	0.66	1.79	0.50
12	0	0	1.41	0.66	1.77	0.43
13	0	0	1.58	0.66	1.73	0.47

Çizelge 9’da gıda veri setinin değişken türü ve tanım aralığı açıklamaları görülmektedir. Burada, X_1 kesikli değişken olup X_2 sürekli değişken olarak tanımlanmıştır.

Çizelge 9. Gıda veri seti için girdi değişkenleri, değişken türü ve tanım kümesi

<i>Girdi Değişkenleri</i>	<i>Değişken Türü</i>	<i>Tanım Kümesi</i>
Sistein jel dokusu (X_1)	Kesikli	{-1.414, -1, 0, 1, 1.414}
Kalsiyum klorür (X_2)	Sürekli	[-1.414, 1.414]

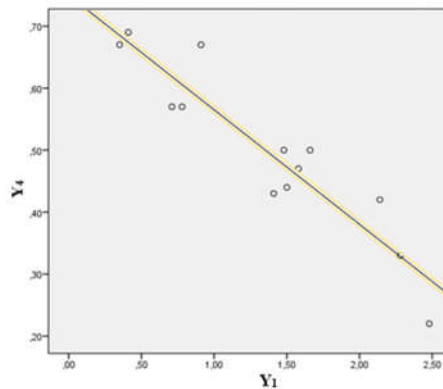
Yanıtlar arasındaki doğrusal ilişkinin incelenmesinden önce yanıtların Normal dağılım varsayımı test edilmiştir. Bu amaçla, Shapiro-Wilk (küçük örneklerde) testi uygulanmıştır. Shapiro-Wilk testi sonucu Çizelge 10’da özetlenmiştir. Çizelge 10’daki p değerlerine bakıldığında %95 güven düzeyinde Y_1 ve Y_4 yanıtlarının Normal dağılımlı olduğu (p -değeri>0.05), Y_2 ve Y_3 yanıtlarının dağılımının ise Normal dağılım göstermediği söylenir (p -değeri<0.05).

Çizelge 10. Gıda veri seti için yanıt değişkenlerinin Normallik testi

<i>Yanıt Değişkenleri</i>	<i>Shapiro Wilk p-değeri</i>
Y_1	0.549
Y_2	0.010
Y_3	0.023
Y_4	0.652

$Y_1 - Y_4$ yanıt değişkenleri için Pearson korelasyon katsayısı -0.932 olup bu yanıtlar arasında ters yönlü, güçlü ve anlamlı bir ilişki olduğu söylenir. Spearman korelasyon testine göre, $Y_1 - Y_3$, $Y_2 - Y_3$ ve $Y_3 - Y_4$ yanıtlarının da sırasıyla, 0.674, 0.69 ve -0.702 büyüklüklerinde doğrusal ilişkili olduğu söylenir. $Y_1 - Y_2$ ve $Y_2 - Y_4$ yanıtlarının ise doğrusal ilişkisiz olduğu görülmüştür. Bu çalışmada Normallik varsayımını sağladığı ve aralarında anlamlı ve güçlü bir ilişki olduğu için $Y_1 - Y_4$ yanıtları ile çalışılmıştır. Ayrıca, $Y_2 - Y_3$ yanıtları arasında anlamlı, pozitif yönlü doğrusal bir ilişki olması nedeniyle bu yanıtların da eş anlı optimizasyonu ile ilgilenilmiştir.

$Y_1 - Y_4$ yanıtlarına ilişkin saçılım grafiği Şekil 5’te verilmiştir. Yanıtlar arasındaki ilişki, ikinci dereceden polinomsal fonksiyonlar kullanılarak SUR yöntemi ile modellenmiştir. Çizelge 11’de, $Y_1 - Y_4$ yanıtlarının SUR yöntemi ile elde edilen parametre tahminleri yer almaktadır.



Şekil 5. $Y_1 - Y_4$ yanıt değişkenine ait saçılım grafiği

Çizelge 11. Y_1 ve Y_4 yanıtlarının SUR yöntemi ile parametre tahminleri

Model Terimi	SUR			
	Y_1	p -değeri	Y_4	p -değeri
Sabit	1.526 (.065)	.001	0.468 (.013)	.001
X_1	-0.575 (.051)	.001	0.131 (.010)	.001
X_2	-0.524 (.051)	.001	0.073 (.010)	.001
X_1^2	-0.171 (.055)	.018	0.026 (.011)	.055
X_2^2	-0.098 (.055)	.12	0.024 (.011)	.076
X_1X_2	0.318 (.073)	.003	-0.083 (.015)	.001

*Standart hatalar parantez içinde belirtilmiştir.

Çizelge 11'den, Y_1 ve Y_4 yanıtları için elde edilen tahmini yanıt modelleri sırasıyla

$$\hat{Y}_1 = 1.526 - 0.575X_1 - 0.524X_2 - 0.171X_1^2 + 0.318X_1X_2 \quad (26)$$

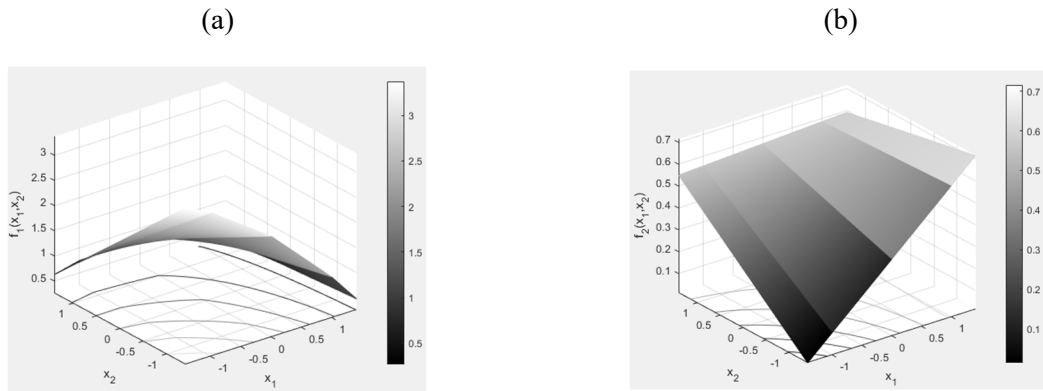
ve

$$\hat{Y}_4 = 0.468 + 0.131X_1 + 0.073X_2 - 0.083X_1X_2 \quad (27)$$

dır. Optimal değeri elde edilmek istenilen her bir yanıt fonksiyonu, bir amaç fonksiyonu olarak değerlendirildiğinde çok yanıtlı optimizasyon problemi

$$\begin{aligned} \max f_1 &= \hat{Y}_1(\mathbf{x}) \\ \min f_2 &= \hat{Y}_4(\mathbf{x}) \\ \mathbf{x} &\in S \end{aligned} \quad (28)$$

biçiminde ÇAO problemine dönüşecektir. Burada, $\{X_1, X_2\}$ değişkenlerinin aldığı değerler vektörü \mathbf{x} olarak belirtilmiştir. f_1 ve f_2 amaç fonksiyonları, sırasıyla \hat{Y}_1 ve \hat{Y}_4 tahmini yanıt fonksiyonlarına karşılık gelmektedir. Elde edilen tahmini yanıt fonksiyonları için yüzey grafikleri Şekil 6.(a)-(b)'de görülmektedir. Buna göre, Şekil 6. (a)'da bir maksimizasyon problemi ve Şekil 6.(b)'de bir minimizasyon problemi ile ilgilenildiği açıktır.



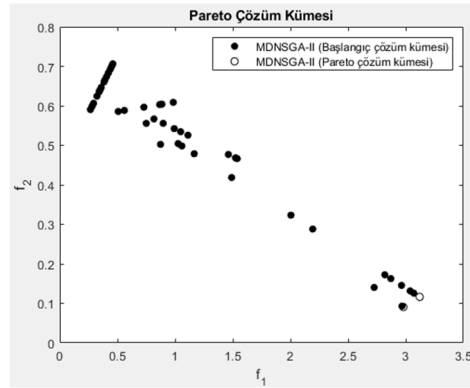
Şekil 6. (a) f_1 amaç fonksiyonu, (b) f_2 amaç fonksiyonu için yüzey grafikleri

Çizelge 12'de, bu çalışmada uygulanan MDNSGA-II'nin ayarlanabilir parametreleri ve parametrelerin aldığı değerler yer almaktadır.

Çizelge 12. MDNSGA-II'nin ayarlanabilir parametreleri ve parametre değerleri

<i>Ayarlanabilir parametreler</i>	<i>Parametre değerleri</i>
Girdi değişken sayısı (p)	2
Popülasyon büyüklüğü (N)	50
Yineleme sayısı (n_{gen})	100
Çaprazlama operatörü	SBX
Mutasyon operatörü	Polinomsal
Seçim operatörü	Turnuva
Çaprazlama olasılığı (Pr_{cr})	0.90
Çaprazlama indeksi (η_{cr})	20
Mutasyon olasılığı (Pr_{mut})	1/2
Mutasyon indeksi (η_{mut})	20

ÇAO probleminin f_1 ve f_2 amaç fonksiyonlarına ilişkin MDNSGA-II ile elde edilen başlangıç ve Pareto çözüm kümeleri Şekil 7'de verilmiştir.

**Şekil 7.** f_1 ve f_2 amaç fonksiyonları için başlangıç ve Pareto çözüm kümeleri

Çizelge 13'te amaç fonksiyonlarını eş anlı minimize eden optimal girdi değişken değerleri yer almaktadır.

Çizelge 13. f_1 ve f_2 amaç fonksiyonlarını eş anlı minimize eden optimal girdi değişken değerleri

<i>Girdi Değişkenleri</i>		<i>Amaç fonksiyonları</i>
X_1	X_2	$[f_1 \ f_2]$
-1.0000	-1.4140	[3.1206 0.1164]
-1.4140	-1.0086	[2.9791 0.0908]

Y_2 ve Y_3 yanıtları için SUR yöntemi ile elde edilen parametre tahminleri Çizelge 14'te verilmiştir.

Çizelge 14. Y_2 ve Y_3 yanıtlarının SUR yöntemi ile parametre tahminleri

<i>Model Terimi</i>	<i>SUR</i>			
	Y_2	p -değeri	Y_3	p -değeri
<i>Sabit</i>	0.66 (.007)	.001	1.776 (.016)	.001
X_1	-0.092 (.005)	.001	-0.25 (.013)	.001
X_2	-0.010 (.005)	.106	-0.078 (.013)	.001
X_1^2	-0.096 (.006)	.001	-0.156 (.014)	.001
X_2^2	-0.058 (.006)	.001	-0.078 (.014)	.001
X_1X_2	-0.070 (.007)	.001	0.01 (.018)	.602

*Standart hatalar parantez içinde belirtilmiştir.

Çizelge 14'ten, Y_2 ve Y_3 yanıtları için elde edilen tahmini yanıt modelleri sırasıyla

$$\hat{Y}_2 = 0.660 - 0.092X_1 - 0.096X_1^2 - 0.058X_2^2 - 0.070X_1X_2 \quad (29)$$

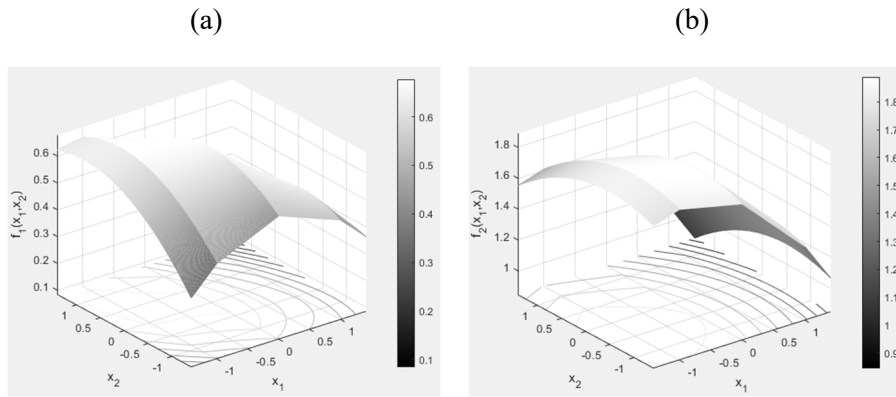
ve

$$\hat{Y}_3 = 1.776 - 0.250X_1 - 0.078X_2 - 0.156X_1^2 - 0.079X_2^2 \quad (30)$$

dır. Optimal değeri elde edilmek istenilen her bir yanıt fonksiyonu, bir amaç fonksiyonu olarak değerlendirildiğinde çok yanıtlı optimizasyon problemi

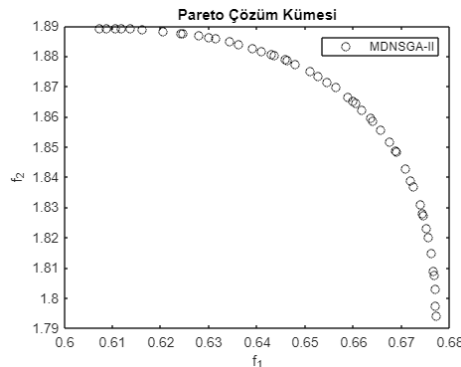
$$\begin{aligned} \max f_1 &= \hat{Y}_2(\mathbf{x}) \\ \max f_2 &= \hat{Y}_3(\mathbf{x}) \\ \mathbf{x} &\in S \end{aligned} \quad (31)$$

biçiminde ÇAO problemine dönüşecektir. Burada, $\{X_1, X_2\}$ değişkenlerinin aldığı değerler vektörü \mathbf{x} olarak belirtilmiştir. f_1 ve f_2 amaç fonksiyonları, sırasıyla \hat{Y}_2 ve \hat{Y}_3 tahmini yanıt fonksiyonlarına karşılık gelmektedir. Elde edilen tahmini yanıt fonksiyonları için yüzey grafikleri Şekil 8.(a)-(b)'de görülmektedir. Şekil 8.(a)-(b)'ye bakıldığında maksimizasyon problemleriyle ilgilenildiği görülmektedir.



Şekil 8. (a) f_1 amaç fonksiyonu (b) f_2 amaç fonksiyonu için yüzey grafikleri

ÇAO probleminin MDNSGA-II ile elde edilen Pareto çözüm kümesi Şekil 9'da verilmiştir. Gıda veri seti için f_1 ve f_2 amaç fonksiyonlarına ilişkin Pareto çözüm değerleri Ek-2'de verilmiştir. Ek-2'de görüldüğü gibi MDNSGA-II ile $\{X_1\}$ değişkeni tanım kümesinden kesikli değerler olarak Pareto çözümlerin elde edilmesi sağlanmıştır.



Şekil 9. Gıda veri seti için f_1 ve f_2 amaç fonksiyonlarına ilişkin Pareto çözüm kümesi

5. Sonuç ve öneriler

Bu çalışmada, karma veri içeren çok yanıtlı problemlerin modellenmesi ve optimizasyonu ile ilgilenilmiştir. Yanıt değişkenlerinin modellenmesinde GLM ve SUR modeller kullanılmıştır. Karma veri içeren çok yanıtlı modellerin tahmini yanıt fonksiyonları, amaç fonksiyonları olarak ele alınıp problem ÇAO problemi biçiminde değerlendirilmiştir. ÇAO için bir yapay zeka optimizasyon algoritması olan NSGA-II'ye dayalı algoritma önerilmiştir. Bu amaçla, NSGA-II'nin değişken gösterimi, başlangıç popülasyonunun oluşturulması ve genetik operatörlerin uygulanması aşamalarında indeksleme yapılarak kesikli değer alan değişkenlerle optimizasyon yapabilmek için NSGA-II modifiye edilmiştir. Önerilen algoritma, çalışma kapsamında MDNSGA-II olarak adlandırılmıştır. Çalışmada, kesikli değişkenlerin indeks değerleri dikkate alınarak karma veri içeren çok yanıtlı problemler için MDNSGA-II ile Pareto çözüm kümesinin elde edilebilir olduğu gösterilmiştir. Uygulamada kullanılan UCI Repository veri tabanından enerji verimliliği veri seti ile literatürde tanımlı gıda alanından deneysel karma veri seti için elde edilen sonuçlar Ek-1 ve Ek-2'de verilmiştir. Ek-1 ve Ek-2 incelendiğinde, uygulamada kullanılan veri setleri için MDNSGA-II ile Pareto çözüm kümesinin elde edilebildiği görülmüştür. Bununla birlikte, Enerji veri setinde $\hat{Y}_1 - \hat{Y}_2$ ve gıda veri setinde $\hat{Y}_2 - \hat{Y}_3$ tahmini yanıt fonksiyonları için Pareto çözüm kümesinde çözüm çeşitliliği sağlanırken, gıda veri setinin $\hat{Y}_1 - \hat{Y}_4$ tahmini yanıt fonksiyonları için Pareto çözüm kümesinde çözüm çeşitliliği yeterince sağlanamamıştır. Pareto çözüm kümesinde çözüm çeşitliliğini artırmak için referans noktalarına dayanan NSGA-III'ün uyarlanmasıyla bu sorunun giderilmesi sonraki çalışma planı olarak öngörülmektedir.

Ek-1 Enerji veri seti için f_1 ve f_2 amaç fonksiyonlarına ilişkin Pareto çözüm kümesi

N	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	f_1	f_2
1	0.62	610.6571	380.9550	113.8490	3.5	3	0.40	4	9.8763	83.3668
2	0.62	610.9978	408.3545	115.5357	3.5	3	0.40	4	10.1452	68.0712
3	0.62	612.4556	393.1433	115.6923	3.5	3	0.40	4	9.9712	76.7932
4	0.62	610.7042	390.4315	113.9442	3.5	3	0.40	4	9.9704	78.0666
5	0.62	610.6187	381.6393	113.8251	3.5	3	0.40	4	9.8837	82.9782
6	0.62	610.8081	386.8392	113.9304	3.5	3	0.40	4	9.9329	80.0928
7	0.62	610.6168	385.6920	113.9586	3.5	3	0.40	4	9.9243	80.7085
8	0.62	610.7253	384.3239	113.8813	3.5	3	0.40	4	9.9090	81.4898
9	0.62	612.3562	394.9505	115.7381	3.5	3	0.40	4	9.9908	75.7672
10	0.62	612.4023	401.1725	115.4784	3.5	3	0.40	4	10.0523	72.2893
11	0.62	610.8283	388.5117	113.9288	3.5	3	0.40	4	9.9493	79.1590
12	0.62	611.0447	406.3511	115.5784	3.5	3	0.40	4	10.1245	69.1996
13	0.62	611.4364	401.3444	115.4862	3.5	3	0.40	4	10.0685	72.0578
14	0.62	612.3345	397.0036	115.4902	3.5	3	0.40	4	10.0116	74.6144
15	0.62	612.2230	394.1300	115.7882	3.5	3	0.40	4	9.9846	76.2081
16	0.62	610.9330	387.8107	113.9694	3.5	3	0.40	4	9.9407	79.5663
17	0.62	612.3493	396.3216	115.5053	3.5	3	0.40	4	10.0046	74.9984
18	0.62	610.7242	385.0456	114.0231	3.5	3	0.40	4	9.9162	81.0855
19	0.62	610.6901	382.5880	113.6679	3.5	3	0.40	4	9.8921	82.4570
20	0.62	612.3872	395.5429	115.7546	3.5	3	0.40	4	9.9962	75.4398
21	0.62	611.3824	403.5865	115.3821	3.5	3	0.40	4	10.0917	70.7947
22	0.62	611.0526	405.6636	115.5859	3.5	3	0.40	4	10.1175	69.5858
23	0.62	610.7707	383.1573	113.9089	3.5	3	0.40	4	9.8966	82.1494
24	0.62	612.3377	397.4010	115.5058	3.5	3	0.40	4	10.0155	74.3923
25	0.62	611.0458	404.8979	115.5219	3.5	3	0.40	4	10.1099	70.0136
26	0.62	610.8196	387.4233	113.9322	3.5	3	0.40	4	9.9385	79.7673
27	0.62	610.7164	389.9160	113.9293	3.5	3	0.40	4	9.9650	78.3570
28	0.62	610.7170	389.5886	113.9400	3.5	3	0.40	4	9.9617	78.5404
29	0.62	611.0235	406.7547	115.6112	3.5	3	0.40	4	10.1288	68.9707
30	0.62	612.2376	395.8547	115.7415	3.5	3	0.40	4	10.0016	75.2442
31	0.62	611.1244	402.5708	115.3929	3.5	3	0.40	4	10.0855	71.3277
32	0.62	611.0496	405.4035	115.5863	3.5	3	0.40	4	10.1149	69.7310

33	0.62	612.4173	393.7951	115.7586	3.5	3	0.40	4	9.9783	76.4228
34	0.62	610.7554	385.3684	113.9496	3.5	3	0.40	4	9.9190	80.9091
35	0.62	612.1254	397.7169	115.3760	3.5	3	0.40	4	10.0219	74.1857
36	0.62	612.2661	399.6804	115.6109	3.5	3	0.40	4	10.0394	73.1058
37	0.62	611.4462	402.0085	115.3470	3.5	3	0.40	4	10.0750	71.6873
38	0.62	610.7435	389.0879	113.9321	3.5	3	0.40	4	9.9563	78.8245
39	0.62	610.7011	383.6942	113.8565	3.5	3	0.40	4	9.9030	81.8390
40	0.62	610.7108	383.7182	113.7821	3.5	3	0.40	4	9.9031	81.8269
41	0.62	611.3962	403.7842	115.3813	3.5	3	0.40	4	10.0935	70.6859
42	0.62	611.0798	403.8909	115.4860	3.5	3	0.40	4	10.0993	70.5823
43	0.62	612.4827	400.2912	115.5557	3.5	3	0.40	4	10.0423	72.7941
44	0.62	612.5187	400.9726	115.5398	3.5	3	0.40	4	10.0485	72.4176
45	0.62	612.3157	398.7933	115.3863	3.5	3	0.40	4	10.0298	73.6096
46	0.62	612.4558	398.2394	115.6069	3.5	3	0.40	4	10.0222	73.9394
47	0.62	612.5211	399.6069	115.5796	3.5	3	0.40	4	10.0349	73.1827
48	0.62	611.0621	404.0529	115.4698	3.5	3	0.40	4	10.1012	70.4891
49	0.62	611.1210	404.5703	115.4520	3.5	3	0.40	4	10.1055	70.2076
50	0.62	612.3504	400.3724	115.4825	3.5	3	0.40	4	10.0451	72.7301

Ek-2 Gıda veri seti için f_1 ve f_2 amaç fonksiyonlarına ilişkin Pareto çözüm kümesi

N	X_1	X_2	f_1	f_2
1	-1	0.6034	0.6771	1.7942
2	-1	-0.4937	0.6073	1.8893
3	-1	-0.3945	0.6194	1.8885
4	-1	-0.4679	0.6105	1.8892
5	-1	0.2400	0.6695	1.8467
6	-1	-0.3555	0.6238	1.8877
7	-1	0.2790	0.6710	1.8421
8	-1	-0.3324	0.6263	1.8872
9	-1	0.3576	0.6736	1.8320
10	-1	0.0050	0.6563	1.8696
11	-1	0.1756	0.6665	1.8539
12	-1	-0.3004	0.6297	1.8863
13	-1	0.2136	0.6683	1.8497
14	-1	-0.2504	0.6348	1.8846
15	-1	0.3787	0.6742	1.8291
16	-1	0.1448	0.6649	1.8570
17	-1	0.1933	0.6674	1.8520
18	-1	-0.0680	0.6510	1.8749
19	-1	0.0409	0.6588	1.8667
20	-1	0.0995	0.6624	1.8615
21	-1	-0.0238	0.6543	1.8718
22	-1	0.4018	0.6748	1.8259
23	-1	-0.2750	0.6324	1.8855
24	-1	0.0804	0.6613	1.8632
25	-1	0.3267	0.6727	1.8361
26	-1	-0.2304	0.6368	1.8838
27	-1	0.4253	0.6753	1.8225
28	-1	-0.0911	0.6491	1.8764
29	-1	-0.1434	0.6448	1.8796
30	-1	-0.1522	0.6440	1.8800
31	-1	0.4833	0.6763	1.8139
32	-1	-0.2185	0.6379	1.8833
33	-1	0.0643	0.6603	1.8647
34	-1	0.6022	0.6771	1.7944

35	-1	0.4615	0.6760	1.8172
36	-1	-0.1727	0.6422	1.8811
37	-1	0.5783	0.6771	1.7985
38	-1	-0.0456	0.6527	1.8734
39	-1	-0.1226	0.6465	1.8784
40	-1	0.5224	0.6767	1.8077
41	-1	-0.0535	0.6521	1.8739
42	-1	0.0270	0.6578	1.8678
43	-1	0.4473	0.6757	1.8193
44	-1	0.4751	0.6762	1.8151
45	-1	0.5642	0.6770	1.8009
46	-1	-0.1019	0.6483	1.8771
47	-1	-0.1198	0.6468	1.8782
48	-1	0.5032	0.6765	1.8107
49	-1	0.2967	0.6717	1.8399
50	-1	0.5038	0.6765	1.8106

Kaynaklar

- [1] J. Garrido, J. Zhou, 2009, Full Credibility with Generalized Linear and Mixed Models, *ASTIN Bulletin*, 39(1), 61-80.
- [2] N. P. Jewell, S. Shiboski, 1990, Statistical analysis of HIV infectivity based on partner studies, *Biometrics*, 46, 1133-1150.
- [3] A. Hern, S. Dorn, 2001, Statistical modelling of insect behavioral responses in relation to the chemical composition of test extracts, *Physiological Entomology*, 26, 381-390.
- [4] Y. J. Lee, J. A. Nelder, 2002, Analysis of ulcer data using hierarchical generalized linear models, *Statistics in Medicine*, 21, 191-202.
- [5] M. P. Diaz, A. H. Barchuk, S. Luque, C. Oviedo, 2002, Generalized linear models to study spatial distribution of tree species in Argentinean arid Chaco, *Journal of Applied Statistics*, 29, 5, 685-694.
- [6] Z. W. Yan, S. Bate, R. E. Chandler, V. Isham, H. Wheeler, 2002, An analysis of daily maximum wind speed in northwestern Europe using generalized linear models, *Journal of Climate*, 15, 2073-2088.
- [7] S. Rajeev, C. S. Krishnamoorthy, 1992, Discrete Optimization of Structures Using Genetic Algorithms, *J. Struct. Eng.*, 118(5), 1233-1250.
- [8] C. Y. Lin, P. Hajela, 1992, Genetic algorithms in optimization problems with discrete and integer design variables, *Engineering Optimization*, 19, 4, 309- 327.
- [9] W. Wang, R. Zmeureanu, H. Rivard, 2005, Applying multi-objective genetic algorithms in green building design optimization, *Building and Environment* 40, 1512–1525.
- [10] S. S. Rao, Y. Xiong, 2005, A hybrid genetic algorithm for mixed-discrete design optimization, *ASME Journal of Mechanical Design*, 127, 1100–1112.

- [11] M. Ahmadi, M. Arabi, D. L. Hoag, B. A. Engel, 2013, A Mixed Discrete-Continuous Variable Multiobjective Genetic Algorithm For Targeted Implementation of Nonpoint Source Pollution Control Practices, *Water Resources Research*, 49, 8344–8356.
- [12] B. El-Kribi, A. Houidi, Z. Affi, L. Romdhane, 2013, Application of multi-objective genetic algorithms to the mechatronic design of a four bar system with continuous and discrete variables, *Mechanism and Machine Theory*, 61, 68–83.
- [13] W. Tong, S. Chowdhury and A. Messac, 2014, A new multi-Objective Mixed-Discrete Particle Swarm Optimization Algorithm, *Proceedings of the ASME 2014 International Design Engineering Technical Conferences & Computers and Information in Engineering Conference Buffalo*, New York, USA.
- [14] T. Holzmann, J. C. Smith, 2018, Solving discrete multi-objective optimization problems using modified augmented weighted Tchebychev scalarizations, *European Journal of Operational Research*, 271, 436–449.
- [15] S. Guangyong, Z. Huile, F. Jianguang, L. Guangyao, L. Qing, 2018, A new multi-objective discrete robust optimization algorithm for engineering design, *Applied Mathematical Modelling*, 53, 602-621.
- [16] S. Roy, W. A. Crossley, S. Jain, 2021, *A Hybrid Approach for Solving Constrained Multi-Objective Mixed-Discrete Nonlinear Programming Engineering Problems*, Books, Engineering Problems - Uncertainties, Constraints and Optimization Techniques.
- [17] I. Khuri, B. Mukherjee, B. K. Sinha, M. Ghosh, 2006, Design Issues for Generalized Linear Models: A Review, *Statistical Science*, 21(3), 376-399.
- [18] D. Collins, 2008, *The performance of estimation methods for generalized linear mixed models*, Doctor of Philosophy thesis, University of Wollongong, School of Mathematics and Applied Statistics- Faculty of Informatics, 223, Australia.
- [19] J. A. Nelder, R. W. M. Wedderburn, 1972, Generalized linear models, *Journal of the Royal Statistical Society A-General*, 135, 370–384.
- [20] C. J. Anderson, J. Verkuilen, T. R. Johnson, 2012. *Applied Generalized Linear Mixed Models: Continuous and Discrete Data.*, Springer.
- [21] P. McCullagh, J. A. Nelder, 1989, *Generalized Linear Models, Second Edition*, Chapman and Hall/CRC, 511, London.
- [22] M. Friendly, D. Meyer, 2015, *Discrete Data Analysis with R: Visualization and Modeling Techniques for Categorical and Count Data*, Chapman and Hall/CRC Published.
- [23] J. J. Faraway, 2006, *Extending the Linear Model with R Generalized Linear, Mixed Effects and Nonparametric Regression Models*, Chapman & Hall/CRC Taylor & Francis Group.
- [24] E. Ostertagová, 2012, *Modelling Using Polynomial Regression*, Procedia Engineering, 48, 500-506.
- [25] Ö. Türkşen, 2023, *Optimizasyon Yöntemleri ve Matlab, Python, R Uygulamaları*, Nobel, 1.Basım, 448, Ankara.

- [26] D. E. Goldberg, 1989, *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading, ABD.
- [27] N. Srinivas, K. Deb, 1994, Mulltiobjective Optimization Using Non-Dominated Sorting in Genetic Algorithms, *Evolutionary Computation*, 2, 221-248.
- [28] K. Deb, A. Pratap, S. Agarwal, T. Meyarivan, 2002, A fast and elitist multiobjective genetic algortihm: NSGA-II, *IEEE Transactions on Evolutionary Computation*, 6, 2.
- [29] Ö. Türkşen, F. Akgün, 2018, Genetik-Simpleks hibrit algoritması ile doğrusal olmayan regresyon model parametrelerinin nokta tahmini, *İstatistikçiler Dergisi: İstatistik & Aktüerya*, 2, 81-92.
- [30] Z. Cebeci, 2021, *R ile Genetik Algoritmalar ve Optimizasyon Uygulamaları*, Nobel, 535, Ankara.
- [31] Ö. Türkşen, 2011, *Çok Yanıtlı Yüzey Problemlerinin Çözümüne Bulanık ve Sezgisel Yaklaşım*, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Doktora Tezi, İstatistik Anabilim Dalı.
- [32] B. El-Kribi, A. Houidi, Z. Affi, L. Romdhane, 2013, Application of multi-objective genetic algorithms to the mechatronic design of a four bar system with continuous and discrete variables, *Mechanism and Machine Theory*, 61, 68–83.
- [33] A. Asuncion, D. Newman, UCI Machine Learning Repository. Available online:<https://archive.ics.uci.edu/dataset/242/energy+efficiency>.
- [34] A. Tsanas, A. Xifara, 2012, Accurate quantitative estimation of energy performance of residential buildings using statistical machine learning tools, *Energy and Buildings*, 49, 560-567.
- [35] R. H. Schmidt, R. B. L. Illingworth, J. C. Deng, J. A. Cornell, 1979, Multiple Regression and Response Surface Analysis of the Effects of Calcium Chloride and Cysteine on Heat-Induced Whey Protein Gelation, *J. Agrie. Food Chem.*, 27(3), 529–532.
- [36] S. Tunçel, 2022, *Çok Yanıtlı Deneysel Verilerin Görünüşte İlişkisiz Regresyon Analizi ile Modellenmesi ve Optimal Değişken Değerlerinin Belirlenmesi*, Ankara Üniversitesi, Fen Bilimleri Enstitüsü, Yüksek Lisans Tezi, İstatistik Anabilim Dalı.