

DİPLOİD GENETİK ALGORİTMALARDA MİYOZ BÖLÜNME VE BASKINLIK

Şima ETANER-UYAR, Emre HARMANCI
İstanbul Teknik Üniversitesi Bilgisayar Mühendisliği Bölümü
TR80626 Maslak İstanbul, Türkiye
{etaner,harmanci}@cs.itu.edu.tr

Özet

Bu çalışmada, topluluğu oluşturan bireylerin diploid kromozom yapılarıyla temsil edildikleri diploid genetik algoritmalar için geliştirilen bazı yeni operatörler ele alınmıştır. Bu operatörlerin ilki eşey hücrelerinin oluşması aşamasında kullanılan miyoz bölünmedir. Diğer yeni bir operatör de bireylerin genotiplerinden fenotiplerinin belirlenmesinde kullanılan yeni bir baskınlık yöntemidir. Bu yöntemlerin kullanılması sonucunda oluşabilecek erken yakınsamanın ve yüksek başarılı bireylerin toplumu ele geçirmelerini önlemek amacıyla bir yaşlanma mekanizması eklenmiştir. Bu yeni operatörlere ek olarak klasik genetik algoritmalarda yer alan temel operatörler de kullanılmıştır. Geliştirilen algoritmanın etkinliği basit test fonksiyonları üzerinden klasik genetik algoritma ile kıyaslanarak gösterilmektedir.

Anahtar sözcükler: evrimsel hesap, diploid genetik algoritmalar, miyoz bölünme, baskınlık, erken yakınsama, global arama ve optimizasyon

1 GİRİŞ

1831 yılında Charles Darwin, doğabilimci göreviyle, HMS Beagle adlı gemide beş yıllık bir yolculuğa çıktı. Dönüşünde, 1859 yılında, gözlemlerini, deneyimlerini anlattığı ve geliştirdiği evrim teorisini belgelediği “*The Origin of Species by Means of Natural Selection*“ [1] adlı kitabını yayımladı. Teorisinin temelinde [2], tüm türlerin başka türlerden evrim geçirerek oluştukları yatmaktadır. Darwin’in çalışmaları bu evrimin gerçekten oluştuğunu belgelemekte ve evrimi sağlayan mekanizmaları açıklamaktadır. Evrim iki aşamalı bir süreçtir. İlk aşamada bireyler arasında rasgele farklılıklar oluşmaktadır. Bu farklılıkların bir kısmı kalıtım yoluyla aktarılabilir niteliktedir. Oluşan farklılıkların organizmaya faydalı veya zararlı oldukları ise o farklılığı taşıyan organizmanın hayatta kalabilme başarısı ve

üreme hızı ile belli olur. İkinci aşama ise bireyler ile çevrenin bir etkileşimi sonucu oluşan doğal seçimdir. Doğal seçim sonucunda iyi özellikler içeren bireyler daha uzun yaşar ve daha çok yavrular. Böylelikle bu bireylerin taşıdığı farklılıklar diğer kuşaklara aktarılmış olur. Uzun vadede evrim farklı ortamlarda, farklı koşullar altında yaşayan organizma grupları arasında farklılıklar oluşmasına sebep olur. Evrim kuramı, biyoloji ve ilgili diğer bilimlerin dayandığı en önemli temel ilkelere biri olarak görülmektedir.

Bir optimizasyon probleminde temel amaç en iyi çözümün veya bazı durumlarda kabul edilebilir iyilikte bir çözümün bulunmasıdır. Stokastik global optimizasyon yöntemlerinin bir alt sınıfı sayılabilen genetik algoritmalar ise problemleri, doğal evrim teorisi ve genetik biliminin bazı mekanizmalarını modelleyerek çözme yaklaşımını kullanır. Doğada iyi özellikler taşıyan ve içinde yaşadığı ortama iyi uyum sağlamış organizmalar daha büyük olasılıkla hayatta kalırlar ve yavrulayarak bu iyi özelliklerini sonraki kuşaklara aktarırlar. Bu olgu doğal bir optimizasyon süreci gibi ele alınabilir. Benzetim yoluyla, doğada yer alan bazı mekanizmalar yapay optimizasyon problemlerinin çözümünde kullanılabilir. Bu mekanizmalardan hangilerinin kullanılacağı ise çözülecek problemin özelliklerine göre değişmektedir.

2 DİPLOİD GENETİK ALGORİTMA

Doğada bulunan karmaşık organizmaların çoğu diploid kromozom yapısına sahiptirler; yani her özellik iki homolog kromozom üzerinde yer alan iki alel tarafından temsil edilmektedir. Organizmanın dıştan görünen özelliği ise bu iki alelin baskın olanı tarafından belirlenmekte, çekinik alel ise organizmanın genotipinde saklı kalmaktadır. Her özelliğin iki alel tarafından temsil edilmesi fazlalık olarak görünse de bu yöntem, doğanın organizma toplulukları içinde genetik çeşitlilik sağlamak ve bir tür genetik bellek tutmak için kullandığı yöntemdir. Baskınlık-çekiniklik mekanizmasının işlemesi sayesinde, organizmaya o sırada faydalı olmayan ama ileride, değişen ortam koşulları sebebiyle faydalı olabilecek bazı özellikler maskelenirler ve doğal seçim sonucunda kaybedilmemeleri sağlanır.

Diploid genetik algoritmalar ve baskınlık mekanizmaları ile yapılmış çalışmalar [3] numaralı referansta özetlenmiştir. Daha yeni bazı çalışmalar ise [4], [5], [6], [7] numaralı referanslarda bulunabilir. Günümüzde dinamik ortamlarda optimizasyon problemlerinin genetik algoritmalarla çözümü [8] yoğun ilgi görmektedir. Genetik algoritmalarda bu tip problemlerin çözümünde kullanılan yaklaşımlardan biri de diploid kromozom yapılarının kullanılmasıdır.

2.1 Algoritmanın Tanıtımı

Bu çalışmada birey, genotipini oluşturan bir kromozom çifti, fenotipini gösteren bir yapı, bir yaş bilgisi ve bir başarımlar bilgisi ile temsil edilmektedir. Burada kromozom çifti ve fenotip birer tamsayı dizisi ile gerçekleştirilmiştir. Kromozomlar üzerindeki genler ve fenotip değerleri 0 veya 1 değerlerini alabilirler. Yaş alanı o bireyin kaç kuşaktır hayatta kaldığını, başarımlar alanı ise optimizasyon kriterine göre bireyin ne kadar başarılı olduğunu göstermektedir. Geliştirilen algoritmada [3] numaralı referansta anlatılan basit genetik algoritmaya ilişkin temel mekanizmaların yanısıra bazı yeni mekanizmalar ve operatörler de kullanılmıştır. Geliştirilen diploid algoritmanın akışı Algoritma 1'de verilmiştir.

Algoritma 1 Diploid genetik algoritma

```
Başlat();  
do  
    ÜremeHavuzuSeç();  
    EşeyleriOluştur();  
    Çiftleştir();  
    MutasyonUygula();  
    for her_ölen_birey  
        YeniBireyYarat();  
    YeniKuşağıSeç();  
    YeniBaskınlıkDizisiHesapla();  
until kuşaklar_sonu;  
Bitir();
```

2.1.1 İlk Değer Atama

Algoritmanın Baslat() aşamasında bireyin kromozomlarına ve baskınlık değerlerinin yer aldığı global diziyeye ilk değerler yüklenir. Kromozomlar üzerindeki genlerin ilk değerleri rasgele olarak belirlenir. İleride işlevi açıklanacak olan baskınlık dizisinin tüm elemanlarına da 0.5 atanır.

2.1.2 Üremeye Girecek Çiftlerin Belirlenmesi

Bu çalışmada bireylerin cinsiyetleri yoktur, yani üreme bir seçim yöntemiyle belirlenen birey çiftleri üzerinde gerçekleştirilmektedir. Ana döngü başında üremeye girecek bireylerin yer aldığı üreme havuzu oluşturulmaktadır. Bunun için bireyler

arasından her bireyin başarımıyla orantılı olarak bir seçim yapılır. Bu seçim yöntemi temel genetik algoritmada da [3] kullanılan rulet çarkı seçim yöntemidir. Havuzda yer alan bireyler ikişer ikişer rasgele olarak eşleştirilir.

2.1.3 Yapay Miyoz Bölünme ve Eşey Hücrelerinin Oluşturulması

Yukarıda akışı verilen algoritmada seçilen bireylerin ikişer ikişer eşleşmesinden sonra eşey hücrelerinin oluşturulması amacıyla, herbir eş yapay miyoz bölünme adımına girer. Bu bölünmenin ilk aşaması kromozomların birer kopyasının çıkarılmasıdır. İlk aşama sonunda oluşmuş dört kromozomun her biri kromatid adını alır. Birbirinin kopyası olmayan kromatidler arasında çaprazlaşma gerçekleşebilir. Bu çalışmada iki-noktalı-çaprazlaşma kullanılmıştır. Önceden belirlenen bir olasılığa göre eğer belirli bir kromatid çifti arasında çaprazlaşma oluşacaksa, çaprazlaşmanın olacağı konumu belirleyen iki nokta seçilir ve bu noktalar arasında kalan segmanlar kromatidler arasında yer değiştirilir. Her kromatid en fazla bir kere çaprazlaşmaya girebilir.

İkinci aşamada dört kromatid birbirinden ayrılarak herbiri birer eşey hücresi durumuna gelir. Her eşleşmeden iki yavru oluşacağından bu dört eşey hücresinden rasgele iki tanesi seçilerek biri bir yavruya diğeri de diğeri yavruya olacak şekilde aktarılır. Her iki bireyden birer kromozom alan yavrular diploid yapılarını tamamlamış olurlar.

Üreme süreci bittiğinde topluluktaki birey sayısı, eski ve yeni bireylerle birlikte başlangıçtaki iki katına çıkmış olur.

2.1.4 Mutasyon Aşaması

Mutasyon, topluluğu oluşturan tüm bireylerin genotiplerinde yer alan genler üzerinde, önceden belirlenen bir olasılıkla gerçekleştirilir. Mutasyon sonucunda bireyin fenotipinde bir değişiklik olabilir.

2.1.5 Yaşlanma ve Ölüm

Erken yakınsamaya ve yüksek başarılı bireylerin toplumu ele geçirmelerine karşı bireylerin yaşlanması kavramı getirilmiştir. Bireyler, yaşlarıyla orantılandırılmış bir olasılıkla ölürler ve yerlerine yeni bireyler rasgele olarak yaratılır. Bir bireyin ölme olasılığı $k(yas)^2$ bağıntısıyla hesaplanır. Burada k problemin gerektirdiği şekilde belirlenen ve $[0, 1]$ aralığında reel değerler alabilen bir katsayıdır ve yas ise bireyin yaşı bilgisidir.

2.1.6 Yeni Kuşakın Oluşturulması

Topluluktaki birey sayısının sabit olması istendiğinden, üreme sonucunda sayıları iki katına çıkan bireylerden bir sonraki kuşakta yer alacakların belirlenmesi gerekmektedir. Bu da her bireyin başarımı ile orantılı bir yaşama şansına sahip olduğu, rulet çarkı benzeri bir yöntemle yapılır. Ancak rulet çarkındakinden farklı olarak bu yöntemde her birey en fazla bir kere seçilebilir. Seçilen birey içinde bulunduğu topluluktan silinir ve bir sonraki kuşağa taşınır. Bu seçim, istenen topluluk boyuna ulaşılan kadar tekrarlanır. Seçilen her bireyin yaş sayacı artırılır .

2.1.7 Fenotip Belirlenmesi ve Baskınlık Dizisi Hesaplanması

Bireyin başarımlı değeri fenotipi kullanılarak hesaplandığından her aşama sonunda bireylerin genotiplerinden fenotiplerinin belirlenmesi için bir dönüşüm yöntemi gerekmektedir. Bu dönüşüm baskınlık mekanizması aracılığıyla sağlanmaktadır. Baskınlık mekanizmasının belirlenmesi, diploid genetik algoritmalarda önemli araştırma konularından biridir ve bu konuda yapılan çalışmalar [3] numaralı referansta detaylı olarak açıklanmıştır. Bu çalışmada kullanılan yaklaşım, her kuşaktaki topluluk bireylerinin başarımlı değerlerinin istatistiksel incelenmesine dayanmaktadır.

Fenotip belirlenmesi aşamasında iki durum ortaya çıkar. O fenotip konumuna karşılık düşen her iki gen de ya aynıdır ya da farklıdır. Aynı olmaları durumunda fenotipin ilgili konumunda da o alel değeri görülür. Farklı olmaları durumunda ise ilgili fenotip değerinin belirlenmesi için bir yöntem geliştirilmesi gerekmektedir. Doğada fenotipin belirlenmesinde alellerin baskın veya çekinik olmaları kullanılır. Bu çalışmada kromozomlar ile aynı boyda olan ve her bir elemanı $[0, 1]$ aralığında reel değerler alabilen bir baskınlık dizisi kullanılmıştır. Bu dizideki i . konumdaki değer, kromozomların üzerinde aynı konumdaki alel değerlerinden 1 olanın 0 olanına göre baskınlık derecesini göstermektedir.

Baskınlık dizisi de bireylerin geçirdiği değişimlere bağlı olarak bir değişim geçirir. Her kuşakta yer alan topluluğu oluşturan bireyler incelenerek, her kuşak için yeni bir baskınlık dizisi (1) numaralı eşitlik kullanılarak belirlenir.

$$Bas_i = \frac{\sum_j (f_{ij} * b_j)}{\sum b_j}, \quad i = 1, 2, \dots, boy \quad j = 1, 2, \dots, nufus \quad (1)$$

Burada Bas_i baskınlık dizisinin i . konumunu, f_{ij} topluluktaki j . bireyin fenotipinin i . konumundaki değerini, b_j topluluktaki j . bireyin başarımlı değerini, boy kromozom boyunu ve $nufus$ da topluluktaki birey sayısını göstermektedir. Bu

Parametreler	Standart GA	Diploid GA
Kuşak Sayısı	1000	1000
Nüfus	250	250
Çaprazlaşma Olasılığı	0.9	0.9
Mutasyon Olasılığı	0.009	0.009
Yaşlanma ve Ölme Faktörü (k)	-	0.001

Tablo 1: Parametreler

eşitlik kullanılarak her kuşak sonunda baskınlık dizisi değerleri, o topluluğu oluşturan bireyler için belirlenir. Dolayısıyla o kuşaktaki toplulukta yer alan bireylerin fenotiplerinin i . konumlarında 1 değeri olanları daha yüksek başarımlarına sahipse Bas_i değerinin de daha yüksek olması beklenir. Yani optimal çözümde fenotipte 1 olan konumlar için baskınlık dizisi değerinin de 1.0'a ve 0 olduğu konumlar için de 0.0'a yakınsaması beklenir.

2.1.8 Algoritmanın Sonlanması

İstenen sayıda kuşak tamamlandığında algoritma sonlanır. Tüm kuşaklar boyunca tutulan, o kuşak içinde en yüksek başarımlarına sahip bireylerin en iyisi, aranan optimal çözüm olarak belirlenir.

3 TEST FONKSİYONLARI VE SONUÇLAR

Çalışmada sunulan diploid genetik algoritma ile standart genetik algoritma [3] iki fonksiyon üzerinden kıyaslanmıştır. Kıyaslamada kullanılan test fonksiyonları ve elde edilen sonuçlar aşağıda verilmiştir. Her iki testte de algoritmalar yüz kere çalıştırılmış ve sonuçlar bu yüz çalışmanın ortalaması olarak belirlenmiştir. Her seferinde algoritmalar aynı parametreler ve farklı başlangıç toplulukları ile çalıştırılmıştır. Seçilen parametreler Tablo 1'de görülmektedir.

Sonuçlar iki aşamalı olarak karşılaştırılacaktır. İlk aşamada algoritmaların çevrimdışı başarımları incelenecektir. DeJong tarafından tanımlanan çevrim-içi ve çevrimdışı başarımlar kavramları [3] numaralı kaynakta açıklanmıştır. İkinci aşamada ise sonuç istatistiklerine ilişkin bir tablo verilecektir. Bu tabloda *Ortalama Başarımlar* yüz çalışmanın her birinde elde edilen en iyi başarımların ortalamasını, σ yüz çalışmada elde edilen en iyi başarımların standart sapmasını, *En İyi Başarımlar* ve *En Kötü Başarımlar* sırasıyla yüz çalışma boyunca elde edilen en iyi ve en kötü başarımların ortalamalarını, *Oran* standart sapmanın ortalamaya oranını ve

Test 1	Standart GA	Diploid GA
Ortalama Başarım	32	32
σ	0.0	0.0
Oran	0.0	0.0
En İyi Başarım	32	32
En Kötü Başarım	32	32
Ortalama Adım	29	30

Tablo 2: Birinci Test Sonuçları

Ortalama Adım da yüz çalışma boyunca en iyi başarımların elde edildiği kuşak sayısının ortalamasını göstermektedir.

3.1 Birinci Test

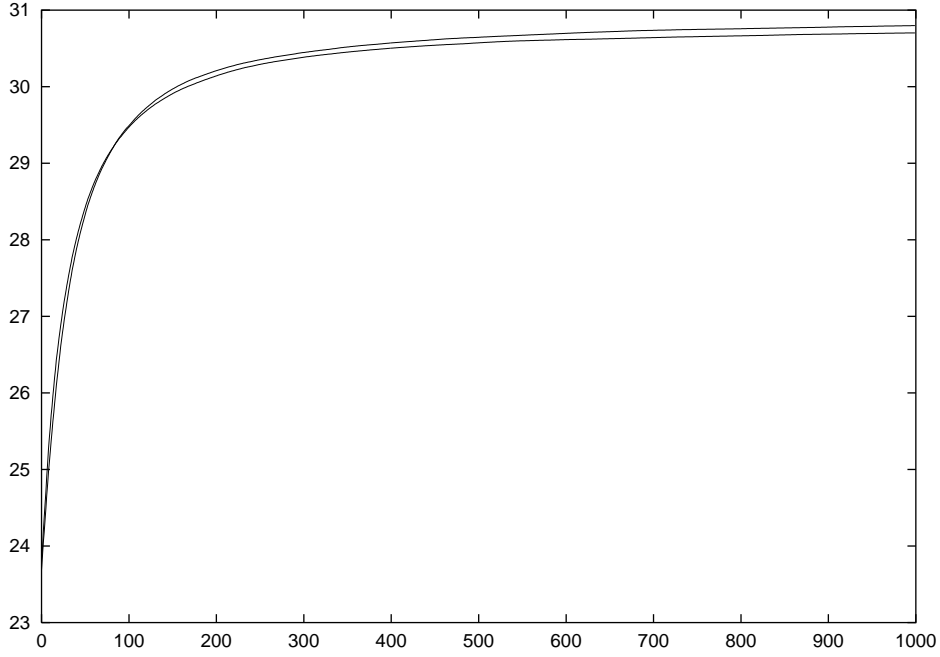
Bu testte kullanılan fonksiyon standart genetik algoritma için kolay bir fonksiyondur. Bu testin amacı, bu tip bir problemde de önerilen algoritmanın en az standart algoritma kadar bir başarımlar elde ettiğini göstermektir.

3.1.1 Problemin Tanımı

Kromozom boyunun 32 olduğu durumda 1 alel değerine sahip genlerin sayısının maksimize edilmesi istenmektedir. Bu durumda en iyi başarımlarına sahip bireyin fenotipinde tüm gen konumlarında 1 değeri, en kötü başarımlarına sahip olanında da 0 değeri olacaktır. Bireyin başarımlarını 1'lerin sayısının toplamı şeklinde hesaplanmaktadır.

3.1.2 Sonuçlar

Şekil 1'de görüldüğü gibi, algoritmaların çevrim-dışı başarımlar grafikleri aynı eksen takımlarına çizilmiştir. Burada x-ekseni kuşak sayısını, y-ekseni de başarımlar değerini göstermektedir. Şekilde üstteki, yani daha iyi olan grafik diploid algoritmaya ait olanıdır. Grafikler ve Tablo 2'de verilen sonuçlardan görülüyor ki diploid algoritma bu tip bir problemde standart algoritma kadar iyi bir başarımlar göstermektedir.



Şekil 1: 100 Çalışma Üzerinden Ortalaması Verilen Çevrim-Dışı Başarım

3.2 İkinci Test

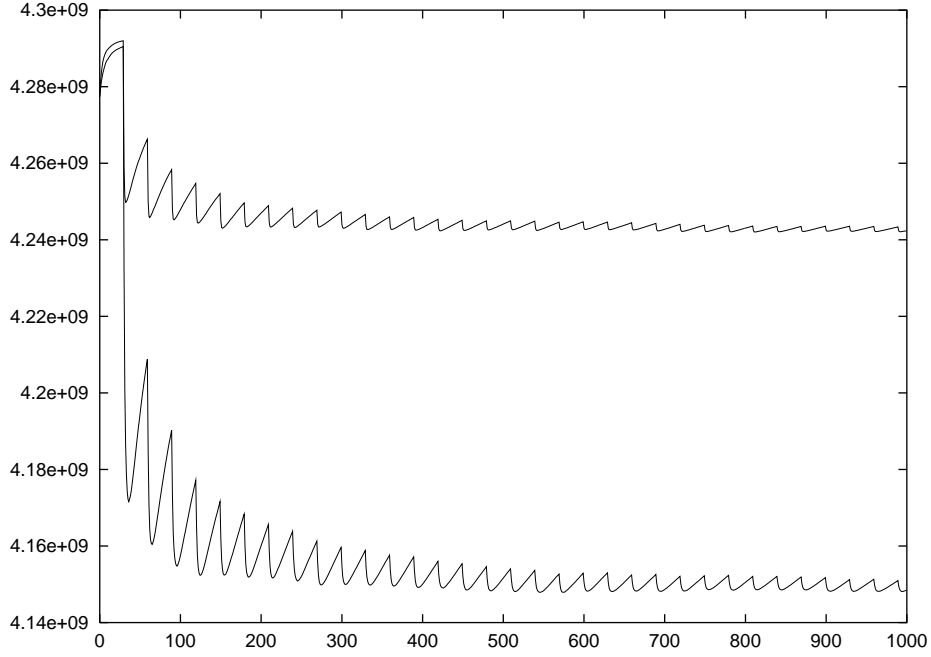
Bu testte kullanılan fonksiyon standart genetik algoritma için zor kabul edilen problemler sınıfındandır. Bu çalışmada önerilen algoritmanın bu test fonksiyonu için standart genetik algortmadan daha iyi sonuçlar verdiği gösterilecektir.

3.2.1 Problemin Tanımı

Bu problemde de kromozom boyu 32 olarak alınmıştır. Başarım hesaplamada kullanılan yöntem her 30 adımda bir, iki fonksiyon arasında değişmektedir. Birinci fonksiyon kromozomda yer alan 1'lerin sayısını maksimize etmeye çalışırken, ikinci fonksiyon da 0'ların sayısını maksimize etmeye çalışmaktadır.

3.2.2 Sonuçlar

Şekil 2'de görüldüğü gibi, algoritmaların çevrim-dışı başarım grafikleri aynı eksen takımlarına çizilmiştir. Burada x-ekseni kuşak sayısını, y-ekseni de başarım değerini göstermektedir. Şekilde üstteki, yani daha iyi olan grafik diploid algoritmaya ait olanıdır. Grafikler ve Tablo 3'de verilen sonuçlardan görülüyor ki bu test



Şekil 2: 100 Çalışma Üzerinden Ortalaması Verilen Çevrim-Dışı Başarım

problemde diploid algoritma standart algoritmadan daha iyi bir başarımlı göstermektedir. Ayrıca Tablo 3'deki sonuçlar incelendiğinde bu test problemi için diploid algoritmanın daha güvenilir olduğu görülür. Bu tabloda yer alan *Oran* satırına bakıldığında diploid algoritmanın sadece bir çalışma sonunda bile iyi ve güvenilir sonuçlar vereceği söylenebilir. Bu durum önemli başka bir özelliği de göstermektedir. Farklı çalışmalarda bulunan en iyi değerler arasındaki standart sapmanın yüksek olması, bu problem için, standart genetik algoritmanın diploid algoritmaya göre, başlangıç toplumunda yer alan bireylere daha bağımlı olduğu sonucunu ortaya çıkarmaktadır.

Test 2	Standart GA	Diploid GA
Ortalama Başarım	4294950144	4294966272
σ	34972	2636
Oran	0.000814	0.000061
En İyi Başarım	4294967296	4294967296
En Kötü Başarım	4294702080	4294958080
Ortalama Adım	147	260

Tablo 3: İkinci Test Sonuçları

4 SONUÇ

Bu bildiride, önerilen diploid algoritmanın tanıtımı yapılmış ve belirlenen iki test fonksiyonu aracılığıyla sonuçların ve algoritmanın başarımının standart genetik algoritma ile karşılaştırması yapılmıştır. Başka tip test problemleri kullanılarak yeni karşılaştırmalar üzerinde halen çalışılmaktadır.

Kaynaklar

- [1] DARWIN, Charles. *The Origin of Species*, Penguin Books Ltd. 1985
- [2] CURTIS, Helena. BARNES N. *Invitation to Biology 3rd ed.*, Worth Publishers Inc. 1981.
- [3] GOLDBERG, David E. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison Wesley. 1989.
- [4] RYAN, Conor. “The Degree of Oneness”, in *Proceedings of the 1994 ECAI Workshop on Genetic Algorithms*. Springer Verlag. 1994.
- [5] Collingwood, Emma. Corne, David. Ross, Peter. “Useful Diversity via Multiplicity”, *AISB Workshop on Evolutionary Computation*. 1996.
- [6] Greene, F.. “A Method for Utilizing Diploid/Dominance in Genetic Search”, in *Proceedings of the First IEEE Conference on Evolutionary Computation*. 1994.
- [7] Kim, Young-il. Kim, JongKyou. Lee, Seung-Soo. Cho, Choong-Ho. Lee-Kwang, Hyung. “Winner Take All Strategy for a Diploid Genetic Algorithm”, in *Proceedings of the First Asia-Pacific Conference on Simulated Evolution and Learning*. 1996.
- [8] Branke, J.. *Evolutionary Algorithms for Dynamic Optimization Problems - A Survey*, Forschungsbericht 387, Institut fuer Angewandte Informatik und Formale Beshreibungsverfahren, Universitaet Karlsruhe. Februar 1999.