

Nicel Birliktelik Kural Madenciliği İçin Baskın Olmayan Sıralama Genetik Algoritma-II'nin Duyarlılık Analizi

Araştırma Makalesi/Research Article

 Elif VAROL ALTAY,  Bilal ALATAŞ*

Yazılım Mühendisliği, Fırat Üniversitesi, Elazığ, Türkiye

evarol@firat.edu.tr, balatas@firat.edu.tr

(Geliş/Received:26.12.2018; Kabul/Accepted:06.01.2020)

DOI: 10.17671/gazibtd.503349

Özet— İkili ya da kesikli değerlere sahip veri kümelerine odaklanan birçok birliktelik kural madenciliği çalışması vardır. Ancak, gerçek dünya uygulamalarındaki veriler genellikle nicel değerlerden oluşmaktadır. Nicel veriler için keşfedilecek kurallarda hangi niteliklerin olacağı ve hangilerinin kuralın solunda hangilerinin sağında olacağını belirlemek, ilgili nicel aralıkların en uygun şekilde otomatik ayarlanması; kuralların yoğun nesne kümeleri üretilmeden tek aşamada anlaşılabilir, doğru, güvenilir, ilginç, sürpriz vb. özelliklere sahip olacak şekilde bulunması ve tüm bu işlemlerin her veri tabanı için önceden belirlenmesi gereken metriklere ihtiyaç duyulmadan ayarlanması zor bir problemdir. Yakın zamanda bazı araştırmacılar, nicel birliktelik kural madenciliğini, farklı kriterleri aynı anda en iyi şekilde karşılayacak şekilde, çok amaçlı bir problem olarak düşünmüşlerdir. Bu makalede nicel birliktelik kural madenciliği problemi için anlaşılabilirlik, ilginçlik ve performansı en üst düzeye çıkarmayı amaçlayan çok amaçlı evrimsel algoritmalarından baskın olmayan sıralama genetik algoritma-II temelli QAR-CIP-NSGA-II'nin parametre analizi yapılmıştır. Bu amaçla; nitelikleri nicel değerler alan beş gerçek dünya verisinde QAR-CIP-NSGA-II'nin değerlendirme sayısı, popülasyon sayısı, mutasyon olasılığı, genlik ve eşik değeri gibi parametrelerinin; elde edilen kural sayısı, ortalama destek, güven, lift, kesinlik faktörü, netconf ve kapsanan kayıt sayısını nasıl değiştirdiği kapsamlı bir şekilde bildiğimiz kadarıyla ilk kez bu çalışmada gerçekleştirilmiştir. Detaylı analiz sonuçları karşılaştırmalı tablolar ile sunulmuştur ve yorumlanmıştır.

Anahtar Kelimeler— nicel birliktelik kural madenciliği, çok amaçlı optimizasyon, baskın olmayan sıralama genetik algoritma-II

Sensitivity Analysis of Non-Dominated Sorting Genetic Algorithm-II for Quantitative Association Rules Mining

Abstract—There are many association rules mining studies that focus on datasets with binary or discrete values. However, the data in real-world applications are generally composed of quantitative values. In association rules discovered within quantitative data, it is very hard to determine which attributes will be included in the rules to be discovered and which ones will be on the left of the rule and which ones on the right; to automatically adjust of most relevant ranges for numerical attributes; to rapidly discover the reduced high-quality rules directly without generating the frequent itemsets; to ensure the rules to be comprehensible, surprising, interesting, accurate, confidential, and etc.; to adjust all of these processes without the need for the metrics to be pre-determined for each dataset. Recently, some researchers have considered quantitative association rule mining as a multi-objective problem that best meets different criteria at the same time. In this paper, the parameter analysis of non-dominated sorting genetic algorithm-II based QAR-CIP-NSGA-II, which aims to maximize comprehensibility, interestingness, and performance for quantitative association rule mining problem, has been performed. For this purpose, to the best of our knowledge the effects of the parameters of QAR-CIP-NSGA-II such as the number of evaluations, population number, mutation probability, amplitude and threshold value to the number of rules obtained, average support, confidence, lift, certainty factor, netconf, and the number of records covered in five real-world data whose attributes consist of quantitative values have been carried out for the first time in this study. Detailed sensitivity analysis results are presented and interpreted in comparative tables.

Keywords— quantitative association rules mining, multi-objective optimization, non-dominated sorting genetic algorithm-II

1. GİRİŞ (INTRODUCTION)

Veri madenciliği geniş veri yığınlarının bulunduğu veri tabanlarından daha önceden keşfedilmemiş doğru, anlaşılabilir ve ilginç bilgilerin otomatik ya da yarı otomatik yollarla çıkarılıp kullanılması olarak tanımlanabilir. Birliktelik kurallarının keşfi büyük veri setlerinde ilginç bilgileri ortaya çıkarmak için kullanılan en yaygın veri madenciliği yöntemlerinden biridir. Literatürde, farklı tipte verileri barındıran ve özellikle nümerik ya da nicel veri içeren veri tabanlarında birliktelik kural madenciliği yöntemlerinin başarısı düşüktür. Çünkü nicel verilerde birliktelik kural keşfi için çoğunlukla veriler kesikli hale getirilmekte ve bu şekilde kategorik verilerde çalışan algoritmaların işleyebileceği hale dönüştürülmektedir. Ancak bu şekilde bir yaklaşım birliktelik kural madenciliğinin otomatik kural bulma özelliğine uygun değildir. Kesikli hale getirilen verilerin ilgili sınırlarının önceden belirlenmesiyle aslında veri tabanı değişmiş olmaktadır. Böylece bilgi kayıpları olabilmekte ve değiştirilmiş verideki birliktelik kuralları bulunmaktadır. Hangi aralıklara göre kesikli hale getirilme işleminin yapılması zor bir problemdir, çünkü ilgili aralıklar dışında doğru ve ilginç kuralları barındırma olasılığı bulunan verinin bu şekilde değiştirilmesi mantıksız bir ön işlemdir. Çünkü nitelikler arası etkileşim de göz önüne alındığında kesikli hale getirilen ilgili niteliklerin bu sınırları dışındaki önemli kurallar (belki de çok doğru ve ilginç) bulunamayacaktır. Yani kesikleştirme ya da ayrıklaştırma gibi bir ön işlem yapılmadan bu aralıkların veri madenciliği süreci içinde otomatik olarak bulunması gerekmektedir.

Literatürde birliktelik kural madenciliği için kullanılan algoritmalar iki aşamalı olarak çalışmaktadır. İlk aşamada yoğun nesne kümesi bulunmakta, ikinci aşamada ise bu yoğun nesne kümelerinden kurallar ortaya çıkarılmaktadır. Bu iki aşamayı tek aşama olarak ve gereksiz sayıda çok fazla kuraldan arınmış az sayıda kaliteli ve anlaşılabilir kuralların bulunması şekline getirecek kaliteli modellerin önerilip kullanılması ihtiyaç olarak görülmektedir.

Son yıllarda evrimsel algoritmalar özellikle genetik algoritmalar, bazı araştırmacılar tarafından nicel değerlere sahip veri setlerinden birliktelik kural madenciliği yapmak için kullanılmıştır. Genetik algoritmaları bilgi çıkarma görevlerine uygulamak için temel motivasyon, aday çözümler yerine küresel bir arama yapan güçlü ve uyarlamalı arama algoritmaları olmalarıdır. Son zamanlarda bazı araştırmacılar mevcut yaklaşımların sınırlarının bazılarını kaldırarak birliktelik kurallarının çıkarılmasını tek bir amaç yerine çok amaçlı bir problem olarak ortaya koymuşlardır. Birliktelik kurallarının çıkarılması sürecinde daha ilginç ve doğru kurallar içeren bir dizi elde etmek için çeşitli amaçlar göz önünde bulundurulur. Bu şekilde, kullanılan veri setine ve ondan elde edilebilecek bilgi türüne bağlı olarak destek, güven vb. gibi ölçütler birlikte optimize edilebilmektedir.

Yakın zamanda çok amaçlı evrimsel algoritmalarından “Baskın Olmayan Sıralama Genetik Algoritması-II” [1]

nicel birliktelik kural madenciliği için uyarlanmış ve QAR-CIP-NSGA-II adıyla önerilmiştir [2]. QAR-CIP-NSGA-II yönteminde ilginçlik, anlaşılabilirlik ve performans kriterleri göz önüne alınarak veri tabanlarında etkili çok amaçlı nicel birliktelik kural madenciliği gerçekleştirilmiştir.

Bu makalede nicel birliktelik kural madenciliği problemi için; anlaşılabilirlik, ilginçlik ve performansı en üst düzeye çıkarmayı amaçlayan QAR-CIP-NSGA-II yönteminin parametre analizi yapılarak hangi parametrenin hangi aralıklarda daha iyi sonuçlar verdiği ilk kez analiz edilmiştir. QAR-CIP-NSGA-II'nin fonksiyon değerlendirme sayısı, popülasyon sayısı, mutasyon olasılığı, genlik değeri ve eşik değeri gibi parametrelerinin farklı gerçek dünya verilerinden Basketball, Bodyfat, Pollution, Quake ve Stock Price'da çok amaçlı nicel birliktelik kural madenciliği problemi için belirlenen amaçlara etkisi incelenmiştir.

Bu makalenin organizasyonu şu şekildedir. Birinci bölümde genel bir giriş yapılmış olup problemin tanımı analiz edilmiştir. İkinci bölümde nicelik birliktelik kuralları ele alınmış olup çalışmada kullanılan ölçütler ve literatür çalışması yer almaktadır. Üçüncü bölümde kullanılan algoritma anlatılmış olup, dördüncü bölümde kullanılan veri seti ve deney sonuçları sunulmuştur. Bu makalenin son bölümü olan beşinci bölümde makale sonuçlandırılmış olup ileride yapılabilecek çalışmalara yer verilmiştir.

2. NİCEL BİRLİKTELİK KURALLARI (QUANTITATIVE ASSOCIATION RULES)

Birliktelik kural madenciliği, geçmiş verileri analiz edip bu veriler içindeki birliktelik davranışlarını tespit ederek geleceğe yönelik çalışmalar yapılmasını destekleyen bir yaklaşımdır [3]. Birliktelik kuralı, veri kümesindeki öğeler arasındaki bağımlılıkları göstermek ve tanımlamak için kullanılmaktadır. Birliktelik kurallarının ifade ediliş şekli $X \rightarrow Y$ biçimindedir. Burada X ve Y nesne kümeleridir ve $X \cap Y = \emptyset$ 'dir. İkili ya da kesikli değerlere sahip veri kümelerine odaklanan önceden yapılmış birçok birliktelik kural madenciliği çalışması vardır. Ancak, gerçek dünya uygulamalarındaki veriler genellikle nicel değerlerden oluşmaktadır. Bu bölümde nicel birliktelik kuralları için ölçütler ve literatürde yapılan çalışmalar incelenmiştir.

2.1. Ölçütler (Measures)

Destek ve güven birliktelik kurallarında en çok kullanılan ölçütlerdir. $X \rightarrow Y$ kuralı için bu ölçütler Denklem 1 ve Denklem 2'de gösterilmiştir.

$$\text{Destek}(X \rightarrow Y) = \frac{DES(XY)}{|D|} \quad (1)$$

$$\text{Güven}(X \rightarrow Y) = \frac{DES(XY)}{DES(X)} \quad (2)$$

Burada $DES(XY)$ X ve Y 'nin veri setinde birlikte geçtiği kayıt sayısı, $|D|$ toplam kayıt sayısı, $DES(X)$ ise X 'in yer aldığı kayıt sayısını ifade etmektedir.

Lift ölçütü [4], kuralın güven değeri ile kuralın beklenen güven değeri arasındaki oranı temsil eder. $[0, \infty)$ aralığında değer alabilen bu ölçütte, 1'den küçük değerler negatif bağımlılık anlamına gelirken, 1 bağımsızlık ve 1'den yüksek değerler pozitif bağımlılık anlamına gelir. $X \rightarrow Y$ kuralı için lift ölçütü Denklem 3'teki gibi ifade edilmektedir:

$$Lift(X \rightarrow Y) = \frac{Güven(X \rightarrow Y)}{DES(Y)} = \frac{DES(XY)}{DES(X)DES(Y)} \quad (3)$$

Kesinlik faktörü (KF) [5], yalnızca X 'in mevcut olduğu işlemleri göz önüne aldığımızda, Y 'nin bir işlemde olması olasılığındaki bir değişimin ölçütü olarak yorumlanmaktadır. Bu ölçüt, güven ölçütü tarafından algılanmayan yanıtıcı kuralların keşfedilmesini önlemektedir. $[-1, 1]$ aralığında değerler alır. Burada 0'dan küçük değerler negatif bağımlılığı temsil etmekte, 0 bağımsızlığı temsil etmekte ve 0'dan yüksek değerler pozitif bağımlılığı temsil etmektedir. $X \rightarrow Y$ kuralı için KF ölçütü üç şekilde tanımlanmaktadır:

$$\begin{cases} KF(X \rightarrow Y) = \frac{Güven(X \rightarrow Y) - DES(Y)}{1 - DES(Y)}, & Güven(X \rightarrow Y) > DES(Y) \\ KF(X \rightarrow Y) = \frac{Güven(X \rightarrow Y) - DES(Y)}{DES(Y)}, & Güven(X \rightarrow Y) < DES(Y) \\ KF(X \rightarrow Y) = 0, & Aksi Taktirde \end{cases} \quad (4)$$

Netconf [6] ölçütü, kuralın desteğine ve kuralın sol kısmının ve sağ kısmının desteğine dayanan birliktelik kurallarının ilginçliğini değerlendirmektedir. KF ölçüsünde olduğu gibi, *netconf*'da da güven değeri ile algılanmayan yanıtıcı kurallar tespit edilebilir. Bu ölçüt, pozitif değerlerin pozitif bağımlılığı, negatif değerlerin negatif bağımlılığı ve 0'ın bağımsızlığı temsil ettiği $[-1, 1]$ aralığında değerlerden oluşmaktadır. $X \rightarrow Y$ kuralı için *netconf* ölçütü Denklem 5'teki gibi ifade edilmektedir.

$$netconf = DES(XY) - \frac{DES(X)DES(Y)}{DES(X)(1-DES(X))} \quad (5)$$

2.2. Literatür Çalışması (Literature Work)

Agrawal ve Srikant, ilk önce nitelik alanını küçük aralıklara bölmekte ve daha sonra birleşik aralıkların yeterli desteğe sahip olmasını sağlamak için daha büyük bir alanla birleştirmektedir. Böylece nicel birliktelik kural madenciliği problemi Boolean kurallarına dönüştürülmektedir. Değişkenlerin ayrıştırılması genellikle bilgi kaybına neden olmaktadır. Ayrıca bu yaklaşımlarda diğer nitelikler dikkate alınmadan bir nitelik ayrıştırılır ve nitelik etkileşim sorununa neden olur [7].

Chiu ve arkadaşları nicel verilerde birliktelik kural madenciliği için bulanık küme tabanlı bir yöntem önermiştir. Bu yöntem veri setini tek bir kez tarayarak kümeleme tabanlı bulanık küme tabloları oluşturmakta ve

sonra kayıtlar, kayıt uzunluğu olan k dikkate alınarak k 'ncü küme tablosuna kümelenebilir [8].

Alataş ve arkadaşları nicel birliktelik kural madenciliğini çok amaçlı bir optimizasyon problemi olarak ele almış ve diferansiyel evrimsel algoritmasını bu problem için çözüm yöntemi olarak uyarlamıştır. MODENAR adını verdikleri bu yöntemde Pareto optimallik temel alınmış ve algoritmanın tek bir çalıştırma sonunda az sayıda kaliteli doğru ve anlaşılır birliktelik kurallarının çıkarılması amaçlanmıştır [9].

Chen ve arkadaşları hem üyelik fonksiyonlarını hem de birliktelik kurallarını nicel işlemlerden ayırmak için çok amaçlı genetik bulanık algoritma önermişlerdir [10].

Yan ve arkadaşlarının çalışmasında, minimum destek değeri kullanmadan genetik algoritma ile birliktelik kurallarının bulunması amaçlanmıştır. Güven değeri uygunluk fonksiyonu olarak kullanılmıştır. Çalışmanın daha etkili olabilmesi için FP ağaç yapısı kullanılmıştır. Bu yaklaşım; yüksek performanslı birliktelik kural madenciliği ve sistem otomasyonu konusunda fayda sağlamıştır. Gerçek dünya verisi kullanılarak deneyler yapılmış olup iyi sonuçlar elde edilmiştir [11].

Beiranvand ve arkadaşları nicel birliktelik kurallarının keşfi için çok amaçlı Pareto temelli PSO algoritması önermiştir. Bu algoritmanın etkinliğini belirtmek için çok amaçlı optimizasyon yaklaşımına güven, anlaşılabilirlik ve ilginçlik özellikleri eklenmiştir. Bu yöntemin iyi sonuçlar verdiği belirtilmiştir [12].

Martin ve arkadaşları anlaşılır, ilginç ve yüksek performanslı nicel birliktelik kural keşfi için çok amaçlı evrimsel bir model önermişlerdir [13]. Piri ve Dey nicel veri tabanından yararlı ve ilginç kurallar çıkarmak için güven, anlaşılabilirlik, ilgi çekicilik gibi amaçları içeren Pareto tabanlı parçacık sürüsü optimizasyonu kullanmışlardır [14].

Almasi ve Abadeh, Rare-PEAR adlı yeni çok amaçlı evrimsel bir algoritma önermişlerdir. Yöntemin diğer yöntemlerden ayrılan özelliği, iki önemli amaç olarak ilginçliğe ve gereksizliğe dikkat etmiş olmalarıdır. Rare-PEAR algoritması ilginç, nadir veya nadir ve ilginç kurallar bulmuştur [15].

Kahvazadeh ve Abadeh, nicel verilerde birliktelik kurallarını çıkarmak için MOCANAR isimli meta sezgisel bir algoritma önermişlerdir. MOCANAR, nicel veri kümelerinden yüksek kaliteli birliktelik kurallarını ayıklayan Pareto tabanlı çok amaçlı bir guguk kuşu arama algoritmasıdır. Destek, güven, ilgi çekicilik ve anlaşılabilirlik, MOCANAR'da ele alınan amaçlardır. MOCANAR her çalıştırmada az sayıda yüksek kaliteli kurallar çıkarmaktadır [16].

Martinez-Ballesteros ve arkadaşları çok-amaçlı evrimsel algoritma olan NSGA-II'nin geliştirilmiş versiyonunu

nicelik birliktelik kurallarını keşfetmek için önermişlerdir [17].

Kuo ve arkadaşları nicel verilerde birliktelik kural madenciliği çıkarmak için Pareto optimal stratejisine dayanan uyarlamalı bir arşiv ızgarası kullanarak çok amaçlı bir parçacık sürüsü optimizasyonu önermişlerdir. Önerdikleri yöntem kural keşfi için güven, anlaşılabilirlik ve ilginçliği optimize etmeyi amaçlamaktadır. Bu yöntem ayrıklaştırma gibi önemli bir ön işlem gerektirmez. Minimum destek ve güven ön koşul değildir [18].

Nicel birliktelik kural madenciliği için büyük veri kümelerinde paralel parçacık yönelimli ve veri yönelimli PSO algoritmaları da önerilmiştir [19]. Bu çalışmada, daha ölçeklenebilir ve verimli sonuçlar elde etmek için PSO paralel hale getirilerek hesaplanabilir maliyetin azaltılması amaçlanmıştır. Kurt avı davranışına dayanan Kurt Arama Algoritması nicel birliktelik kural madenciliği problemi için kullanılan başka bir sürü tabanlı optimizasyon algoritmasıdır [20]. Bu çalışmada deney sonucu veya simülasyon bulunmamaktadır sadece metodoloji teorik olarak açıklanmıştır.

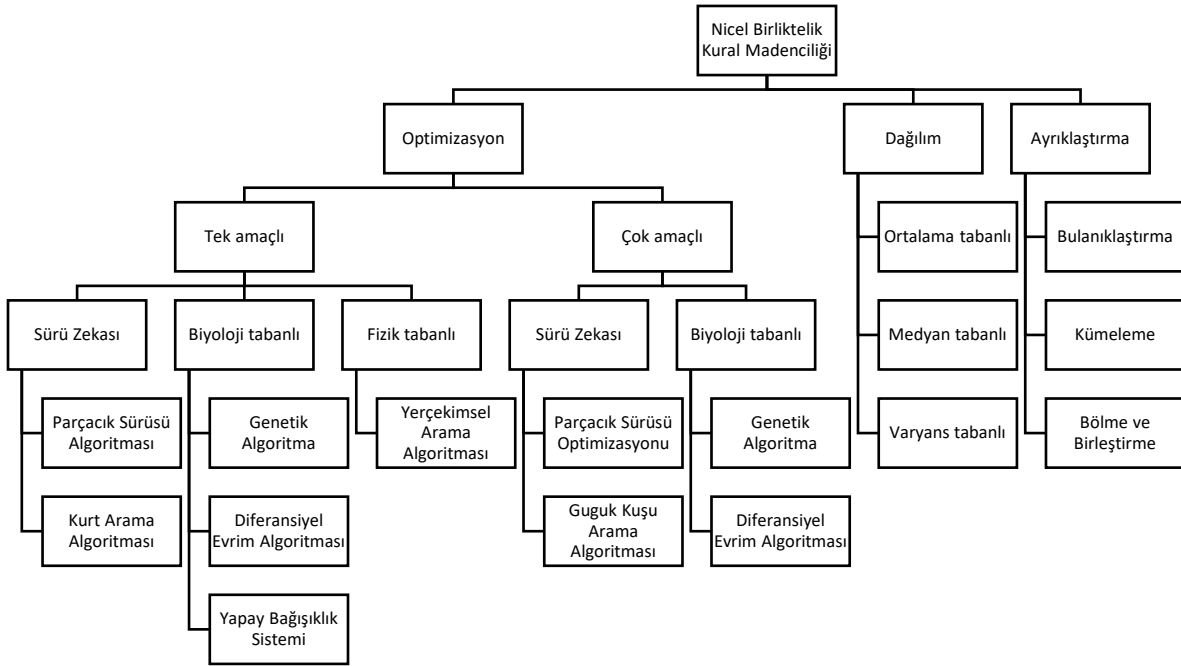
Badhon ve arkadaşları yakın zamanda önerilen çok amaçlı evrimsel algoritmaların (ÇAEA) kategorik, nicel ve bulanık kural keşfinde; kromozom temsili, genetik

operatörler ve başlangıç popülasyonu açısından sistematik yapısal bir incelemesini sunmuşlardır. ÇAEA temelli çeşitli yaklaşımlarla ilgili net bir karşılaştırmalı çalışma, hesaplama karmaşıklığı ve bu araştırma bağlamında çeşitli uygulamaları içermektedir [21].

Altay ve Alataş nicel birliktelik kural madenciliğini tek amaçlı ve çok amaçlı bir optimizasyon problemi olarak kabul eden algoritmaları incelemişler ve bu algoritmaların performans analizini belirli metrikler kullanarak bulmuşlardır [22, 23].

Zhang ve arkadaşları, ikili, kategorik ve nicel veriler için uygun olan evrimsel algoritma tabanlı birliktelik kural madenciliği için genel bir çerçeve önermişlerdir. Önerilen çerçeve P-EAARM, veri setinin taranmasında etkili stratejiler geliştirmiş ve arşiv olarak daha fazla baskın olmayan çözüm kaydedecek bir esnek dağıtılmış veri kümesi uygulaması kullanır. Hesaplama maliyeti ve hızlandırması analiz edilmiştir. Sonuçlar, paralel PSO algoritmasında belirtilen tarama yöntemiyle karşılaştırıldığında, P-EAARM'in hesaplanan maliyet açısından daha iyi performansa sahip olduğu görülmektedir [24].

Nicel birliktelik kurallarını keşfetmeye yönelik üç ana yaklaşım Şekil 1'de gösterilmiştir.



Şekil 1. Nicel birliktelik kural madenciliği problemi için yöntemler
(Methods for quantitative association rules mining problem)

3. QAR-CIP-NSGA-II YÖNTEMİ (QAR-CIP-NSGA-II METHOD)

3.1. Kodlama şekli ve gen havuzu (Encoding scheme and initial gene pool)

Nicel birliktelik kural madenciliği problemi için her kromozom, kuralın özelliklerini ve aralıklarını temsil eden bir gen vektörüdür. Konumsal bir kodlama kullanılmıştır. Aralığın öğrenilmesiyle koşul seçimini birleştirmek için her gen üç bölümden oluşur [2].

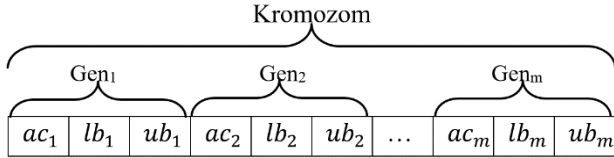
İlk kısım ac , bir genin kurala dahil olup olmadığını göstermektedir. Bu bölüm 1 veya 0 ise bu nitelik kuralın sağ kısmı ya da kuralın sol kısmının bir parçasıdır. Eğer -1 ise bu nitelik kuralda yer almaz. İlk kısımlarında 1 olan genler kuralın sağ kısmını oluştururken 0 olan genler sol kısmını oluşturacaktır.

İkinci kısım lb , nitelik aralığının alt sınırını temsil etmektedir. Üçüncü kısım ub , nitelik aralığının üst sınırını temsil etmektedir [2].

Öznitelik nominal ise lb ve ub eşit olacaktır ve nominal özneliğin sadece tek bir değerini temsil edecektir. Böylece bir kromozom C_T Denklem 6 ve Denklem 7'deki gibi kodlanır. Burada m veri kümesindeki özneliklerin sayısıdır. Kromozomun şeması Şekil 2'de gösterilmektedir.

$$Gen_i = (ac_i, lb_i, ub_i), i = 1, \dots, m \quad (6)$$

$$C_T = Gen_1 Gen_2 \dots Gen_m \quad (7)$$



Şekil 2. Bir kromozomun temsil biçimi
(Representation scheme of a chromosome)

Aralıkları toplam alana yayılana kadar artırmaktan kaçınmak için, genlik belirlenen bir niteliğin elde edebileceği maksimum boyut olarak tanımlanır. Böylece i niteliğin genliği Denklem 8'deki gibi tanımlanır [2]:

$$Genlik_i = (Max_i - Min_i) / \delta \quad (8)$$

Burada Max_i ve Min_i i niteliğinin sırasıyla maksimum ve minimum değerleridir. δ ise kuralların genelleştirilmesi ve özgüllüğü arasındaki dengeyi belirleyen sistem uzmanı tarafından verilen bir değerdir.

Bu algoritmada başlangıç gen havuzu ya da popülasyonu veri setini iyi bir şekilde kapsayan ve kuralın sağ bölümünde yalnızca bir nitelik içeren kural kümesini temsil etmektedir. Başlangıç popülasyonunu oluşturmak için ilk önce kuralların sağ kısmı ve kuralların sol kısım nitelikleri rastgele seçilmektedir. Daha sonra veri setinden rastgele işaretli bir örnek seçilir ve her bir niteliğin genliğinin %50'sine eşit bir boyutta ve her birinin ortasında seçilen örneğin değeriyle oluşturulur. Aralığın alt ve üst limitleri aşmaması için düzenleme yapılmıştır. Son olarak bu kural için kapsanan örnekler veri setinde işaretlenmiştir. Bu işlem başlangıç popülasyonu tamamlanmaya kadar tekrar edilir. Eğer örneklerin tamamı işaretlenmiş ve başlangıç popülasyonu tamamlanmamışsa örneklerin tamamı tekrar işaretlenmeyecektir ve başlangıç popülasyonu tamamlanmaya kadar işlemin

tekrarlanacağına dikkat edilmelidir. QAR-CIP-NSGA-II'nin sözde kodu Şekil 3'te verilmiştir.

Giriş: Popülasyon boyutu, değerlendirme sayısı, mutasyon olasılığı, genlik faktörü, eşik değeri
Çıkış: Harici popülasyon

Adım 1: Başlangıç popülasyonunu üret ve değerlendir. Baskın olmayan çözümleri bul ve harici popülasyonu başlat.

Adım 2: Pareto sıralamasına ve kalabalık ölçütünü baz alarak ikili turnuva seçimine göre bir çift ata çözüm seç. Çaprazlama, mutasyon ve düzeltme operatörlerini uygula.

Adım 3: Yeni bireyleri değerlendir.

Adım 4: Mevcut ve çocuk popülasyonları tek bir popülasyonda birleştir.

Adım 5: Birleştirilmiş popülasyonun baskın olmayan bütün Pareto cephelemlerini oluştur ve kalabalık ölçütünü hesapla ve en iyisini seç.

Adım 6: Harici popülasyonu güncelle.

Adım 7: Popülasyonu tekrar başlatma kriterini test et.

Adım 8: Maksimum değerlendirme sayısına ulaşılmadıysa Adım 2'ye git.

Adım 9: Harici popülasyondaki fazlalığı at, başkalarının alt kromozomları olan bireyleri sil.

Adım 10: Harici popülasyonu döndür.

Şekil 3. QAR-CIP-NSGA-II'nin sözde kodu
(Pseudo-code of QAR-CIP-NSGA-II)

3.2. Amaçlar (Objectives)

Bu problem için üç amaç belirlenmiştir. Bunlar ilginçlik, anlaşılabilirlik ve performanstır. Performans, yerel ve genel kurallar arasında iyi bir dengeyle doğru kural bulmamıza izin veren destek ve KF 'nin çarpımı olarak düşünülmüştür. Nesnelere arasındaki pozitif bağımlılığı temsil eden ve destek dezavantajını kaldıran yalnızca çok güçlü kurallarla ilgilenilmektedir. Böylece bir $X \rightarrow Y$ kuralının aşağıdaki şartları sağlaması beklenmektedir:

$$KF(X \rightarrow Y) > 0.$$

$$Destek(X \rightarrow Y) > Minimum Destek.$$

$$\neg(Destek(X \rightarrow Y) > (1 - Minimum Destek)).$$

Bu ölçü [0,1] aralığında değerler alır. Performans değeri 1'e yakın olan bir kural kullanıcı için daha yararlı olmaktadır [2].

İlgincilik, kuralın ne kadar ilginç olduğunu ölçer. Bu, yalnızca kullanıcıların ilgisini çekebilecek kuralları çıkarmamızı sağlar. Bu durumda, ilginçlik ölçütü lift ölçütü olarak kullanılmıştır.

Son olarak anlaşılabilirlik, kuralın anlaşılma kolaylığını ölçmeye çalışmaktadır. Oluşturulan kurallar, anlamakta zorlanılan çok sayıda nitelik içerebilir. Oluşturulan kurallar

kullanıcı için anlaşılabilir değilse, kullanıcı bunları kullanmayacaktır. Burada amaç $X \rightarrow Y$ kuralının içerdiği nitelik sayısına göre anlaşılabilirliği ölçmektir. Anlaşılabilirlik Denklem 9'daki gibi tanımlanmaktadır:

$$\text{anlaşılabilirlik } X \rightarrow Y = 1/\text{Nit}_{X \rightarrow Y} \quad (9)$$

Burada $\text{Nit}_{X \rightarrow Y}$ kuralın sol kısmında yer alan niteliklerin sayısıdır. Bu algoritma ile kuralın sağ kısmında sadece bir tane nitelik içeren kurallar ele alınmaktadır.

3.3. Algoritmanın Karmaşıklığı (Algorithm Complexity)

QAR-CIP-NSGA-II algoritması tarafından üretilen aday çözümlerin veri setiyle karşılaştırılması, algoritmik karmaşıklık hesabındaki en kritik kısımdır ve bu algoritmanın genel karmaşıklığı büyük ölçüde beş parametreye bağlıdır: veri setindeki niteliklerin sayısı (NS), veri setindeki kayıt sayısı (KS), aday çözümlerin sayısı (AS), verilen bir evrim sürecinde iterasyon (İS) sayısı ve elde edilecek birliktelik kural sayısı (BKS). Aday çözümlerin sayısı genellikle veri setindeki kayıt sayısından daha azdır. Dolayısıyla, aday çözümler üzerinde birkaç tarama yapmak, veri kayıtlarını pek çok kez kullanmaktan daha etkilidir. Veri setleri bir diskte depolanır ve aday çözümler bilgisayar belleğinde bulunur. Eşleştirme, popülasyondaki her bir aday çözümün veri setindeki her bir veri kaydına karşı bir genler arası karşılaştırmayı içermektedir ve algoritmik karmaşıklık (AK) hesabında kullanılacak bu tür karşılaştırmaların toplam sayısı

$$\text{AK} = \text{NS} \times \text{KS} \times \text{AS} \times \text{İS} \times \text{BKS} \quad (10)$$

olarak hesaplanmaktadır.

4. DENEY SONUÇLARI (EXPERIMENTAL RESULTS)

Çok amaçlı evrimsel algoritmalarından baskın olmayan sıralama genetik algoritma-II temelli QAR-CIP-NSGA-II'nin parametre analizi yapılarak hangi parametrenin daha iyi performans verdiği analiz edilmiştir. Bu bölümde performans analizi yapmak için beş gerçek dünya verisi kullanılmıştır. Bu veri setleri Bilkent Üniversitesi Function Approximation Repository [25]'ten alınmıştır. Veri setlerinin özellikleri Tablo 1'de, algoritmanın varsayılan parametreleri ise Tablo 2'de gösterilmektedir. Tüm deneyler için algoritma beşer kez çalıştırılmıştır.

İlk deneyde popülasyon sayısı, mutasyon olasılığı, genlik ve eşik değeri ilgili çalışmada [2] verilen varsayılan değerler olup Tablo 2'deki gibidir. Değerlendirme sayısı 10000 ile 1000000 arasında olabilmektedir. Bu deneyde de değerlendirme sayısı 10000, 30000, 50000, 100000 ve 1000000 olarak seçilmiş olup; beş veri setinde

değerlendirme sayısı arttıkça kural sayısı, ortalama güven değeri, ortalama lift, ortalama KF , ortalama netconf değerlerinin arttığı Tablo 3'te görülmektedir.

Tablo 1. Veri setinin özellikleri
(Features of the dataset)

Veri Seti	Kayıt	Nitelik
Basketball	96	5
Bodyfat	252	18
Pollution	60	16
Quake	2178	4
Stock price	950	10

Tablo 2. Algoritmanın varsayılan parametreleri
(Default parameters of the algorithm)

Algoritma	Parametre
QAR-CIP-NSGA-II	Değerlendirme sayısı: 50000, Popülasyon sayısı: 100, Mutasyon olasılığı: 0.1, Genlik (δ): 2.0, Eşik değeri: 5.0

İkinci deneyde değerlendirme sayısı, mutasyon olasılığı, genlik ve eşik değeri makalede verilen varsayılan değerler olup Tablo 2'deki gibidir. Popülasyon sayısı 2 ile 1000 arasında olabilmektedir. Bu deneyde popülasyon sayısı 10, 50, 100 ve 500 olarak seçilmiştir. Beş veri setinde de kural sayısı, ortalama güven değeri, ortalama lift, ortalama KF ve ortalama netconf değerleri popülasyon sayısı arttıkça artmaktadır. Beş veri setinde popülasyon sayısına göre elde edilen ölçütlerin değerlerindeki değişim Tablo 4'te görülmektedir.

0 ile 1 arasında değişebilen mutasyon olasılığına göre elde edilen sonuçlar Tablo 5'te görülmektedir. Yapılan bu analize göre mutasyon olasılık değerinin 0.1 olduğu durumlarda genel olarak küçük boyutlu veri setlerinde en yüksek güven değerlerinin elde edildiği görülmektedir. Nispeten daha büyük boyutlu veri setlerinde mutasyon olasılığının artışı ile güven değerinde de az da olsa bir artış görülmektedir.

2 ile 20 arasında değerler alabilen genlik değerine göre yapılan parametre analizi sonuçları Tablo 6'da verilmiştir. Genlik değeri 2 seçildiğinde, ortalama destek değerinin tüm veri setlerinde en yüksek değere ulaştığı görülmektedir. Basketball, Bodyfat ve Pollution veri setlerinde genlik değerinin 5, 10, 15 ve 20 şeklinde değişmesinin ortalama güven değerine bir etkisi olmamıştır. Eşik değerinin 5, 10, 25, 50 ve 100 değerlerine bağlı olarak belirlenen ölçütlerden elde edilen değerler Tablo 7'de listelenmiştir. Bu tablodan da görüldüğü gibi, eşik değeri arttıkça, veri setlerinden elde edilen kural sayısı ve kuralların ortalama güven ve ortalama kesinlik faktörü değerleri azalmıştır. Quake veri seti hariç, genel olarak diğer veri setlerinde eşik değeri 10 olarak belirlendiğinde, hemen hemen tüm ölçütler cinsinden daha iyi sonuçlar elde edilmiştir.

Tablo 3. Değerlendirme sayısına göre analiz

(Analysis based on the number of evaluation)

Değ. sayısı	Kural sayısı	Ort. Destek	Ort. Güven	Ort. Lift	Ort. KF	Ort. netconf	Kapsanan kayıt sayısı
Basketball							
10000	69	0.06	0.93	58.12	0.91	0.87	89.59
30000	130	0.04	0.98	75.23	0.97	0.94	92.71
50000	190	0.03	0.99	80.09	0.98	0.95	95.84
100000	274	0.03	0.99	83.56	0.99	0.96	98.96
1000000	570	0.03	1.0	86.16	1.0	0.98	100.0
Bodyfat							
10000	51	0.1	0.91	59.79	0.89	0.86	85.32
30000	81	0.11	1.0	136.72	0.99	0.98	97.62
50000	129	0.1	1.0	151.87	1.0	0.99	99.61
100000	205	0.09	1.0	173.94	1.0	0.99	100.0
1000000	1061	0.06	1.0	181.45	1.0	1.0	100.0
Pollution							
10000	64	0.1	0.98	33.88	0.97	0.9	98.34
30000	170	0.06	1.0	48.71	0.99	0.96	95.0
50000	230	0.05	1.0	51.16	0.99	0.97	100.0
100000	431	0.05	1.0	54.39	1.0	0.99	100.0
1000000	2265	0.03	1.0	58.05	1.0	1.0	100.0
Quake							
10000	66	0.06	0.86	262.84	0.82	0.7	54.37
30000	124	0.09	0.9	376.45	0.86	0.62	73.24
50000	153	0.09	0.92	469.73	0.88	0.61	74.52
100000	167	0.1	0.92	570.68	0.87	0.56	80.17
1000000	438	0.05	0.97	1215.66	0.96	0.35	79.48
Stock Price							
10000	74	0.06	0.91	81.36	0.91	0.89	59.9
30000	123	0.06	0.93	176.12	0.92	0.9	60.74
50000	104	0.08	0.94	52.6	0.94	0.91	64.11
100000	109	0.1	0.94	50.26	0.93	0.9	74.64
1000000	241	0.08	0.97	145.05	0.97	0.95	84.22

Tablo 4. Popülasyon sayısına göre analiz

(Analysis based on number of population)

Pop. sayısı	Kural sayısı	Ort. Destek	Ort. Güven	Ort. Lift	Ort. KF	Ort. netconf	Kapsanan kayıt
Basketball							
10	172	0.03	0.98	76.18	0.97	0.95	100.00
50	170	0.03	0.99	81.51	0.98	0.96	96.88
100	190	0.03	0.99	80.09	0.98	0.95	95.84
500	228	0.03	0.99	85.34	0.99	0.96	94.80
Bodyfat							
10	75	0.13	0.98	127.46	0.97	0.95	100.0
50	125	0.13	1.00	134.61	1.00	0.96	99.61
100	129	0.10	1.00	151.87	1.00	0.99	99.61
500	256	0.06	1.00	212.64	1.00	0.99	99.61
Pollution							
10	117	0.05	0.99	51.58	0.98	0.97	100.00
50	204	0.06	1.00	51.54	0.99	0.97	100.00
100	230	0.05	1.00	51.16	0.99	0.97	100.00
500	384	0.04	1.00	56.08	1.00	0.99	100.00
Quake							
10	64	0.04	0.91	712.15	0.89	0.53	74.57
50	115	0.08	0.91	525.65	0.87	0.59	71.22
100	153	0.09	0.92	469.73	0.88	0.61	74.52
500	170	0.08	0.92	303.08	0.89	0.71	68.32
Stock Price							
10	64	0.05	0.93	71.90	0.93	0.90	95.79
50	79	0.08	0.93	61.69	0.92	0.89	79.58
100	104	0.08	0.94	52.60	0.94	0.91	64.11
500	205	0.04	0.98	428.23	0.98	0.97	59.58

Tablo 5. Mutasyon olasılığına göre analiz

(Analysis based on mutation probability)

<i>Mut. olasılığı</i>	<i>Kural sayısı</i>	<i>Ort. Destek</i>	<i>Ort. Güven</i>	<i>Ort. Lift</i>	<i>Ort. KF</i>	<i>Ort. netconf</i>	<i>Kapsanan kayıt</i>
Basketball							
0.1	190	0.03	0.99	80.09	0.98	0.95	95.84
0.3	204	0.04	0.99	78.86	0.98	0.95	95.84
0.5	223	0.04	0.99	77.3	0.98	0.95	97.92
0.7	193	0.04	0.98	77.12	0.97	0.94	100.0
0.9	162	0.06	0.98	66.65	0.96	0.91	96.88
Bodyfat							
0.1	129	0.10	1.00	151.87	1.00	0.99	99.61
0.3	105	0.12	1.00	140.43	1.00	0.98	98.42
0.5	122	0.12	1.00	146.34	1.00	0.95	98.81
0.7	122	0.20	0.99	106.21	0.98	0.93	99.21
0.9	82	0.19	0.99	89.02	0.98	0.91	98.42
Pollution							
0.1	230	0.05	1.00	51.16	0.99	0.97	100.0
0.3	225	0.07	0.99	48.68	0.98	0.96	98.34
0.5	196	0.08	0.99	49.82	0.99	0.96	100.0
0.7	153	0.09	0.99	46.64	0.99	0.96	100.0
0.9	109	0.12	0.99	43.18	0.98	0.93	100.0
Quake							
0.1	153	0.09	0.92	469.73	0.88	0.61	74.52
0.3	147	0.09	0.92	566.56	0.89	0.58	81.46
0.5	175	0.08	0.95	548.45	0.93	0.60	83.52
0.7	148	0.09	0.95	731.85	0.93	0.53	77.65
0.9	145	0.1	0.95	534.05	0.92	0.63	78.06
Stock Price							
0.1	104	0.08	0.94	52.6	0.94	0.91	64.11
0.3	145	0.08	0.96	180.44	0.95	0.92	61.27
0.5	135	0.11	0.95	36.46	0.94	0.91	79.37
0.7	133	0.11	0.94	61.44	0.93	0.91	88.53
0.9	116	0.08	0.94	73.49	0.93	0.91	95.06

Tablo 6. Genliğe göre analiz

(Analysis based on amplitude)

<i>Genlik</i>	<i>Kural sayısı</i>	<i>Ort. Des</i>	<i>Ort. Güven</i>	<i>Ort. Lift</i>	<i>Ort. KF</i>	<i>Ort. netconf</i>	<i>Kapsanan kayıt</i>
Basketball							
2	190	0.03	0.99	80.09	0.98	0.95	95.84
5	337	0.02	1.00	92.02	1.00	0.99	98.96
10	362	0.02	1.00	94.67	1.00	1.00	100.0
15	375	0.02	1.00	95.39	1.00	1.00	100.0
20	394	0.02	1.00	95.0	1.00	1.00	100.0
Bodyfat							
2	129	0.10	1.00	151.87	1.00	0.99	99.61
5	88	0.02	1.00	22.48	1.00	1.00	58.34
10	124	0.01	1.00	245.71	1.00	1.00	35.32
15	126	0.01	1.00	245.6	1.00	1.00	34.53
20	123	0.01	1.00	245.68	1.00	1.00	33.34
Pollution							
2	230	0.05	1.0	51.16	0.99	0.97	100.0
5	302	0.03	1.0	57.68	1.0	0.99	100.0
10	321	0.02	1.0	59.34	1.0	1.0	95.0
15	309	0.02	1.0	59.07	1.0	1.0	100.0
20	296	0.02	1.0	59.13	1.0	1.0	95.0
Quake							
2	153	0.09	0.92	469.73	0.88	0.61	74.52
5	131	0.03	0.94	1063.33	0.93	0.43	47.89
10	144	0.01	0.97	1561.44	0.97	0.25	32.97
15	159	0.01	0.97	1547.72	0.96	0.24	23.56
20	174	0.01	0.99	1758.92	0.99	0.17	22.09
Stock Price							
2	104	0.08	0.94	52.6	0.94	0.91	64.11
5	65	0.03	0.92	205.14	0.92	0.91	38.74
10	37	0.02	0.92	240.98	0.92	0.9	25.16
15	31	0.01	0.91	373.45	0.91	0.89	8.32
20	24	0.01	0.89	518.94	0.89	0.88	8.43

Tablo 7. Eşik değerine göre analiz
(Analysis based on threshold value)

Eşik değeri	Kural sayısı	Ort. Des	Ort. Güven	Ort. Lift	Ort. KF	Ort. netconf	Kapsanan kayıt
Basketball							
5	190	0.03	0.99	80.09	0.98	0.95	95.84
10	200	0.03	0.99	83.05	0.98	0.96	97.92
25	175	0.04	0.99	80.95	0.98	0.96	92.71
50	135	0.04	0.97	72.72	0.96	0.93	93.75
100	62	0.07	0.93	52.92	0.90	0.84	91.67
Bodyfat							
5	129	0.10	1.00	151.87	1.00	0.99	99.61
10	140	0.08	1.00	171.21	1.00	0.99	99.21
25	109	0.11	0.99	151.17	0.99	0.96	99.21
50	62	0.10	0.96	115.4	0.96	0.92	100.0
100	56	0.11	0.95	76.73	0.94	0.91	98.42
Pollution							
5	230	0.05	1.00	51.16	0.99	0.97	100.00
10	219	0.05	1.00	53.38	0.99	0.98	100.00
25	205	0.07	0.99	51.2	0.99	0.97	100.00
50	149	0.07	0.99	49.49	0.98	0.94	100.00
100	37	0.12	0.96	31.09	0.93	0.88	98.34
Quake							
5	153	0.09	0.92	469.73	0.88	0.61	74.52
10	126	0.10	0.90	524.74	0.85	0.55	72.41
25	103	0.05	0.93	767.97	0.91	0.52	72.09
50	38	0.07	0.82	291.34	0.77	0.63	74.75
100	38	0.06	0.85	306.17	0.81	0.61	71.54
Stock Price							
5	104	0.08	0.94	52.60	0.94	0.91	64.11
10	101	0.08	0.95	51.11	0.94	0.92	71.48
25	128	0.07	0.95	171.18	0.95	0.92	89.27
50	81	0.07	0.93	88.79	0.92	0.89	92.32
100	61	0.08	0.93	34.16	0.93	0.89	92.11

5. SONUÇ (CONCLUSION)

Her ne kadar ikili ya da kesikli değerlere sahip veri tabanlarında etkili olarak çalışabilen birliktelik kural madenciliği çalışması olsa da, nicel değerler alabilen verilerde birliktelik kurallarının keşfi zor bir problemdir ve başarılı sonuçlar veren çalışma sayısı azdır. Yakın zamanda, nicel birliktelik kural madenciliğini, farklı kriterleri aynı anda en iyi şekilde karşılayacak şekilde, çok amaçlı bir problem olarak ele alan birkaç çalışma önerilmiştir. Bu makalede de nicel birliktelik kural madenciliği problemi için anlaşılabilirlik, ilginçlik ve performansı en üst düzeye çıkarmayı amaçlayan çok amaçlı evrimsel algoritmalarından baskın olmayan sıralama genetik algoritma-II temelli QAR-CIP-NSGA-II'nin nicel beş gerçek dünya verisinde değerlendirme sayısı, popülasyon sayısı, mutasyon olasılığı, genlik ve eşik değeri gibi parametrelerinin; elde edilen kural sayısı, ortalama destek, güven, lift, kesinlik faktörü, netconf ve kapsanan kayıt sayısına etkisi bildiğimiz kadarıyla ilk kez incelenmiştir.

QAR-CIP-NSGA-II'de değerlendirme sayısı ve popülasyon sayısı artışının tüm veri setlerinde kural sayısı, ortalama güven değeri, ortalama lift, ortalama KF, ortalama netconf değerlerini arttırdığı gözlemlenmiştir. Mutasyon olasılık değerinin 0.1 olduğu durumlarda genel olarak küçük boyutlu veri setlerinde en yüksek ortalama güven değerlerinin elde edildiği görülmüştür. Nispeten daha büyük boyutlu veri

setlerinde mutasyon olasılığının artışı kurallardaki ortalama güven değerinde az da olsa bir artış göstermiştir. Genlik değerinin değişiminin ortalama güven değerlerine pek bir etkisinin olmadığı görülmüştür. Eşik değeri arttıkça, veri setinden elde edilen kural sayısı ve kuralların ve ortalama güven ve ortalama kesinlik faktörü azalmıştır.

İleride nicel birliktelik kural madenciliği problemi için etkili sonuçlar alabilmek amacıyla, en güncel akıllı arama ve optimizasyon algoritmalarının çok amaçlı hibrit versiyonlarının önerilmesi ve farklı alanlarda uygulamalarının gerçekleştirilmesi hedeflenmektedir.

KAYNAKLAR (REFERENCES)

- [1] K. Deb, S. Agrawal, A. Pratap, T. Meyarivan, "A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II", **In International Conference on Parallel Problem Solving From Nature**, Springer, Berlin, Heidelberg, 849-858, 2000.
- [2] D. Martín, A. Rosete, J. Alcalá-Fdez, F. Herrera, "QAR-CIP-NSGA-II: A new multi-objective evolutionary algorithm to mine quantitative association rules", *Information Sciences*, 258, 1-28, 2004.
- [3] B. Oğuz Yolcular, U. Bilge, M. K. Samur, "Kulak burun boğaz taburcu notlarından birliktelik kurallarının çıkartılması", *Bilişim Teknolojileri Dergisi*, 11(1), 35-42, 2018.

- [4] S. Ramaswamy, S. Mahajan, A. Silberschatz, "On the discovery of interesting patterns in association rules", In: **24rd International Conference on Very Large Data Bases**, San Francisco, CA, USA, 1998.
- [5] E. Shortliffe, B. Buchanan, "A model of inexact reasoning in medicine", *Mathematical Biosciences*, 23(3-4), 351-379, 1975.
- [6] K. I. Ahn, J. Y. Kim, "Efficient mining of frequent itemsets and a measure of interest for association rule mining", *Journal of Information & Knowledge Management*, 3(3), 245-257, 2004.
- [7] R. Srikant, R. Agrawal, "Mining quantitative association rules in large relational tables", In: *Proceedings of ACM SIGMOD*, 1-12, 1996.
- [8] H. P. Chiu, Y. T. Tang, K. L. Hsieh, "A cluster-based method for mining generalized fuzzy association rules", In **First International Conference on Innovative Computing, Information and Control**, 2, 519-522, 2006.
- [9] B. Alatas, E. Akin, A. Karci, "MODENAR: Multi-objective differential evolution algorithm for mining numeric association rules", *Applied Soft Computing*, 8, 646-656, 2008.
- [10] C. H. Chen, T. P. Hong, V. S. Tseng, L. C. Chen, "A multi-objective genetic-fuzzy mining algorithm", In **IEEE International Conference on Granular Computing**, IEEE, GrC 2008, 115-120, 2008.
- [11] X. Yan, C. Zhang, S. Zhang S, "Genetic algorithm-based strategy for identifying association rules without specifying actual minimum support", *Expert Systems with Applications*, 36(2), 3066-3076, 2009.
- [12] V. Beiranvand, M. M. Kashani, A. A. Bakar, "Multi-Objective PSO algorithm for mining numerical association rules without a priori discretization", *Expert Systems with Application*, 41, 4259-4273, 2014.
- [13] D. Martin, A. Rosete, A. J. Fdez, F. Herrera, "QAR-CIP-NSGA-II: A new multi-objective evolutionary algorithm to mine quantitative association rules", *Information Sciences*, 258, 1-28, 2014.
- [14] J. Piri, R. Dey, "Quantitative association rule mining using multi-objective particle swarm optimization", *International Journal of Scientific & Engineering Research*, 5(10), 155-161, 2014.
- [15] M. Almasi, M. S. Abadeh, "Rare-PEARs: A new multi objective evolutionary algorithm to mine rare and non-redundant quantitative association rules", *Knowledge-Based Systems*, 89, 366-384, 2015.
- [16] I. Kahvazadeh, M. S. Abadeh, "MOCANAR: A multi-objective cuckoo search algorithm for numeric association rule discovery", *Computer Science & Information Technology*, 99-113, 2015.
- [17] M. Martínez-Ballesteros, A. Troncoso, F. Martínez-Álvarez, J. C. Riquelme, "Improving a multi-objective evolutionary algorithm to discover quantitative association rules", *Knowledge and Information Systems*, 49(2), 481-509, 2016.
- [18] R. J. Kuo, M. Gosumolo, F. E. Zulvia, "Multi-objective particle swarm optimization algorithm using adaptive archive grid for numerical association rule mining", *Neural Computing and Applications*, 1-14, 2017.
- [19] D. Yan, X. Zhao, R. Lin, D. Bai, "PPQAR: Parallel PSO for quantitative association rule mining", In **IEEE International Conference on Big Data and Smart Computing (BigComp)**, 163-169, 2018.
- [20] I. E. Agbehadji, S. Fong, R. Millham, "Wolf search algorithm for numeric association rule mining", In **IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)**, 146-151, 2016.
- [21] B. Badhon, M. M. J. Kabir, S. Xu, M. Kabir, "A survey on association rule mining based on evolutionary algorithms", *International Journal of Computers and Applications*, 1-11, 2019.
- [22] E. V. Altay, B. Alatas, "Performance analysis of multi-objective artificial intelligence optimization algorithms in numerical association rule mining", *Journal of Ambient Intelligence and Humanized Computing*, 1-21, 2019.
- [23] E. V. Altay, B. Alatas, "Intelligent optimization algorithms for the problem of mining numerical association rules", *Physica A: Statistical Mechanics and its Applications*, 540, 123142, 2020.
- [24] T. Zhang, M. Shi, J. Wang, G. Yang, "P-EAARM: A generic framework based on spark for eas-based association rule mining", In **IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA)**, IEEE, 99-104, 2019.
- [25] Internet: H. A. Guvenir, I. Uysal, Bilkent university function approximation repository, <http://funapp.cs.bilkent.edu.tr/DataSets>, 15.12.2018.