

# Veri Madenciliği Uygulamaları

---

## Hafta 13

Yrd. Doç.Dr. Nilüfer YURTAY

## Kümeleme-Grid Temelli Algoritmalar

### 13. 1 Giriş

Clustering (kümeleme) özetle; elimizdeki dataset’den yola çıkarak datayı kümelemeyi, demetlemeyi amaçlamak olarak düşünülebilir. Bazı ingilizce kavramların türkçeye çevrilmesinde yanlışlıklar ortaya çıkabilmektedir. Burada da Clustering’in karşılığı olarak Sınıflandırma terimini düşünmek yanlış olacaktır. Sınıflandırma ile kümeleme türkçede yakın anlamları ifade etse de, İş Zekası dünyasında esasen farklı anlamlara ve karşılıklara sahiptirler<sup>1</sup>.

Kümelemeye dayalı modellerin, Sınıflandırmaya dayalı modellerden farkı eldeki mevcut verilerin daha önceden belirli olan bir sınıflandırmaya göre değil de belirli olmayan bir sınıfa göre gruplandırmasıdır. Yani kümeleme analizine dayalı bir modelde daha önceden belirlememiş kıstaslara göre verilere gruplara ayrılıp kümelenebilmektedir.

Grid temelli kümeleme yaklaşımı çok çözümlü grid veri yapısını kullanır. Kümeleme yapılacak alanın sonlu sayıda hücrelere bölünmesiyle oluşur. Ana avantajı genelde birbirinden bağımsız sayıda veri nesnelerinde hızlı işlem zamanıdır.<sup>2</sup>

Grid temelli yaklaşımın bazı genel örnekleri: STING, grid hücrelerindeki istatistiksel bilgiyi araştırır; Wavecluster,wavelet dönüşüm metodunu kullanan nesneleri kümeler; CLIQUE, yüksek boyutlu veri alanlarını kümelemek için grid ve yoğunluk temelli yaklaşımı temsi eder.

### 13.2 STING (Statistical Information Grid- İstatistiksel bilgi Grid)

STING uzayı dikdörtgensel hücrelere bölen bir tekniktir. Bu hücreler hiyerarşik yapıdadır. Üst seviyedeki bütün hücreler bir sonraki alt seviyede parçalanmış hücrelerden oluşur. Her bir grid hücresindeki niteliklerle ilişkili istatistiksel bilgi (örneğin mean, maksimum veya minumum değerler) önişlenir veya tutulur.

Bölgeyi dikdörtgen hücrelere bölmek için hiyerarşik bir tekniktir. Grid veri yapısı, her hücre hakkında özet bilgi içerir. Grid-tabanlı istatistiksel bilgi ve hiyerarşik kümeleme yapısı. Quad treeye benzer.

STING tekniğinin özellikleri aşağıdaki gibi listelenebilir:

- Uzaysal alanı dikdörtgen hücrelere(cells) böl.
- Farklı hücre seviyeleri farklı kararlara karşılık gelir ve bu hücreler hiyerarşik bir yapıya sahiptir.
- Daha yüksek seviyedeki her hücre, sonrasında daha düşük seviyedeki hücrelere bölünür.
- Her hücrenin istatistiksel bilgisini hesapla ve depola. Sorguları cevaplamak için kullan.

<sup>1</sup> <http://www.gurunlu.com/?tag=/k-means>

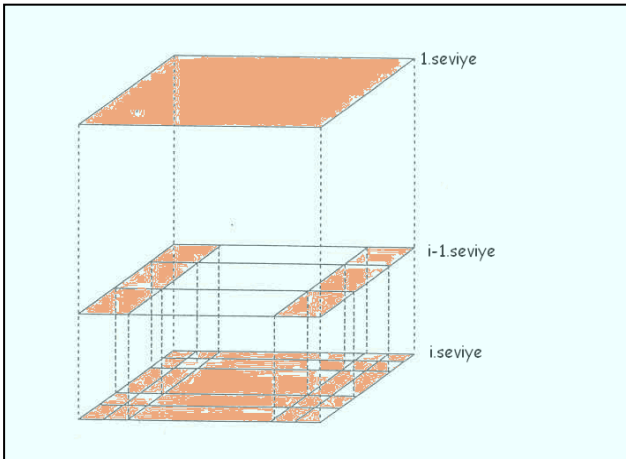
<sup>2</sup> Ayhan Adsız, Metin Madenciliği, Ahmet Yesevi Üniversitesi, Bilişim Sistemleri ve Mühendislik Fakültesi,Dönem Projesi,sayfa:46,2006.

Üst seviyede hücre istatistiksel parametreleri, alt seviyedeki hücre istatistiksel parametrelerden kolayca hesaplanabilir. Bu parametreler şu şekildedir: bağımsız parametre, sayı (count); nitelik-bağımlı parametreler, m (mean), s (standart sapma), min (minimum), max (maksimum) ve hücrelerin nitel dağılımlarındaki dağılım tipi: normal, tek biçimli, üstel, veya hiçbirisi(dağılım bilinmiyor) gibi. Veri , veritabanına kaydedilirken, en alt seviyede hücrelerdeki count, m, s, min ve max parametreleri direkt olarak hesaplanır. Dağılım değeri eğer dağılım tipi biliniyorsa kullanıcı tarafından önceden hesaplanabilir veya  $\chi^2$  testi gibi hipotez testleri tanımlanabilir. Üst seviyedeki dağılım tipi , aynı alt seviyedeki hücrelerin birleştirilerek eşik filtreleme işleminden geçirilerek bulunabilir. Eğer alt seviye hücrelerdeki dağılım birbiriyle uyuşmuyorsa, eşik testi boşa gider ve üst seviyedeki dağılım tipi hiçbirisi olur.

İlk olarak, sorgu-cevap işleminin başlayacağı hiyerarşik yapıdaki seviye belirlenir. Bu katman genelde küçük sayıda hücre içerir. Katmandaki bütün hücreler için, verilen sorguya ilgisine göre güven aralığı hesaplanır. İlgisiz hücreler ilerki adımlar için silinir. Bu işlem en alt seviyeye ulaşılan kadar devam eder. Sorgu şartı sağlanırsa hücrelerdeki ilgili bölgeler döndürülür. İlgili veri, sorgunun gereklerini yerine getirene kadar yeniden düzeltilir ve işlenir.

STING algoritmasının bazı avantajları şu şekilde belirlenmiştir:

- (1) Grid tabanlı hesaplama sorgu bağımsızdır, tüm hücrelerdeki istatistiksel bilgi grid hücresindeki özet bilgileri içerir, sorguya bağlı değildir;
- (2) Grid yapısı, paralel işleme ve güncelleştirmelere uygundur.
- (3) Metodun verimi asıl avantajıdır: STING, hücrelerdeki istatistiksel parametreleri hesaplamak için veritabanına bir kere gider, kümeleri oluşturma zaman karmaşıklığı  $O(n)$ 'dir, n nesnelerin toplam sayısıdır. Hiyerarşik yapıyı oluşturduktan sonra, sorgu işleme zamanı  $O(g)$ , g en alt seviyedeki hücre sayısıdır. (genellikle  $n > g$  olur).



**Şekil 13.1 STING Algoritmasının Hiyerarşik Yapısı**

STING'in kalitesi grid yapısındaki en alt seviye taneciğe bağlıdır. Tanecikler hassas ise işlem maliyeti artar, bununla beraber, en alt seviyedeki grid yapısının kalınlığı kümeleme analiz kalitesini azaltabilir. STING, ana hücrenin çocuk ve komşuları ile ilişkilerini göz önünde bulundurmaz. Kümeleme sınırları dikey veya yataydır, diagonal sınır yoktur. Bu kalite ve doğruluğu düşürür.

### 13.3 WaveCluster: Wavelet Dönüşümü Kullanarak Kümeleme

Wavecluster, çoklu çözüm kümeleme algoritmasıdır. İlk olarak veri uzayını çok boyutlu grid yapısına dönüştürür. Sonra wavelet dönüşümü kullanarak yoğun bölgeleri bularak orjinal uzayda dönüşüm yapar.

Wavelet dönüşüm, işareti alt frekans bandlarına ayırtıran bir işaret işleme tekniğidir. Wavelet model, n defa dönüşüm yaparak bir boyutlu sinyali n boyutlu işaretlere dönüştürebilir. Farklı çözüm seviyelerindeki göreceli mesafe verini wavelet dönüşümü sayesinde saklanır. Bu doğal kümeleri daha çok ayırt edilebilir hale getirir. İlgili alanındaki yoğun bölgeler arayarak kümeler tanımlanabilir.

Denetlemeden kümeleme sağlar. Nokta kümelerin olduğu bölgeleri vurgulayarak şapka şeklinde filtreler kullanır. Aynı zamanda zayıf bilgileri küme sınırları dışına atar. Bu şu anlama gelir; veri kümelerinde otomatik olarak belirler ve bölgelerde temizlik yapar. Wavelet dönüşüm sınır dışındaki verileri otomatik olarak temizler.

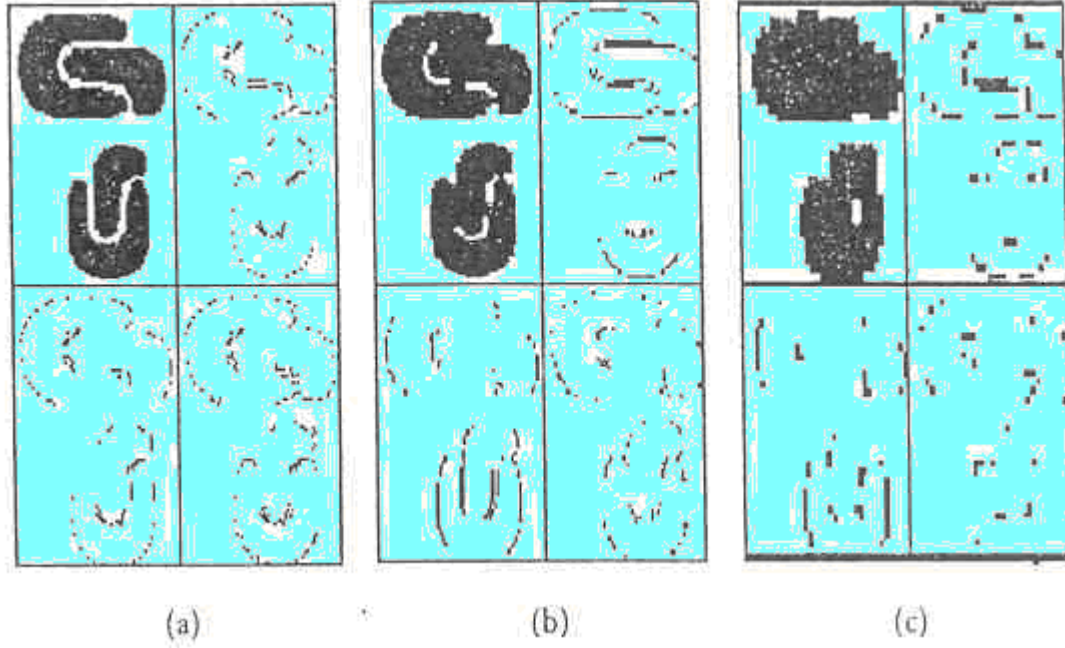
Wavelet dönüşüm kümelerdeki farklı seviyedeki doğruluğu keşfeder. Şekil 13.2'de iki boyutlu bir uzay ve şekildeki bütün noktalar nesnedeki nitelik özellikleri tanımlar. Şekil 13.3'de ise farklı çözünürlükte wavelet dönüşüm sonuçları gösterilmektedir. Tüm seviyeler alt 4 banda ayrılmış durumdadır. Sol üst band her veri noktası üzerindeki ortalama komşuluğu; sağ üst band yatay kenarları, sol alt band dikey kenarları, sağ alt band köşeleri vurgular.



Şekil 13.2 Verilerin orjinal hali<sup>3</sup>

---

<sup>3</sup>Wavecluster: A multi-resolution clustering approach for very large spatial databases, Sheikholeslami, Gholamhosein and Chatterjee, Surojit and Zhang, Aidong (1998) Wavecluster: A multi-resolution clustering approach for very large spatial databases. In Proceedings of the 24th VLDB conference .



**Şekil 13.3 Farklı çözünürlükte wavelet dönüşüm sonuçları**

Wavelet tabanlı kümeleme çok hızlıdır, hesaplama karmaşıklığı  $O(n)$ ,  $n$  veritabanındaki nesne sayısıdır. Algoritmanın paralel implemantasyonu mümkündür.

WaveCluster grid tabanlı ve yoğunluk tabanlı algoritmadır. İyi kümeleme gereksinimlerinin çoğunu karşılar. Büyük veri kümelerini verimli ele alabilir, keyfi şekilli kümeleri keşfedebilir, sınırları başarıyla tutabilir, giriş sırasına duyarlı, giriş parametrelerinin belirlenmesine gerek duymaz. Deneme çalışmalarından, wavecluster'ın BIRCH, CLARANS ve DBSCAN'dan verim ve kümeleme kalitesi olarak daha yüksek performans verdiği bulunmuştur. Ayrıca bu çalışmada wavecluster'ın 20 boyutta veri tutma kapasitesi olduğu görülmüştür.

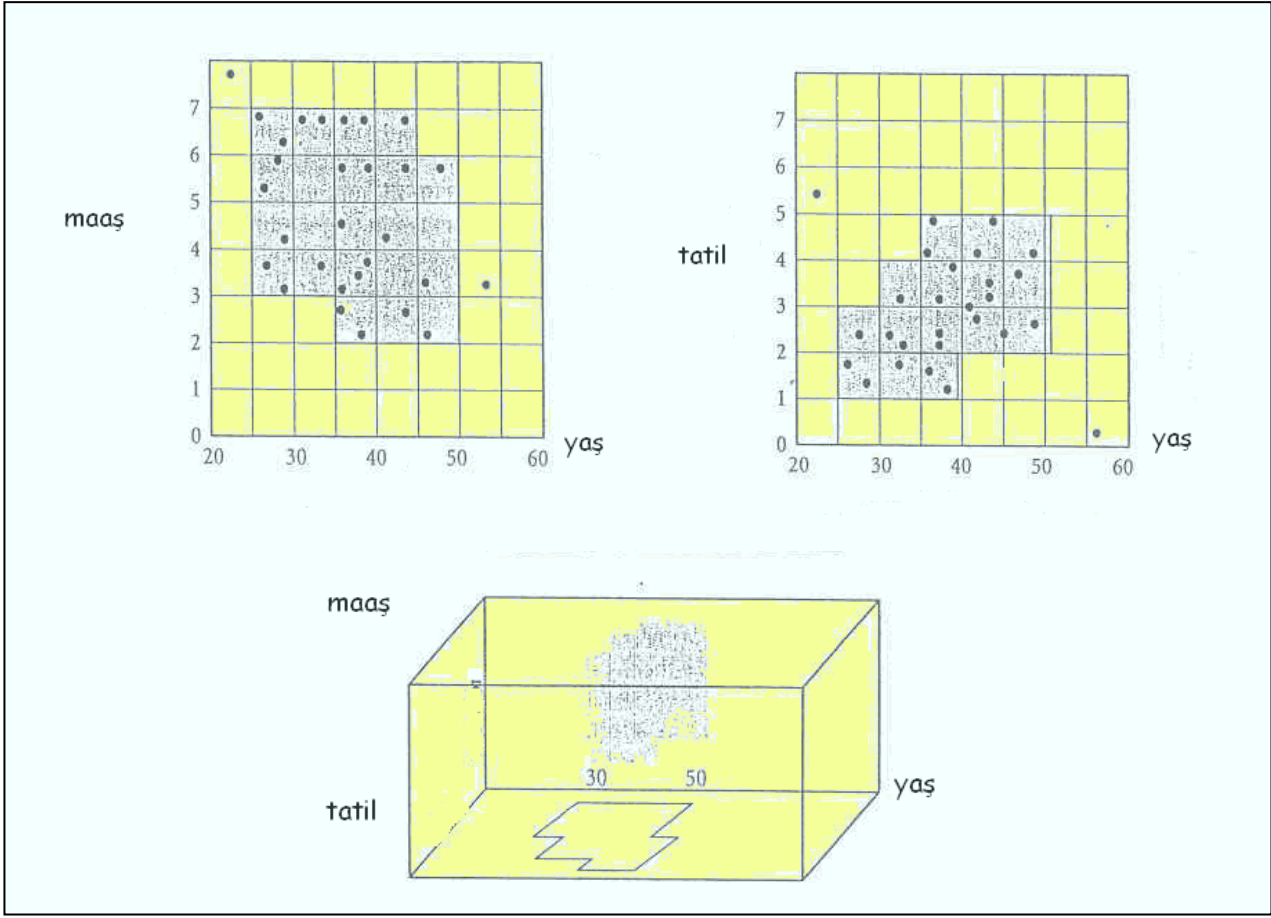
### 13.4 CLIQUE (Clustering High-Dimensional Space- Yüksek Boyutlu Alanda Kümeleme)

CLIQUE, yoğunluk tabanlı ve grid tabanlı kümelerin birleşmesinden oluşan bir algoritmadır. Büyük veritabanlarında yüksek boyutlu veri kümelemek için yararlıdır.

- Verilen büyük küme çok boyutlu veri noktalarında, veri alanı genelde tek biçimli olmaz. CLIQUE kümeleme, dağınık örüntülü veri kümelerinde seyrek ve kalabalık bölge alanlarını tanımlar.
- Toplam veri noktalarının parçası giriş model parametresini aşarsa bölüm yoğundur. CLIQUE'da , küme bağlantılı yoğun bölümün maksimal bir kümesi olarak tanımlanır.

CLIQUE iki adımda çalışır.

İlk adımda CLIQUE, üst üste binmeyecek şekilde  $n$ -boyutlu dikdörtgen şeklinde parçalara ayırır. Örnek olarak Şekil 13.4'de yoğun dikdörtgen bölümleri, yaşa göre maaş ve tatil(hafta) boyutları olarak gösterilmiştir. Alt alanlar, yoğun bölümlerin kesişmesi durumunda bizim arama yapmak için aday olması anlamına gelir.



**Şekil 13.4** CLIQUE Algoritmasının çalışması

CLIQUE için uyarlanan özellik, şu şekilde açıklanır: Eğer k-boyutlu bölüm yoğun ise k-1 boyutlu bölme bakılır. k-1 boyutlu bölümde yoğunluk yoksa k-boyutlu bölümde de yoğunluk olmaz. Bu nedenle, k-1 boyutlu alanların yoğun bölümlerinden bulunan , k boyutlu alanların yoğun bölümleri içinde potansiyel veya aday yoğunlukları üretebiliriz. Genelde sonuçlanan alan, orjinal alandan daha küçük olacaktır.

İkinci adımda, CLIQUE her küme için minimal bir tanımlama yapar. Her küme için maksimum bölge tanımlanır. Maksimum bölge bağlantılı yoğun bölümlerin kümesini kapsar. Sonra her küme için minimal kapsamı tanımlar.

CLIQUE otomatik olarak en yüksek boyutlu alt alanları bulduğu gibi her alt alandaki yüksek yoğunlukta kümeleri de bulur. Girilen satırların sırasına duyarlıdır ve herhangi geleneksel veri dağılımını tahmin edemez. Giriş boyutuna göre lineer ölçeklenir, verideki boyut sayısı arttıkça iyi ölçeklenme özelliğine sahiptir. Basitlik masrafına göre kümeleme doğruluğu düşebilir.

### 13.5 Genetik Algoritmalar

**Genetik algoritmalar**, doğada gözlemlenen evrimsel sürece benzer bir şekilde çalışan arama ve eniyileme yöntemidir. Karmaşık çok boyutlu arama uzayında en iyinin hayatta kalması ilkesine göre

bütünsel en iyi çözümü arar<sup>4</sup>. Genetik algoritmaların temel ilkeleri ilk kez Michigan Üniversitesi'nde John Holland tarafından ortaya atılmıştır. Holland 1975 yılında yaptığı çalışmaları “Adaptation in Natural and Artificial Systems” adlı kitabında bir araya getirmiştir. İlk olarak Holland evrim yasalarını genetik algoritmalar içinde eniyileme problemleri için kullanmıştır.

Genetik algoritmalar problemlere tek bir çözüm üretmek yerine farklı çözümlerden oluşan bir çözüm kümesi üretir. Böylelikle, arama uzayında aynı anda birçok nokta değerlendirilmekte ve sonuçta bütünsel çözüme ulaşma olasılığı yükselmektedir. Çözüm kümesindeki çözümler birbirinden tamamen bağımsızdır. Her biri çok boyutlu uzay üzerinde bir vektördür. Genetik algoritmalar problemlerin çözümü için evrimsel süreci bilgisayar ortamında taklit ederler. Diğer eniyileme yöntemlerinde olduğu gibi çözüm için tek bir yapının geliştirilmesi yerine, böyle yapılardan meydana gelen bir küme oluştururlar. Problem için olası pekçok çözümü temsil eden bu küme genetik algoritma terminolojisinde nüfus adını alır. Nüfuslar vektör, kromozom veya birey adı verilen sayı dizilerinden oluşur. Birey içindeki her bir elemana gen adı verilir. Nüfustaki bireyler evrimsel süreç içinde genetik algoritma işlemcileri tarafından belirlenirler. Problemin bireyler içindeki gösterimi problemden probleme değişiklik gösterir. Genetik algoritmaların problemin çözümündeki başarısına karar vermedeki en önemli faktör, problemin çözümünü temsil eden bireylerin gösterimidir. Nüfus içindeki her bireyin problem için çözüm olup olmayacağına karar veren bir uygunluk fonksiyonu vardır. Uygunluk fonksiyonundan dönen değere göre yüksek değere sahip olan bireylere, nüfustaki diğer bireyler ile çoğalmaları için fırsat verilir. Bu bireyler çaprazlama işlemi sonunda çocuk adı verilen yeni bireyler üretirler. Çocuk kendisini meydana getiren ebeveynlerin (anne, baba) özelliklerini taşır. Yeni bireyler üretilirken düşük uygunluk değerine sahip bireyler daha az seçileceğinden bu bireyler bir süre sonra nüfus dışında bırakılırlar. Yeni nüfus, bir önceki nüfusta yer alan uygunluğu yüksek bireylerin bir araya gelip çoğalmalarıyla oluşur. Aynı zamanda bu nüfus önceki nüfusun uygunluğu yüksek bireylerinin sahip olduğu özelliklerin büyük bir kısmını içerir. Böylelikle, pek çok nesil aracılığıyla iyi özellikler nüfus içersinde yayılırlar ve genetik işlemler aracılığıyla da diğer iyi özelliklerle birleşirler. Uygunluk değeri yüksek olan ne kadar çok birey bir araya gelip, yeni bireyler oluşturursa arama uzayı içersinde o kadar iyi bir çalışma alanı elde edilir. Probleme ait en iyi çözümün bulunabilmesi için;

Bireylerin gösterimi doğru bir şekilde yapılmalı,  
Uygunluk fonksiyonu etkin bir şekilde oluşturulmalı,  
Doğru genetik işlemciler seçilmeli.

Bu durumda çözüm kümesi problem için bir noktada birleşecektir. Genetik algoritmalar, diğer eniyileme yöntemleri kullanılırken büyük zorluklarla karşılaşılacak, oldukça büyük arama uzayına sahip problemlerin çözümünde başarı göstermektedir. Bir problemin bütünsel en iyi çözümünü bulmak için garanti vermezler. Ancak problemlere makul bir süre içinde, kabul edilebilir, iyi çözümler bulurlar. Genetik algoritmaların asıl amacı, hiçbir çözüm tekniği bulunmayan problemlere çözüm aramaktır. Kendilerine has çözüm teknikleri olan özel problemlerin çözümü için mutlak sonucun hızı ve kesinliği açısından genetik algoritmalar kullanılmazlar. Genetik algoritmalar ancak;

- Arama uzayının büyük ve karmaşık olduğu,
- Mevcut bilgiyle sınırlı arama uzayında çözümün zor olduğu,
- Problemin belirli bir matematiksel modelle ifade edilemediği,
- Geleneksel eniyileme yöntemlerinden istenen sonucun alınmadığı alanlarda etkili ve kullanışlıdır.

Genetik algoritmalar parametre ve sistem tanılama, kontrol sistemleri, robot uygulamaları, görüntü ve ses tanıma, mühendislik tasarımları, planlama, yapay zeka uygulamaları, uzman sistemler,

---

<sup>4</sup> [http://tr.wikipedia.org/wiki/Genetik\\_algoritma](http://tr.wikipedia.org/wiki/Genetik_algoritma)

fonksiyon ve kombinasyonel eniyileme problemleri ağ tasarım problemleri, yol bulma problemleri, sosyal ve ekonomik planlama problemleri için diğer eniyileme yöntemlerinin yanında başarılı sonuçlar vermektedir.

Biyolojik Altyapı olarak aşağıdakilerin anlaşılmış olması gerekmektedir<sup>5</sup>.

**Kromozom:** Tüm yaşayan organizmalar hücrelerden oluşur. Her hücrede aynı kromozom kümeleri bulunur. Kromozomlar DNA dizileri olup, tüm organizmanın örneği olarak hizmet ederler. Bir kromozom gen adı verilen DNA bloklarından oluşur. Her gen belirli bir proteini kodlar. Basitçe, her genin, örneğin göz rengi gibi bir özelliği kodladığı söylenebilir. Bir özellik için olası ayarlar, (Örn. Mavi, Yeşil) alel olarak adlandırılır. Her gen kromozom üzerinde kendine ait bir konuma sahiptir. Bu konuma yörünge ("locus") adı verilir. Tüm genetik malzeme kümesine (tüm kromozomlar) genom adı verilir. Genom üzerindeki belli gen kümelerine genotip adı verilir. Genotipler, doğumdan sonra gelişmeyle fenotiplere - canlının göz rengi, zekâ v.b. fiziksel ve zihinsel özellikleri- dönüşür.

**Tekrar Üretim:** Tekrar üretim sırasında, yeniden birleşme (veya çaprazlama) ilk önce ortaya çıkar. Atalardan gelen genler yepyeni bir kromozom üretmek için bir araya gelirler. Bu yeni yaratılmış nesil daha sonra mutasyona uğrayabilir. Mutasyon DNA elemanlarının değişmesidir. Bu değişimler genellikle atalardan gen kopyalanması sırasındaki hatalardan kaynaklanır. Bir organizmanın uygunluğu ("fitness") organizmanın yaşamındaki başarısıyla (hayatta kalma) ölçülür.

Genetik algoritmalar Darwin'in Evrim teorisinden esinlenilerek üretilmiştir. Bir problemin çözümü evrimsel süreç kullanılarak çözülmektedir. Algoritma toplum adı verilen ve kromozomlarla temsil edilen bir çözüm kümesi ile başlamaktadır. Bir toplumdaki çözümler yeni toplumların üretilmesinde kullanılmaktadır. Bu işlem, yeni toplumun eskisinden daha iyi olacağı umuduyla yapılmaktadır. Yeni çözümler (yavru) üretmek için alınan çözümler uygunluklarına (fitness) göre seçilmektedir. Daha uygun olan tekrar üretim için daha fazla şansa sahiptir. Bu süreç belli bir durum (örneğin belli sayıda toplum veya en iyi çözümün gelişmesi) karşılanana kadar tekrar edilmektedir.

Basit Genetik Programlama Taslağı şu şekilde özetlenebilir:

1. n kromozom oluşan rasgele toplum oluşturulur (problemin olası çözümleri)
2. Toplumdaki her x kromozomu için  $f(x)$  uygunluk değeri değerlendirilir.
3. Aşağıdaki adımlar izlenerek yeni toplum üretilir;
  - Toplumdan uygunluklarına göre iki ata seçilir (daha uygun olanın seçilme şansı daha fazladır)
  - Çaprazlama olasılığı ile ataları yeni yavru oluşturmak için birbirleriyle eşleştirilir. Eğer çaprazlama yapılmazsa, yavru ataların tıpatıp aynısı olacaktır.
  - Mutasyon olasılığı ile yeni yavru üzerinde her yörünge için mutasyon işlemi yapılacaktır.
  - Yeni yavru, yeni topluma eklenir.
  - Yeni toplum algoritmanın tekrar işlenmesinde kullanılır.
  - Eğer bitiş durumu sağlandıysa, durup toplumdaki en iyi çözüm döndürülür.
  - Adım 2'ye gidilir.

Çaprazlama ve mutasyon genetik algoritmanın en önemli kısımlarıdır. Başarım en çok bu iki işleçten etkilenir. Bu işleçlerden bahsetmeden önce kromozomlardan daha fazla bahsetmek gereklidir.

---

<sup>5</sup> <http://www.yapay-zeka.org/modules/wiwiwmod/index.php?page=GA>



Bir kromozom temsil ettiđi çözüm hakkında bir şekilde bilgi içermelidir. En çok kullanılan kodlama ikili karakter dizisidir. Bu yöntemle kromozom şu şekilde görölmektedir:

Kromozom 1 : 1101100100110110

Kromozom 2 : 1101111000011110

Her kromozom ikili karakter dizisi şeklinde temsil edilmektedir. Karakter dizisindeki her bit çözümün bir özelliđini temsil eder. Bir başka olasılık tüm karakter dizisinin bir sayıyı temsil etmesidir. Elbette, birçok başka kodlama yöntemi vardır. Kodlama daha çok çözülen probleme bađlıdır. Örneđin bazı problemler için tamsayı veya gerçek sayı şeklinde kodlamak gerekirken, bazı problemlerde permütasyon şeklinde kodlamaya ihtiyaç vardır.

Kodlamaya karar verdikten sonra, çaprazlama işlemiyle devam edebiliriz. Çaprazlama, atalardaki seçili genler üzerinde işlem yapar ve yeni yavrular oluşturur. Bunun en basit şekli, rasgele bir kesme noktası (çaprazlama noktası) seçip, bu noktadan önceki her şeyi ilk atadan, sonraki her şeyi ikinci atadan alıp birleştirerek yavruyu oluşturmaktır.

Çaprazlama aşağıdaki şekilde gösterilebilir: ( | kesme noktasıdır):

Kromozom 1: 11011 | 00100110110

Kromozom 2: 11011 | 11000011110

Yavru 1 : 11011 | 11000011110

Yavru 2 : 11011 | 00100110110

Çaprazlamanın birçok yolu mevcuttur, örneđin birden fazla kesme noktası seçilebilir. Çaprazlama daha da karmaşık olabilir ve tamamen kromozomların kodlanmasına bađlıdır. Özel problemler için yapılmış özel çaprazlamalar genetik algoritmanın başarımını arttırabilir.

Çaprazlama işlemi gerçekleştirildikten sonra, mutasyon işlemi yapılır. Mutasyonun amacı, toplumdaki tüm çözümlerin çözülen problemlerin bir yerel uygun deđerine düşmesinin önüne geçmektir. Mutasyon işlemi çaprazlama sonucu oluşan yavruyu rasgele deđiştirmektedir. İkili kodlamada rasgele seçilmiş bir kaç biti 1'i 0'a, 0'ı 1'e şeklinde deđiştirmek bir mutasyondur.

Asıl Yavru 1: 1101111000011110

Mutasyon Geçirmiş Yavru 1:1100111000011110

Asıl Yavru 2: 1101100100110110

Mutasyon Geçirmiş Yavru 2:1101101100110110

Mutasyon tekniđi (çaprazlama tekniđi de) kromozomların kodlamasına çođunlukla bađlıdır. Örneđin permütasyon şeklinde kodlamada mutasyon rasgele seçilen iki genin yer deđiştirmesi olarak gerçekleştirilir.

Genetik algoritmaların kümeleme işleminde çeşitli tekniklerle kullanımı vardır. Mauli ve Bandyopadhyay tarafından geliştirilen bir teknikden bahsedilebilir. Bu teknik temel olarak veri tabanını k adet kümeye ayırarak genetik algortima eşliđinde küme merkezlerini hesaplar. Aynı zamanda küme merkezleri minimum yapılmaya çalışılır. Küme merkezi kümeyi oluşturan elemanların birbirleri ile olan Öklid mesafesi toplamıdır.

## KAYNAKLAR

- Wavecluster: A multi-resolution clustering approach for very large spatial databases, Sheikholeslami, Gholamhosein and Chatterjee, Surojit and Zhang, Aidong (1998) Wavecluster: A multi-resolution clustering approach for very large spatial databases. In Proceedings of the 24th VLDB conference .
- Veri Madenciliği Yöntemleri, Yalçın Özkan 06'2008
- Veri Madenciliği ,Gökhan Silahtaroglu 06'2008
- İstanbul Ticaret Üniversitesi Dergisi Veri Madenciliği Modelleri Ve Uygulama Alanları (Serhat ÖZEKES)
- [www.bilmuh.gyte.edu.tr/~htakci/vm/verimadenciligi.doc](http://www.bilmuh.gyte.edu.tr/~htakci/vm/verimadenciligi.doc)
- <http://www.gurunlu.com/?tag=/k-means>
- Ayhan Adsız, Metin Madenciliği, Ahmet Yesevi Üniversitesi, Bilişim Sistemleri ve Mühendislik Fakültesi,Dönem Projesi,sayfa:46,2006.
- [http://tr.wikipedia.org/wiki/Genetik\\_algoritma](http://tr.wikipedia.org/wiki/Genetik_algoritma)
- <http://www.yapay-zeka.org/modules/wiwimod/index.php?page=GA>