Journal of the Faculty of Engineering and Architecture of Gazi University 39:1 (2024) 563-575

Mühendislik Mimarlık Fakültesi Dergisi
Journal of The Faculty of Engineering and Architecture of Gazi University
Elektronik / Online ISSN : 1304 - 4915
Basılı / Printed ISSN : 1300 - 1884

Point of interest coverage with distributed multi-unmanned aerial vehicles based on distributed reinforcement learning

Fatih Aydemir[1,2]*, Aydın Çetin[2]
[1]STM Defence Technologies Engineering and Trade. Inc., Ankara, 06560, Türkiye
[2]Department of Computer Engineering, Faculty of Technology, Institute of Science, Gazi University, 06500, Ankara, Türkiye

Highlights:

- Dynamic area coverage with unmanned aerial vehicles
- Positioning with high fairness index
- Energy efficient movement model

Keywords:

- Unmanned Aerial Vehicle
- Multi-agent system
- Reinforcement learning
- Dynamic area coverage
- Grid Decomposition

Article Info:

Research Article
Received: 07.09.2022
Accepted: 19.03.2023

DOI:

10.17341/gazimmfd.1172120

Correspondence:

Author: Fatih Aydemir
e-mail: faydemir@yandex.com
phone: +90 544 585 3105

Graphical/Tabular Abstract

In the study, a group of unmanned aerial vehicle was modelled as a multi-agent reinforcement learning system, which can adapt to dynamic environment to cover point of interests. Agents create an abstract rectangular plane containing the area to be covered, and then decompose the area into grids. An agent learns to locate on a center of grid that are closest to it, which has the largest number of POIs to plan its path. This planning helps to achieve a high fairness index by reducing the number of common covered POIs. In order to maximize covered area with minimum energy consumption under connectivity constraints between agents, three reward strategies were designed with collaborative approach. The proposed model archived better results than previous studies. The architecture of the proposed system is shown in Figure A.

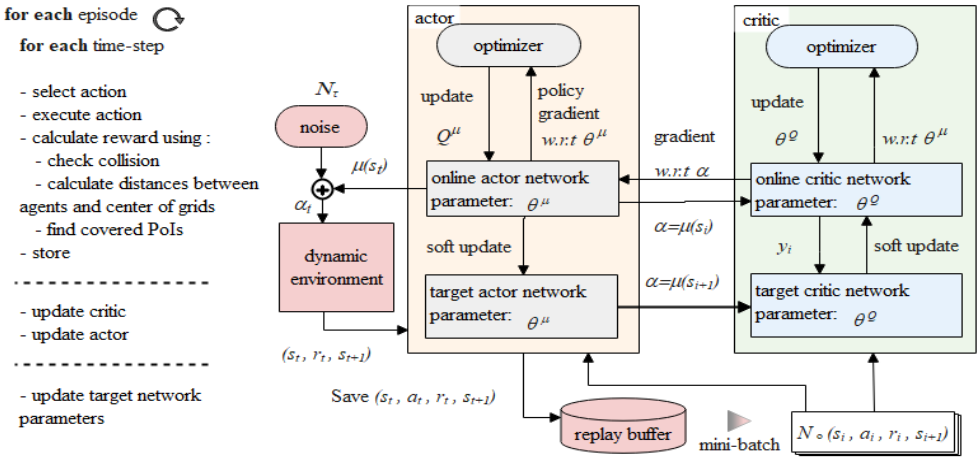


Figure A. The architecture of the proposed system

Purpose: The aim of the study is to design a distributed control solution to place a group of UAVs in the dynamic area in order to maximize covered point of interests with minimum energy consumption and a high fairness value.

Theory and Methods: The proposed method is modeled on the basis of multi-agent deep reinforcement learning so that agents can produce behavioral strategies in dynamic environments. The central learning-decentralized execution scheme was preferred to obtain a system which acts independently of each other but produces policies for the common goal of the group. In addition, a collaborative reward policy, which is shared in common, has been achieved with the reward strategy designed. The collaborative reward structure depends on the appropriate positioning of all agents, and thus has a collective nature of its own.

Results: In experimental studies, the proposed method was compared with similar research in literature using three metrics: (i) energy consumption, (ii) coverage, (iii) fairness index. According to the comparison, the proposed method achieved better results than the other models.

Conclusion: The proposed model which is based on local observations independent of central control helps to solve the area coverage problems with multi-agent systems in real applications.

Journal of the Faculty of Engineering and Architecture of Gazi University 39:1 (2024) 563-575

Mühendislik Mimarlık Fakültesi Dergisi
Journal of The Faculty of Engineering and Architecture of Gazi University

Elektronik / Online ISSN : 1304 - 4915
Basılı / Printed ISSN : 1300 - 1884

Dağıtık pekiştirmeli öğrenme tabanlı çoklu insansız hava aracı ile ilgi çekici nokta kapsama

Fatih Aydemir[1,2]*, Aydın Çetin[2]

[1]STM Savunma Teknolojileri ve Mühendislik A.Ş., Ankara, 06560, Türkiye
[2]Gazi Üniversitesi, Fen Bilimleri Enstitüsü, Teknoloji Fakültesi Bilgisayar Mühendisliği Anabilim Dalı, Ankara, 06500, Türkiye

ÖNEÇIKANLAR

- İnsansız hava araçlarıyla dinamik alan kapsama
- Yüksek adillik indisine sahip konumlanma
- Enerji verimli hareket modeli

Makale Bilgileri

Araştırma Makalesi
Geliş: 07.09.2022
Kabul: 19.03.2023

DOI:
10.17341/gazimmfd.1172120

Anahtar Kelimeler:
İnsansız hava aracı, çok ajanlı sistem, pekiştirmeli öğrenme, dinamik alan kapsama, ızgara ayrıştırma

ÖZ

Mobil araçlar haritalama, trafiğin izlenmesi, arama-kurtarma operasyonları gibi çeşitli alan kapsama uygulamalarında yaygın olarak kullanılmaktadır. Kapsama sürecini geliştirmek için uygun konumlandırma modeli ve etkili öğrenme stratejisi gereklidir. Mobil araçlar hareket modeli içeren yönlendirme mekanizması ile dinamik ortamlara uyum sağlayabilir ve en uygun konumları bulabilirler. Konumlandırma sürecinin çok ajanlı mobil sistem temelinde yönetildiği çalışmalarda algılama, veri toplama ve gözetim gibi görevleri birden fazla ajanın işbirlikçi yaklaşım ile tamamlaması gerekir. Öğrenmeye dayalı bu süreç, bir görevi gerçek zamanlı optimize etmeyi öğrenebilen mobil ajanlar vasıtasıyla yürütülebilir. Bu çalışmada, bir grup insansız hava aracının (İHA) öğrenebilen çok ajanlı sistem temelinde modellenerek dinamik ortamda ilgi çekici noktaları (İÇN) etkin şekilde kapsaması hedeflenmektedir. Hedef alan, İÇN kapsamını en üst düzeye çıkarmak ve enerji tüketimini en aza indirmek için ızgaralara ayrıştırılır. Ayrıştırma, hedef alanın konumu ve mobil ajan olarak modellenen İHA'ların iletişim mesafesi göz önünde bulundurularak gerçekleştirilir. Bununla birlikte ızgaralara gidiş planlanması yapan mobil ajanlar çarpışmadan kaçınmayı da öğrenirler. Önerilen yöntem benzetim ortamında test edilmiş ve sonuçlar benzer çalışmalar ile kıyaslanarak sunulmuştur. Sonuçlar, önerilen yöntemin mevcut benzer çalışmalara göre daha iyi performans gösterdiğini ve alan kapsama uygulamaları için uygun olduğunu göstermektedir.

Point of interest coverage with distributed multi-unmanned aerial vehicles based on distributed reinforcement learning

HIGHLIGHTS

- Dynamic area coverage with unmanned aerial vehicles
- Positioning with high fairness index
- Energy efficient movement model

Article Info

Research Article
Received: 07.09.2022
Accepted: 19.03.2023

DOI:
10.17341/gazimmfd.1172120

Keywords:
Unmanned aerial vehicle, multi-agent system, reinforcement learning, dynamic area coverage, grid decomposition

ABSTRACT

Mobile vehicles are widely used in various area coverage applications such as mapping, traffic monitoring, search and rescue operations. Appropriate positioning model and effective learning strategy are required to improve the coverage process. Mobile vehicles can adapt to dynamic environments and find the optimum locations with the navigation mechanism that includes a motion model. In studies, where the positioning process is managed on the basis of a multi-agent mobile system, multiple agents should complete tasks such as detection, data collection and surveillance with a collaborative approach. This learning-based process can be carried out through mobile agents that can learn to optimize a task in real time. In this study, it is aimed to effectively cover points of interest (PoI) in a dynamic environment by modeling a group of unmanned aerial vehicles (UAVs) on the basis of a learning multi-agent system. The target area is decomposed into grids to maximize PoI coverage and minimize energy consumption. Decomposition is performed by considering the location of the target area and the communication distance of the UAVs modeled as mobile agents. However, mobile agents planning to go to grids also learn to avoid collisions. The proposed method has been tested in a simulation environment and the results are presented by comparing with similar studies. The results show that the proposed method outperforms existing similar studies and is suitable for area coverage applications.

*Sorumlu Yazar/Yazarlar / Corresponding Author/Authors : *faydemir@yandex.com, acetin@gazi.edu.tr / Tel: +90 544 585 3105

1. Giriş (Introduction)

İnsansız hava aracı (İHA), üzerinde pilot bulunmayan, otonom veya uzaktan kumanda ile yönetilen bir hava aracıdır. Robotik teknolojilerindeki ilerleme, belirli bir ihtiyaca yönelik İHA'ların üretimini kolaylaştırmıştır. Bu kolaylık, esneklik ve çeviklik, faydalı yük, algılama kabiliyeti, düşük maliyet gibi birçok gelişmeyi beraberinde getirmiştir. Kullanılan teknolojinin evrimi hem askeri hem de sivil alanlarda İHA kullanım talebinin artmasına sebep olmuştur.

Alan gözetleme [1], harita oluşturma ve keşif [2], arama ve kurtarma [3] gibi birçok çalışma dinamik alan kapsama temelinde yapılmıştır. İlgilenilen alanın ne kadarının izlendiği, sürecin hangi hızda yönetildiği ve hata toleransının ne kadar yüksek olduğu gibi konular kapsam kalitesini belirler. Mobil olmayan duyargaların kullanıldığı uygulamalarda duyargalar, ilk kurulumdan sonra sabit kalır. Bu durum, ilgilenilen alanda değişikliğe karşı adaptasyonu zorlaştırır ve duyargaların hata toleransını düşürür. Dinamik alan kapsama sürecinde, düşük maliyet-yüksek esneklik elde edilebilmek için duyargalar ile donatılmış mobil araçlar yani İHA'lar kullanılabilir [4].

Kapsama yapılacak alanda tek mobil aracın kullanılabilmesi için aracın karmaşık yapı ve kontrol modülleri ile tasarlanmış olması gerekir. Bu yapı, mobil ajanın herhangi bir alt bileşeninde meydana gelecek bir problemin tüm sistemi etkilemesine sebep olabilir. Aynı yetenek, bir grup basit mobil aracın iş birliği ile de elde edilebilir. Ek olarak, basit araçların kullanımı bakım maliyetini düşürür. Problemlerin büyüklüğü ve karmaşıklığı sebebiyle bir grup mobil araç; birden fazla sistemin koordineli, etkileşimli, bağımlı veya bağımsız olarak yürüttükleri süreçler ile çok ajanlı sistem olarak ele alınabilir [5]. Bu sistemler, N adet ajan ile "kolektif zekâ" olarak adlandırılan ortak bir amaca yönelik davranışı geliştiren ajanlar yığını olarak tanımlanabilir [6]. Duyarga düğümleri kullanılarak en üst seviyede ilgi çekici nokta (İÇN) kapsama ve alan gözetleme gibi kapsama temelli çalışmalar yapılmıştır. Gupta vd. [7], tüm hedef noktaları k-kapsama ve duyarga düğümleri m-bağlı olacak şekilde genetik algoritma tabanlı bir yaklaşım ile duyarga düğüm konumlandırma çalışması yapmıştır. Njoya vd. [8], [7]'deki gibi hedef kapsama problemini ele almıştır. Duyargalar arası bağlantıyı göz önünde bulundurarak duyarga dağıtımını amaçlayan melez bir yaklaşım ile hedef noktaları kapsamayı amaçlamıştır. Yu vd. [9], bağlantı ve kapsama alanını artırmak için geliştirilmiş yapay arı kolonisi (YAK) algoritması önermiştir. Çalışma içerisinde önerilen algoritma, rastgele dağılım ve genetik algoritma (GA) ile karşılaştırılarak sonuçları sunulmuştur. Bu sonuçlara göre önerilen algoritma daha uzun süreli iletişim ve daha yüksek kapsama oranına sahiptir. [10]'da araştırmacılar, mobil duyargaların en az sayıda hareketi ile ağaç modeli ve karınca arı kolonisine (KAK) dayalı yeni bir algoritma önermiştir. Önerilen algoritma mevcut benzer çalışmalarda elde edilen kapsama oranını daha az hareket ve daha az enerji tüketimi ile elde etmiştir. Shu vd. [11], kapsanan alanı artırmak ve kablosuz duyarga ağların (KDA) ömrünü uzatmak için yeni bir yaklaşım önermiştir. Araştırmacılar, geometrik merkez ve Voronoi diyagramı ile mobil duyarga düğümlerinin hareket alanını kısıtlamıştır. Bu kısıt, düğümlerin birbirleri arasındaki bağlantının korunmasına olanak sağlamıştır. Shi vd. [12], küçük İHA'ların en uygun noktalara konumlandırılması ile kullanıcı kapsamını en üst düzeye çıkarmaya çalışmıştır. 3 boyutlu coğrafik düzlemde, en az enerji-en fazla kapsama elde edilebilmesi için daire paketleme teorisinden (DPT) faydalanılmıştır [13]. Zhang ve Duan [14], İHA ağlarının alan üzerinde konumlandırılması sırasında geçen süreyi düşürmeye çalışmıştır. Ayrıca araştırmacılar, doğrusal yaklaşım ve k-ortalama kümeleme yöntemlerini temel alan bir yöntem geliştirerek en fazla sayıda kullanıcı kapsamını da elde etmeye çalışmıştır [15]. Cabreira vd. [16], düşük enerji tüketen İHA'ların ızgaralara ayrıştırılmış alanda hareketi için bir model önermiş ve bu model yardımıyla düzensiz şekilli alanlarda kapsama planlama problemini çözmeyi amaçlamıştır. Kalantari vd. [17], sistem gereksinim ve kısıtlamalarını göz önünde bulundurarak ihtiyaç duyulan İHA baz istasyon sayısını bulmaya çalışmıştır. Baz istasyonlarını 3 boyutlu bir düzlemde konumlandırabilmek için parçacık sürü optimizasyonu (PSO) temelinde sezgisel bir algoritma önermiştir. Chiu vd. [18], bağlanabilir cihazlarda sırt çantası problemini çözemeye çalışmıştır. Önerilen yöntemde, ajanlar arası çatışmayı en fazla kar atamasıyla ortadan kaldıran çok ajanlı İHA sistemi tasarlanmıştır. Bu çok ajanlı İHA sistemi, bir yörünge etrafında hareket ederek en az enerji tüketimi ile en fazla bilgi toplamayı amaçlamaktadır. Ganganath vd. [19] engelli ve engelsiz ortamda mobil duyarga ağlarının dinamik kapsama yapabilmesi için antifloklamayı temel alan 2 farklı yöntem üzerinde çalışmıştır. Krajnik vd. [20], ajan konumlarını belirlemek için yerleşik kameralar yardımıyla örüntü algılamaya dayanan pekiştirmeli öğrenme (PÖ) algoritması önermiştir. Başka bir çalışmada ise en az enerji tüketimi ile şebeke kapsamını ve ömrünü artırmak için çok-nesneli bölgesel keskin koku işaretleme algoritması sunulmuştur [21]. Liu vd. [22], İHA konumlandırma ve yönlendirme sürecini daha az enerji ile yapabilmek amacıyla DRL-EC3 adında bir öğrenme yöntemi önermiştir. Bu yöntem derin deterministik politika gradyanı (DDPG) [23] algoritmasını temel alarak tasarlanmıştır. Sistem içerisindeki İHA'lar tek bir aktör-eleştirmen ağı ile öğrenme gerçekleştirir. Yazarlar daha sonra DRL-EC3 modelini geliştirerek, sistemin ortam değişikliklerine karşı dayanıklılığını artırmışlardır [24]. Geliştirilmiş bu modelde, her İHA kendi aktör-eleştirmen ağına sahiptir ve en iyi eyleme bireysel olarak karar verebilir. İHA'ların dağıtık olarak çalışmasına izin verilen bu yöntemde, hedef alanda İHA'lar arasındaki bağlantının kurulması da amaçlanmıştır. Aydemir ve Çetin [25], çoklu İHA kullanarak dinamik ortamda alan kapsamayı en yükseğe çıkarmaya çalışmıştır. Derin pekiştirmeli öğrenme ile modellenmiş ajanlar merkezi bir modül ile öğrenme gerçekleştirir; ancak yürütme aşamasında merkezi modül devreden çıkarılır. Dağıtık sistem mimarisine sahip bu çalışma, ortamdaki aktif ajanların bir graf oluşturarak hedef alanda en uygun noktalara konumlanmasını amaçlamıştır.

Önerilen bu yöntemlerde mobil ajanlar hedef alan üzerinde yayılarak kapsama sürecini iyileştirmeye odaklanır. Aralarındaki temel fark konumlandırma stratejisinin nasıl tasarlandığıdır. Bazı sistemler tüm ajanları merkezi bir denetleyici yardımıyla tek bir noktadan yönlendirmeyi temel alır. Bu tür sistemlerde merkezi denetleyicide oluşabilecek hatalar tüm sistemi doğrudan etkiler. Sezgisel yöntemler, dinamik alan kapsama problemlerini optimizasyon temelinde ele alabilir. Sezgisel tabanlı sistemlerde ajanlar sadece kendi eylemlerini inceler, yani gruptaki diğer ajanların eylemlerini optimizasyon sürecine dahil etmezler. Dolayısıyla bu tür sistemlerde en uygun sonuca yakınsama uzun sürebilir. Antifloklama kullanılan yöntemlerde ajanlar genellikle kapsama sürecinin tamamını değil sadece anlık durumu değerlendirir. Böylelikle kapsama sürecinin en uygun sonucundan ziyade anlık durumun en uygun sonucu elde edilir. Yürütme sırasında oluşabilecek ortam değişiklikleri tüm sistem yapısının değişmesine sebep olur. Bu sebeple mobil araçların bulundukları ortama uyum sağlayabilmeleri büyük önem arz etmektedir. Dolayısıyla dinamik ve/veya bilinmeyen alanlarda konumlandırma ancak öğrenme yoluyla gerçekleştirilebilir. Bu çalışmada dinamik kapsama için her bir İHA'nın ajan olarak temsil edildiği çok ajanlı derin pekiştirmeli öğrenme temelli bir yöntem önerilmiştir. Pekiştirmeli öğrenme, amaca yönelik ne yapılması gerektiğini öğrenen ödül güdümlü makine öğrenmesi yaklaşımıdır. Bir pekiştirmeli öğrenme ajanı, çevresi ile etkileşime girer ve eylemlerinin karşılığında pozitif veya negatif ödül puanı alır. Ajan, aldığı ödül puanını maksimuma çıkarmak için davranışını

değiştirmeyi öğrenir. Klasik pekiştirmeli öğrenme yöntemleri bireysel öğrenme sürecidir. Çok ajanlı pekiştirmeli öğrenmede ise birbirinden bağımsız olarak çalışan ajanların eylemleri grubun ortak hedefi için değerlendirilir. Dinamik alan kapsama problemi için oluşturulmuş kolektif davranış ajan grubu için değerlendirildiğinde en yüksek seviyede kapsama sağlayan bir sistemi ifade eder. Bu davranış bir ajan için değerlendirildiğinde ise, sınırlı seviyedeki iletişim yetenekleri ile kendisini konumlandıran bir alt sistemi temsil eder. Önerilen yöntemde ajan grubunun öğrenme-yürütme süreci, merkezi öğrenme-merkezi olmayan yürütme yöntemlerinden birisi olan çok ajanlı derin deterministik politika gradyanı (ÇADDPG) [26] algoritması temel alınarak yapılmıştır. ÇADDPG algoritmasında her ajan, aktörün yerel gözlemlere erişiminin olduğu bir DDPG algoritması tarafından eğitilir. Güncellenen ÇADDPG ile çok sayıda PÖ ajanının anlık durumları ve politikaları göz önünde bulundurulmuş, aktör ve eleştirmen ağları ile kolektif eylemler oluşturulmuştur. Ajanların homojen yapıda olduğu ve daire şeklinde kapsama yaptığı varsayılmıştır. Önerilen yöntemde ajanların, çarpışmadan kaçınarak en kısa sürede en uygun konumlara yönelmesiyle en fazla İÇN kapsama hedeflenmiştir. ÇADDPG algoritmasında ödül işlevi olarak ajanların ödül toplamları, grubun ortalama ödül puanı gibi tüm grubu temel alan yöntemler kullanılır. Bu yaklaşım, ajan bazlı en iyi eylem ile en kötü eylem farkındalığının azalmasına sebep olur. Bu çalışmada önerilen yöntem içerisinde kullanılan ödül işlevi, bağlı ajanların toplam kapsadığı İÇN temelinde tasarlanmıştır. Böylelikle, ajanlar arasındaki mesafe etkileşimde bulunabilecekleri seviyede artırılarak, ortak kapsanan İÇN sayısını en aza indirilir. Buna ek olarak ızgaralara ajan atama yaklaşımı kullanılarak en fazla sayıda İÇN olan bölgeye en hızlı gidiş için yapılan yol planlaması ile enerji tasarrufu sağlanır. Ajanlar, kapsama yapılacak alanın $x_{min}, y_{min}, x_{max}$ ve y_{max} değerlerini kullanarak soyut dörtgen düzlem oluşturur ve sonrasında bu alanı ızgaralara ayırır. Ajanlar kendilerine en yakın olan ve en çok İÇN sayısına sahip ızgaralara konumlanmayı öğrenerek yol planlaması yapar. Bu planlama, ortak kapsanan İÇN sayısını düşürerek yüksek adillik indisi elde etmeye yardımcı olur. Bunun yanında, hedef alanda ajanlar arası bağlantının sağlanması da amaçlanır. Yol planlaması ile enerji ve zaman tasarrufu sağlanırken, ızgara ayrıştırma yöntemi ile düzensiz şekilli alanlar alan kapsama sürecine dahil edilir.

Çalışmanın kalan kısmı şu şekilde düzenlenmiştir. Bölüm 2'de, önerilen yöntem hakkında detaylı bilgi verilerek, dinamik alan kapsama problemleri için çok ajanlı sistem (ÇAS) modeli sunulmuştur. Bölüm 3'te, deneysel çalışmalar ve benzetim sonuçlarına yer verilmiştir. Bölüm 4'te, sonuçlar analiz edilmiş ve sonraki çalışmalar için yol haritası çizilmiştir.

2. Materyal-Metot (Material-Method)

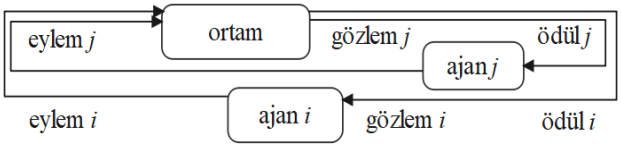
Hedef alanda etkili kapsama için İHA'lar bulundukları ortamlara uygun davranışlar sergilemelidir. Bununla birlikte, hedef alanda İHA'lar arasındaki bağlantıyı korumak ve enerji tasarrufu sağlamak için İHA hareketleri azaltılmalıdır. Bu minvalde yapılan bu çalışmada, bir ajan grubundaki her bir İHA aynı hareket modeline sahip bir ajan olarak modellenmiştir. Birden fazla ajanın ortaklaşa çalışması ile aşağıdaki hedeflerin karşılanması amaçlanmıştır:

- Kendi kendine öğrenebilen ajanların oluşturduğu dağıtık yapıda sistem inşa etmek;
- Yol planlaması ve hedef ızgara ataması ile en fazla sayıda İÇN kapsamak;
- Ajan eylemlerinden kaynaklanan enerji tüketimini en aza indirmek;
- Hedef bölge içerisinde kalarak ajanlar arasındaki bağlantıyı sağlamak,
- Ajan eylemlerini optimize etmek ve öğrenme yoluyla ajanlar arasındaki çarpışmaları önlemek;
- Hata toleransı yüksek bir sistem ile bilinmeyen dinamik ortamlarda görev devamlılığı sağlamak.

Ajanlar öğrenme yetenekleri sayesinde hedef alandaki değişikliklere tepki verebilir ve kolektif davranışlar ile başarı elde edebilir. Ayrıca dağıtık sistem mimarisine uygun tasarım ile de merkezi bir yönlendirmeden bağımsız karar alabilmeleri sağlanabilir. Tüm bunlar göz önünde bulundurulduğunda, önerilen yöntem, merkezi kontrolden bağımsız kolektif başarı üretmeye çalışan PÖ mobil ajanları temelinde modellenmiştir.

2.1. Pekiştirmeli Öğrenme (Reinforcement Learning)

PÖ'de, bir ajan bulunduğu ortamla etkileşime geçer ve bulunduğu ortama göre davranışını optimize eder. Şekil 1'de görüldüğü üzere ajan bulunduğu bir durumdan başka bir duruma geçiş için eylemde bulunur ve ortamdan eylemin kalitesinin gösteren bir ödül alır. Bu süreç Markov Karar Süreci (MKS) olarak adlandırılır [27]. Bir durumda yapılacak eyleme ise politika denir. Ajanın temel amacı, toplam ödülü en yüksek seviyeye çıkaran bir politika bulmaktır.



Şekil 1. PÖ ajanı (RL agent)

MKS esas olarak tek ajanlı sistemler için tasarlanmıştır. Klasik MKS'lerde, bir ajanın tüm sisteme doğrudan erişimi olduğu varsayılır. Kısmen gözlemlenebilir MKS'lerde (KGMKS), sistem, ajan tarafından yerel gözlemler ve eylem geçmişi aracılığıyla çalışır. Ajan birden fazla ise bu sürece çok ajanlı MKS (ÇAMKS) adı verilir. ÇAS'larda ajan grubu her zaman homojen değildir. Ajanlar farklı ödül fonksiyonlarına sahip olsa da grubun ödülünü en üst düzeye çıkarmak için çalışırlar. Ödül fonksiyonu her ajan için özel olabilir ve bu durumda ödül bilgisi diğer ajanlarla paylaşılamaz. Homojen olarak oluşturulan sistemlerde takım çalışmasının elde edilebilmesi daha fazla koordinasyon gerektirir. Denetleyici ödülleri toplayıp, değerlendirip, sonuçları tüm ajanlara dağıtabileceği için genellikle merkezi bir denetleyici ile çok ajanlı pekiştirmeli öğrenme (ÇAPÖ) algoritması kullanılır. Ancak maliyet, ölçeklenebilirlik veya sağlamlık gibi kaygılar nedeniyle böyle bir denetleyici tercih edilmeyebilir. Bu gibi sistemlerde ajanlar, kısa mesafeli ve zamanla değişebilen bir iletişim ağı üzerinden birbirleriyle bilgi paylaşırlar.

Merkezi olmayan KGMKS'ler, ÇAMKS ile işbirlikçi davranışlar elde etmek için kullanılır; ancak ÇAMKS'lerde her ajanın çevre ve sistem durumu için sadece yerel gözlemleri vardır. Ajan, diğer ajanların gözlemlerine erişemediğinden, bir ajanın tüm sistem adına karar verebilmesi mümkün değildir. Ödül fonksiyonu ve politika gibi unsurları Markov takım modeli olarak paylaşan ve merkezi olmayan bir yöntem ile yerel gözlemlerin paylaşıldığı bir sistem modellenebilir. Bu yönteme dağıtık kısmi gözlemlenebilir MKS (DKGMKS) denir. Ajanlar, diğer ajanların gözlemlerine erişemediği için bir ajan tüm takımın yerine karar verebilecek yeterli bilgiye sahip değildir. Bu nedenle, DKGMKS'lerde en kötü durumu çözmek için süper üssel zamana ihtiyaç duyulmaktadır.

2.2. Çok Ajanlı Derin Pekiştirmeli Öğrenme (Multi-agent Deep Reinforcement Learning)

[28]'ye göre derin öğrenme, farklı soyutlama düzeylerine karşılık gelen birden çok temsil düzeyini öğrenen bir makine öğrenme yaklaşımıdır. Basit bir durumda biri giriş sinyali alan, diğeri çıkış

sinyali gönderen iki nöron grubu içerir. Girdi katmanı bir girdi aldığında, girdinin işlenmiş halini bir sonraki katmana iletir. Derin bir ağda, giriş ve çıkış arasında çoklu doğrusal ve çoklu işlem katmanlarından oluşan birçok katman vardır [29]. Çok ajanlı derin PÖ (ÇADPÖ), karmaşık görevleri çözmek amacıyla rekabet eden veya iş birliği yapan PÖ ajanlarından oluşan çok ajanlı bir sistemdir. Tek ajanlı sistemlerde, ajan yalnızca kendi eylemlerinin sonucuyla ilgilenir. Çok ajanlı bir sistemde ise bir ajan yalnızca kendi eyleminin sonuçlarını değil, aynı zamanda diğer ajanların davranışlarını da inceler. Öğrenme süreci karmaşıktır; çünkü tüm ajanlar birbirleriyle etkileşime girerek aynı anda öğrenir. Bu sebeple DKGMKS ile modellenmiş sistemlerde ÇADPÖ ile iyi sonuçlar elde edilmesi mümkündür.

DKGMKS'lerde her ajanın yalnızca yerel gözlemlere doğrudan erişimi vardır. Bu gözlemler pek çok şey olabilir; çevrenin bir görüntüsü, yer işaretlerine göre konumlar ve hatta diğer ajanların göreceli konumları. Ayrıca öğrenme sırasında tüm ajanlar merkezi bir modül veya eleştirmen tarafından yönlendirilir. Her ajanın yalnızca yerel bilgileri ve yerel politikaları olmasına rağmen, tüm ajanların bilgilerini gözden geçiren ve onlara politikalarını nasıl güncelleyecekleri konusunda tavsiyelerde bulunan bir eleştirmen vardır. Ajanlar, öğrenme sürecinde, tüm sistem hakkında bilgi sahibi olan bir modülden yardım alırlar. Bu yaklaşıma merkezi öğrenme denir. Yürütme sırasında ise, merkezi modül kaldırılır ve geriye ajanlar, onların politikaları ve yerel gözlemleri kalır. Bu, artan durum ve eylem uzay etkilerini azaltır. Merkezi modülün yürütme zamanında en uygun yerel politika için yeterli bilgi vermesi amaçlanmaktadır.

2.3. Çok Ajanlı Derin Deterministik Politika Gradyanı (Multi-agent Deep Deterministic Policy Gradient)

DDPG, derin Q-öğrenmenin [31] temel başarısını sürekli eylem alanına uyarlayan bir aktör-eleştirmen yöntemidir. ÇADDPG algoritması, ajanların yalnızca yerel bilgilere erişebildiği ve diğer ajanlarla politikalarını paylaştığı aktör-eleştirmen politika gradyan yöntemlerinin bir uzantısı olarak önerilmiştir. Politika gradyanı (PG) yaklaşımındaki ana fikir, verilen gradyan yönünde adımlar atarak belirli bir hedefi en yükseğe çıkarmak için politika parametresi ayarlamaktır. Sistem içerisinde bir eleştirmen kullanmak, ortamın dinamik durumunu ele almak için yaygın bir çözümdür. Bu nedenle, bu merkezi eleştirmen, yerel gözlemlere sahip ajanların esnekliğini artırmak için güvenilir bir rehber olarak kullanılabilir. ÇADDPG'de her ajanın iki ağı vardır: bir aktör ağı ve bir eleştirmen ağ. Aktör ağı, ajanın bulunduğu duruma göre yürütülecek eylemi hesaplarken, eleştirmen ağ aktör ağının performansını iyileştirmek için eylemin sonuçlarını değerlendirir. Eleştirmen ağ güncellemesi için kullanılan deneyim tekrar arabelleği, eğitim verilerindeki korelasyonları kırmaya ve eğitimi daha kararlı hale getirmeye yardımcı olur. Eğitim aşamasında her ajan, aktörün yerel gözlemlere erişiminin olduğu bir DDPG algoritması tarafından eğitilir. Merkezileştirilmiş eleştirmen ise girdi olarak tüm durum-eylemleri birleştirir ve buna karşılık gelen Q-değerini elde etmek için yerel ödül fonksiyonunu kullanır. Yürütme aşamasında, eleştirmen ağ kaldırılır ve ajanlar sadece aktör ağı kullanır. Bu, yürütmenin merkezi olmadığı anlamına gelir. Aslında ÇADDPG, DDPG'nin çok ajanlı versiyonu olarak düşünülebilir. Temel amaç yürütmeyi merkezden uzaklaştırmaktır. Q-öğrenme ve DDPG, gruptaki diğer ajanların bilgilerini kullanmadıkları için çok ajanlı ortamlarda düşük performans gösterir. ÇADDPG yaklaşımı, tüm ajanların gözlemlerini ve eylemlerini kullanarak bu zorluğun üstesinden gelir.

2.4. Önerilen Yöntem (Proposed Method)

Bu çalışmanın temel amacı; dinamik alan kapsama ile görevli bir grup İHA'nın, Şekil 3'te görüldüğü gibi çok ajanlı derin pekiştirmeli öğrenme (ÇADPÖ) yaklaşımı ile modellendiği bir yöntem inşa etmektir. ÇADPÖ yaklaşımında İHA'lar, dinamik ortamda görev yapan mobil ajanlar olarak kabul edilir ve ortak bir hedef için etkileşime girer. Böylelikle Tablo 5'te verilen algoritma yardımıyla, kapsama en üst düzeye çıkarılırken düşük enerji tüketimi ile yüksek adillik indisi sağlayan stratejiler üretilerek akıllı bir sistem elde edilebilir. Çalışmanın devamında mobil ajanlar, ajan kelimesi ile ifade edilecektir.

Ajan tabanlı kapsama ve enerji tüketimi konularını sadeleştirmek için hedef alan ızgaralara bölünmüş ve her ızgaranın merkezi ızgara merkezi (IM) olarak adlandırılmıştır. Her ajanın makul bir sürede bir IM'ye konumlanmış olması amaçlanmaktadır.

Varsayım 1: Ajan takımı içerisindeki tüm ajanların aynı özelliklere sahip olduğu ve 2 boyutlu bir düzlemde hareket ettikleri varsayılmıştır. Takım içerisindeki her bir ajan A_i ile temsil edilir, $A_i \in A \mid i = 1, \dots, N|$.

Varsayım 2: Her ajanın kendi konumunu bildiği varsayılmıştır. Önerilen yöntemde, hedef alana ulaşmak, hedef alandaki diğer ajanları keşfetmek ve iletişim aralığındaki ajanlarla etkileşimde bulunabilmek için coğrafi yaklaşım kullanır, $(x_i^A(t), y_i^A(t))$ Ajanlar, takımda kaç ajan olduğu bilgisine sahiptir; ancak etkileşimde bulunmadıkları yani iletişim mesafesi içerisinde bulunmayan ajanların konum bilgisine sahip değildir. Ortamdaki tüm ajanların t anındaki konumları $A^N = \{(x_i^A(t), y_i^A(t), \dots, (x_m^A(t), y_m^A(t)\}$ ile gösterilir.

Varsayım 3: Her ajanın, dairesel bir şekle sahip olduğu (çap: $\emptyset_A$) ve dairesel bir algılama bölgesine sahip olduğu varsayılmaktadır. A_i ve A_j arasındaki mesafe $d_{ij} = |A_i A_j|$ şeklinde ifade edilir. Dairesel şekildeki ajanların çap uzunluğu ve dairesel algılama bölgesinin çap uzunluğu 0'dan büyük olmalıdır. Ek olarak iletişim mesafesi, ajanın sahip olduğu dairesel şeklin yarıçapından küçük olmamalıdır. Aksi durumda iletişim kurulamaz ve kolektif davranış üretilemez. A_i ve A_j ajanlarının etkileşime girebilmesi için aşağıdaki eşitsizliğin sağlanması gerekmektedir:

$$\left|(x_i^A(t), y_i^A(t)), (x_j^A(t), y_j^A(t))\right| \leq d_{ij} \quad (1)$$

Her ajanın iletişim/algılama aralığı gibi bağlantı sınırlamaları vardır. Her ajan her bir öğrenme adımında hedef alanı ızgaralara ayrıştırır. İletişim mesafesi, ızgara ayrıştırma mesafesinden daha kısa olduğunda ajanlar arası bağlantı kopacaktır. Bununla birlikte iletişimin yeniden sağlanabilmesi adına yapılabilecek herhangi bir ajan eylemi, diğer ajan konumlarının değişmesine sebep olabilir. Sonuç olarak uygun çözüm bulunabilmesi amacıyla yapılan her bir eylem, enerji tüketimini artırır.

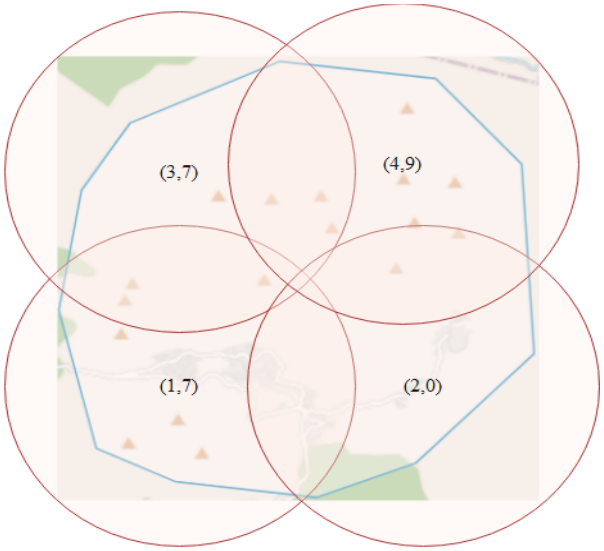
Bu bölümde, hedef alanda bağlı ajanların (bilgi alışverişi yapabilen) yüksek adil indisi şemsiyesi altında en az enerji tüketimi ile en fazla kapsama alanı elde edilebilmesi amacıyla önerilen yöntemin detayları sunulmuştur.

2.4.1. Izgara ayrıştırma (Grid decomposition)

Sistem içerisinde her bir İHA bir ajan tarafından temsil edilir. İki boyutlu bir ortamda ajanlar, hedef alanda ızgaralar oluşturarak merkezlerine konumlanmayı öğrenir. Hedef alan düzenli bir şekil olmayabilir. Düzenli veya düzenli olmayan hedef alan için bir soyut alan oluşturulur. Soyut alan, hedef alanı içeren en küçük düzenli dörtgen alandır. Ajanlar, ızgaralara ayrılmış soyut alan merkezlerine yönlenir. Ajanlar, kendilerine en yakın ve en çok İÇN içeren ızgaraya giderek en hızlı ve en uygun çözümü üretmeye çalışır. Aynı zamanda, hedef alanda ajanlar arası iletişimin sağlanması da amaçlanır. T hedef

alanı ve k^T hedef alanın köşeleri olmak üzere, hedef alanı içeren en küçük düzenli dörtgen alan $\{(x_i^T, y_i^T), \dots, (x_n^T, y_n^T) \; \forall \; x_{max}^T, y_{min}^T \in k^T\}$ için $\{(x_{min}^T, y_{max}^T), (x_{max}^T, y_{max}^T), (x_{max}^T, y_{min}^T), (x_{min}^T, y_{min}^T)\}$ şeklinde ifade edilir. IM'ler ise $j \in \{1, \dots, N\}$ tüm $IM_j \subseteq T$ ile temsil edilir. Hedef alan içerisindeki tüm İÇN'lerin kümesi $\text{İÇN}_T = \{\text{İÇN}_1, \dots, \text{İÇN}_n\}$ şeklinde tanımlanır. Izgara ayrıştırma algoritması Tablo 1'de verilmiştir.

İlk olarak Tablo 1'de görüldüğü üzere hedef alanı içeren en küçük düzenli dörtgen alanın koordinatları bulunur. Daha sonra belirlenen ayrıştırma mesafesi kullanılarak (genellikle ajanın iletişim/algılama mesafesi), ajanın konumlandırılacağı IM'ler bulunur. Şekil 2'de görüldüğü gibi IM'lerde kaç adet İÇN olduğu hesaplanır.



Şekil 2. IM-İÇN ikilisi (Center of grid-PoI pair)

Burada temel amaç ızgara oluşturup merkezine ajan koymak değil; hesaplanan noktalara ajan yerleştirmek ve ajanların iletişim mesafesi kadar bir kapsama alanı oluşturmaktır. Oluşturulan kapsama alanları ile hedef alandaki İÇN'ler kesiştirilir. Son olarak IM'nin indeksi ve içerdiği İÇN sayısını içeren ikili listeye atılır.

2.4.2. Ödül stratejisi (Reward strategy)

Seyir mesafesini en aza indirmenin zamandan, enerjiden ve donanım kullanım süresinden tasarruf etmek gibi birçok getirisi vardır. [30]'dan esinlenilerek hedef ataması gerçekleştirmek ve seyahat mesafesini en aza indirmek amacıyla bir ödül yapısı tasarlanmıştır. Tablo 2'de görüldüğü üzere ödül, her bir ajanın her bir IM'ye uzaklığı ile belirlenir. Her adımda, ajanların IM'lere olan uzaklıkları hesaplanır. Fakat alınan ödülü en yüksek düzeye çıkarmak burada başarısız bir strateji olacaktır. Çünkü elde edilen ödülün en yüksek düzeye çıkarılması, ajanları hedef alandan uzaklaştıracak ve alanların kapsanmasını engelleyecektir. Bu sebeple elde edilen mesafe değerlerinden en küçük olanı seçilir. Sonrasında ise bu uzaklık değeri -1 ile çarpılarak negatif değeri alınır. Bu sayede ajanların ızgaralara ayrıştırılmış alanlara dağılım ödülü negatif değerden 0'a yakınsar. Böylelikle, PÖ'nün temelinde var olan ödül güdümlü öğrenme gerçekleştirilir.

Tablo 2. Ödül işlevi 1 (Reward function 1)

Algoritma Sözde Kodu
Başla: $r_1 = 0$
for ızgara 0'dan IM^N'e kadar do
ajanların ızgaraya mesafesini (d) hesapla
$r_1 = r_1 + (-min(d))$
end for
return r_1

IM'lere doğru hareket eden ajanlar birbirleriyle çarpışırlarsa cezalandırılır. Bu nedenle, ajanlar çarpışmalardan kaçınırken hedef alanları kapsamayı öğrenmelidir. Çarpışma olup olmadığının belirlenebilmesi için her adımda ajanlar arasındaki mesafe hesaplanır. Ajanların daire şeklinde modellendiği düşünüldüğünde mesafe, ajanların yarıçap toplamlarından daha küçük ise çarpışma gerçekleşmiş demektir. Çarpışma önleme algoritması Tablo 3'te verilmiştir.

Tablo 3. Ödül İşlevi 2 (Reward function 2)

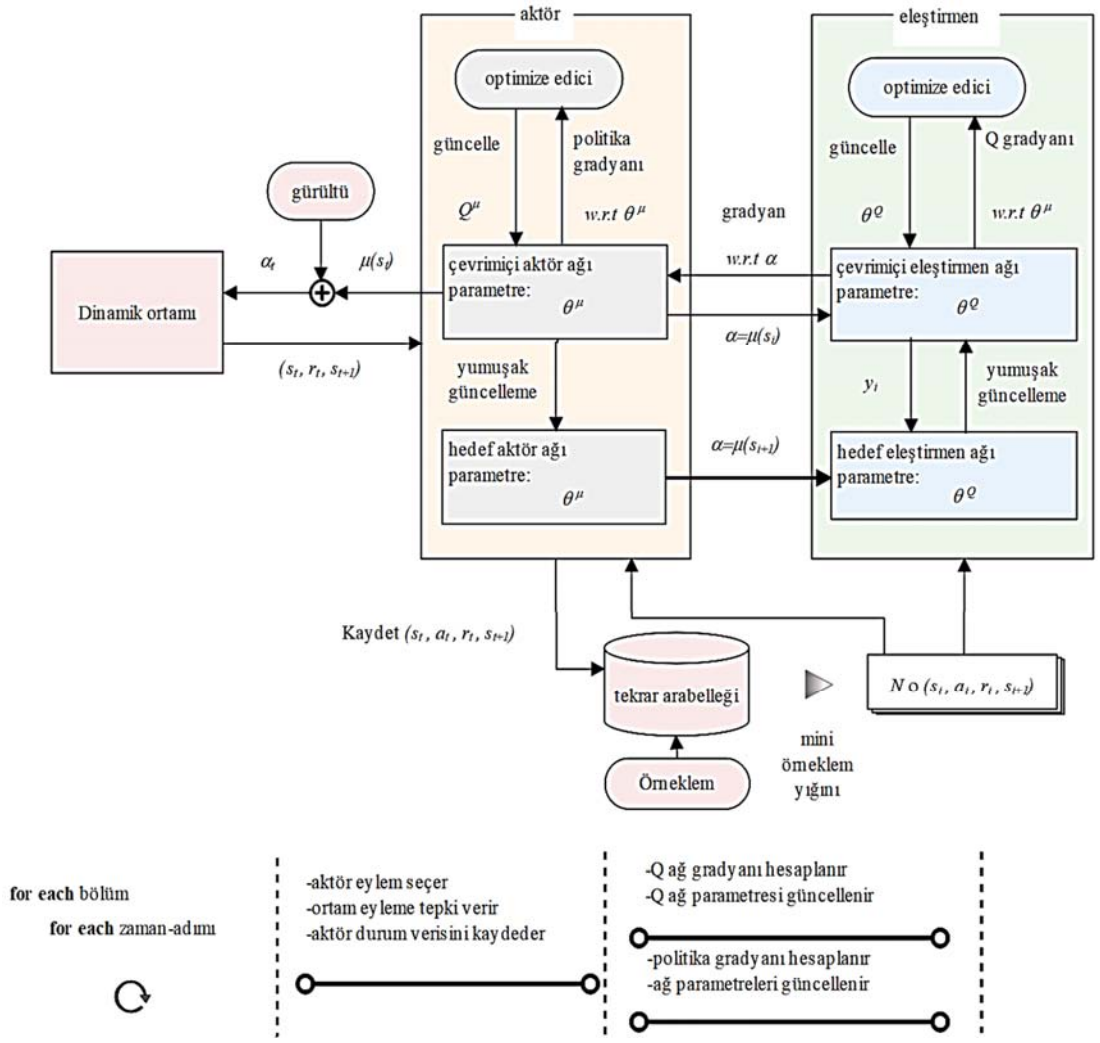
Algoritma Sözde Kodu
Başla: $r_2 = 1, \emptyset_A$
for ajan 0 dan E^N'e kadar do
if $d_{A_i A_j} < \emptyset_A$ then
$r_2 = \gamma_r * r_2$
end if
end for
return r_2

Tablo 3'te sözde kodu verilen çarpışma hesaplama yönteminde γ_r çarpışmadan kaçınma için indirim faktörünü temsil eder:

$$p_i = \prod_j w(ij), j \in \{1, \dots, N\} \text{ tüm } A_j \in E_i^C \quad (2)$$

Tablo 1. Izgara ayrıştırma algoritması (Grid decomposition algorithm)

Algoritma Sözde Kodu
Başla: $L_I(i,k) \longleftarrow (\textit{ızgara indeksi}, \text{İÇ}\textit{N sayısı})$– ızgaralar için boş liste
Eşitle: ızgara ayrıştırma için mesafe m, $m \leq A_m \mid \; A_m$ ajan iletişim mesafesi
Başla: $x_{temp} = x_{min} + (m/2)$
$y_{temp} = y_{min} + (m/2)$, $\text{İÇ}N_T \longleftarrow \textit{hedef alandaki ilgi çekici noktalar}$
for y_{temp}' ten y_{max}' a kadar do
$j = 0$
for x_{temp}'ten x_{temp}' a kadar do
$IM_j = (x_{temp}, y_{temp}), x_{temp} = x_{temp} + m$
$\text{İÇ}N_{temp} = IM_j.buffer(m/2) \cap \text{İÇ}N_T$
$L_I.insert(j, \text{İÇ}N_{temp}.size())$
$j++$
end for
$x_{temp} = x_{temp} + (m/2), y_{temp} = y_{temp} + m$
end for
return L_I



Şekil 3. Önerilen yöntemin çalışma prensibi (Working principle of the proposed method)

$$w(ij) = \begin{cases} 1, (\emptyset_i + \emptyset_j)/2 \leq d_{ij} \\ \gamma_r, (\emptyset_i + \emptyset_j)/2 > d_{ij} \end{cases} \quad (3)$$

Ajanların, İÇN sayısı fazla olan IM'lere doğru yönlenmesini sağlayan ödül işlevi Tablo 4'te verilmiştir:

Tablo 4. Ödül İşlevi 3 (Reward function 3)

Algoritma Sözde Kodu
Başla: $r_3 = 0$
for ajan 0'dan E^N'e kadar do
Bul: A_i'nin erişebildiği ajanları
Bul: Erişilebilir ajanların (A^R) kapsadığı IM'leri
Topla: $r_3 \leftarrow L_I(i,k)$ kullanarak kapsanan İÇN sayıları (c),
$r_3 = \sum_i^n c_i \forall i, n \{A_i, \dots, A_n\} \in A^R$
end for
return r_3

En kısa yol, çarpışmadan kaçınma ve kapsanan İÇN'ler için tasarlanan ödül işlevleri kullanılarak elde edilen kümülatif ödül işlevi aşağıdaki gibidir:

$$r_i = r_1 * r_2 * r_3 \quad (4)$$

Kümülatif ödül işlevinde; kapsama sürecinde çarpışmadan kaçınarak yol planlama karşılığında elde edilen ödül puanı ve kapsanmayan İÇN yüzdesi çarpılır. Yol planlama süreci negatif bir değerdir; kapsanmayan İÇN sayısı ne kadar yüksek ise elde edilecek kümülatif puan o kadar düşük olacaktır. Bu yaklaşım ödül aralığını genişleterek öğrenme sürecini hızlandırır. Böylelikle önerilen yöntemindeki PÖ ajanları, Eş 4. kullanan ödül güdümlü çalışma yaklaşımı ile 0'a yakınsamaya çalışır.

2.4.3. Çok ajanlı sistem modeli (Multi-agent system model)

Ajanlar arasındaki iş birliği karmaşıklığının çözümü için bir ÇAS modeli oluşturulmuştur. 2 boyutlu bir düzlem üzerinde hareket eden ajanların A_i, t zamanındaki konumu $(x_i^A(t), y_i^A(t))$ ile ifade edilmiştir. Hedef alan T, ızgaralara ayrıştırma işlemi sonrasında gidilecek ızgara alanları ise $j \in \{1, \dots, N\}$ tüm $IM_j \subseteq T$ ile temsil edilir. Önerilen bu ÇAS modelinde aşağıdaki hedeflerin karşılanması beklenmektedir.

- Ajanlar, kapsanan İÇN sayısı maksimum olacak şekilde IM'lere dağılmalıdır.
- Her IM yalnızca bir ajan tarafından kapsanmalıdır. Örneğin $\forall i, j, IM_i' \neq IM_j' \, i, j \in IM^T$

Tablo 5. Önerilen yöntem (Proposed method)

Algoritma Sözde Kodu
Başla: öğrenme ve ödül indirim faktörü
for $bölüm = 1$ to M do
Başla: N ve hedef alan T
Bul: Tablo 1 ile IM^T
Al: x' in ilk durumu
for $t = 1$ to maks_bölüm_uzunluğu do
her bir ajan i için, seçilen eylem ve gürültü: $N_t, a_i = \mu_{\theta_i}(o_i) + N_t\ N_t$, politika ve gözlem: $w.r.t$
Eylemleri uygula: $a = (a_1, \dots, a_N)$
Hesapla: Tablo 4 kullanarak ödül r
Belirle: yeni durum x'
Sakla: (x, a, r, x') bilgisini tekrar oynama belleğine D
$x \leftarrow x'$
for ajan $i = 1$'den N'ye kadar do
Örneklem al: tekrar oynatma arabelleğinden D mini-yığın olarak S
Güncelle: Eş. 6 kullanarak eleştirmen
Güncelle: Örneklem alınmış politika gradyanı Eş. 7 ile aktör
end for
Güncelle: her ajan için hedef ağ parametreleri
end for
end for

- Hedef alan içerisindeki IM'lere konumlanmış ajanlar, birbirleriyle iletişim sağlayabilecek mesafede olmalıdır.
- Ajanlar çarpışmadan kaçınmalıdır yani herhangi iki ajan herhangi bir t anında aynı konumda bulunamaz. Örneğin $\forall i, j, \left(x_i^A(t), y_i^A(t)\right) \neq (x_j^A(t), y_j^A(t))$

Ajan politikası μ ile politika parametresi ise θ ile temsil edilir. N ajanlı ÇAS modelinde, durum geçişleri için politikalar $\mu = \{\mu_1, \mu_2, \dots, \mu_N\}$, parametreler ise $\theta = \{\theta_1, \theta_2, \dots, \theta_N\}$ ile ifade edilir. Ayrıca, model, politika bağımsız eğitim için bir deneyim tekrar arabelleği kullanır. Ajan, her bir adımda ortak durumu, sonraki ortak durumu, ortak eylemi ve ajanların her biri tarafından alınan ödülleri gösteren bilgileri $(x, x', a_1, \dots, a_N, r_1, \dots, r_N)$ şeklinde saklar. Ardından, ajanı eğitmek için tekrar arabelleğinden bir örnekleme yapar. Örnekleme alınmış bilgiler kullanılarak ajanın eleştirmeni güncellenir. Böylelikle eleştirmen için örneklenmiş zamansal fark hatası kullanan kayıp fonksiyonu şu şekilde tanımlanır:

$$y = r_i + \gamma Q_i^{\mu'}(x', a_1', \dots, a_N')|_{a_j' = \mu_j'(o_j)} \quad (5)$$

$$L(\theta_i) = \frac{1}{S}\sum_j \left(y^j - Q_i^{\mu}\left(x^j, a_1^j, \dots, a_N^j\right)\right)^2 \quad (6)$$

Yukarıdaki eşitlikte a', γ, S, o ve Q_i^{μ} sırasıyla; sonraki ortak eylemi, indirim faktörünü, tekrar arabelleğinden rastgele seçilen örneklemenin boyutunu, ortamın kısmi gözlemlerini ve merkezi eylem-değer fonksiyonunu gösterir. Aktörün güncellenmesi için örneklenen politika gradyanı aşağıdaki gibi tanımlanır:

$$\nabla_{\theta_i} J \approx \frac{1}{S}\sum_j \nabla_{\theta_i}\mu_i(o_j^i)\nabla_{a_i} Q_i^{\mu}\left(x^i, a_j^i, \dots, a_N^j\right)|_{a_i = \mu_i(o_i^j)} \quad (7)$$

2.4.4. Önerilen Yöntemin İnşası (Construction of proposed method)

ÇAPÖ, öğrenmeye dayalı sistematik bir yaklaşımdır ve bilinmeyen dinamik ortamları ele alma yeteneğine sahiptir. Davranış stratejileri tasarlanırken belirtilen özelliklerden faydalanılabilmesi için önerilen yöntem, derin öğrenme ile ÇAPÖ temelinde modellenmiştir. Ek olarak merkezi öğrenme-merkezi olmayan yürütme şeması tercih edilerek birbirinden bağımsız hareket eden; ancak grubun ortak hedefi için politika üreten çok ajanlı sistem oluşturulmuştur.

Merkezi öğrenme, eğitim katmanında; merkezi olmayan yürütme ise yürütme katmanında ele alınır. Eğitim katmanında politika eğitimi için ajan eylemlerine göre elde edilen ödül puanları kullanılır. Her bir ajan, yalnızca yerel gözlemlere sahiptir ve mevcut durumunu paylaşmak için diğer ajanlarla iletişim kurar. Ajanlar, merkezi bir kontrol mekanizması gerekmeksizin grubun ortak başarısına hizmet eder ve en uygun konumu bulmak amacıyla bir politika öğrenir. ÇADDPG'de ajanlar, gruptaki diğer ajanların aldığı ödüllerin toplamı, ortalaması gibi ortak bir ödül tasarımı kullanırlar. Fakat önerilen ödül stratejisi ile ortak paylaşımda olan kolektif ödül politikası elde edilmiştir. Kolektif ödülün yapısı, tüm ajanların uygun pozisyon almasına bağlıdır ve dolayısıyla kendi başına kolektif bir doğası vardır. İşbirlikçi dinamik bir ortam için r_i, tüm ajanlar tarafından paylaşılan ortak bir ödüldür. r_i ile ajan bazlı ödül ve grup bazlı ödül olmak üzere 2 bilgi elde edilir. Bu, ÇADDPG'deki ödül paylaşım politikasının ötesinde, etkin bir iş birliği ortaya çıkarır.

Önerilen yöntemin davranış stratejisi için yaygın olarak kullanılan "çekicilik" ve "kaçınma" kombinasyonu baz alınmıştır. [32]'ten esinlenerek, "çekicilik" amaçlı, ajanların Tablo 2'de ifade edildiği gibi hedef alanın iç kısmına çekilmeleri yani IM'lere konumlanmaları pozitif bir ödül olarak kabul edilir. Ajanların iletişim mesafesinin dışına çıkması ise negatif ödül kabul edilir. Buna ek olarak Tablo 4'te sunulduğu üzere, ajanlar erişilebilir ajanlarla birlikte kapsadığı İÇN'ler için olumlu ödül alırlar. Bu olumlu ödül, ajanların iletişim mesafesini aşmasını engeller; dolayısı ile "çekicilik" yönünden olumlu sonuç elde edilir. [33]'ten esinlenerek, ajanlar arası çarpışma (kaçınma) aralığı tanımlanır ve bu aralığa dayanarak, herhangi iki ajanın birbirine çok yakın olduğu durumda olumsuz bir ödül verilir. Böylece Tablo 3'e göre, her bir ajan, diğer ajanlarla uygun mesafede olduğu için olumlu bir ödül; çok yakın mesafede olduğu durum için olumsuz bir ödül alır.

Önerilen yöntem model bağımsız öğrenme yapısına sahiptir ve tekrar arabelleğini kullanarak öğrenme sürecini işletir. Merkezileştirilmiş modül, ajanlara eğitim süresi boyunca politikalarını nasıl güncelleyecekleri konusunda rehberlik eder. Eğitim süreci bölümlere

ayrılmıştır ve her bölüm koşulmadan önce ajanların konumu rastgele belirlenir. Ajanlar, eğitim sürecinde senaryoyu yeniden koşmak için tekrar arabelleğine eylem, durum ve ödülden oluşan bilgi grubunu depolar. Her zaman adımında, tekrar arabelleğinden örnekleme yapılır. Eleştirmenler Eş. 6 ile güncellenirken, Eş. 7 ile ajanların politika gradyanı güncellenir. Her ajan, alan kapsamını artırmak için en uygun ortak eylemi belirler. Yürütme sürecinde ise merkezi modül kaldırılır ve yerel gözlemler için aktör ağı kullanılır. Önerilen yöntemin sözde kodu Tablo 5'te verilmiştir.

2.4.5. Eğitim süreci (Training process)

Her ajanın kendine özgü aktör ve eleştirmen ağları vardır. Daha önce açıklandığı gibi yöntem, tekrar arabelleğinde depolanan deneyimlerden (yani eylem, durum ve ödül) öğrenir. Diğer bir deyişle, öğrenme süreci boyunca her t zaman diliminde ağdaki tüm ajanlar için aktörler ve eleştirmenler, rastgele örneklem alınan mini-yığın kullanımıyla deneyimlerden güncellenir.

Tablo 5'te eğitim sürecindeki öğrenme yaklaşımı sözde kodu verilmiştir. Önerilen yöntemde, ajanların başlangıç konumlarından hedef alan üzerindeki konumlanmalarına kadar geçen süreç bir bölüm olarak ifade edilir. Eğitim için her bölüm t zaman dilimlerinden oluşur. Eğitim döngüsünde, sistem s_1 başlangıç durumunu alır ve ortamın başlangıç koşulları oluşturulur. Her ajan i, Q_i gözlemi ile aktör μ_{θ_i}'ye göre bir eylem seçer. Ajanın yerel olarak en uygun politikayı seçmesini ve daha fazla keşif gerçekleştirmesini önlemek için, seçilen eyleme gürültü eklenir. Seçilen eylemi gerçekleştiren ajanlar bir ödül değeri r_t ve yeni bir s_{t+1} durumu elde edecektir. Seçilen eylem, ajanı hedef bölge dışına çıkmaya veya diğer ajanlarla çarpışmaya zorlarsa Eş. 4'e göre cezalandırılır. Dolayısıyla ajan bu eylemden kaçınmayı ve ilgili konumu seçmemeyi öğrenir. Daha sonra (s_t, a_t, r_t, s_{t+1}) 'in son değerleri tekrar oynatma arabelleğinde saklanır. Eğitim sürecinin sonunda, t zaman aralığındaki her ajan, tekrar oynatma arabelleğinden D mini-yığın S rastgele seçer ve ardından Eş. 6 kullanarak eleştirmeni günceller. Bu adımdan sonra, aktör Eş. 7 ile güncellenir. En son aşamada hedef ağ, kayıp işlevi ve öğrenme oranı ile yavaş yavaş güncellenir.

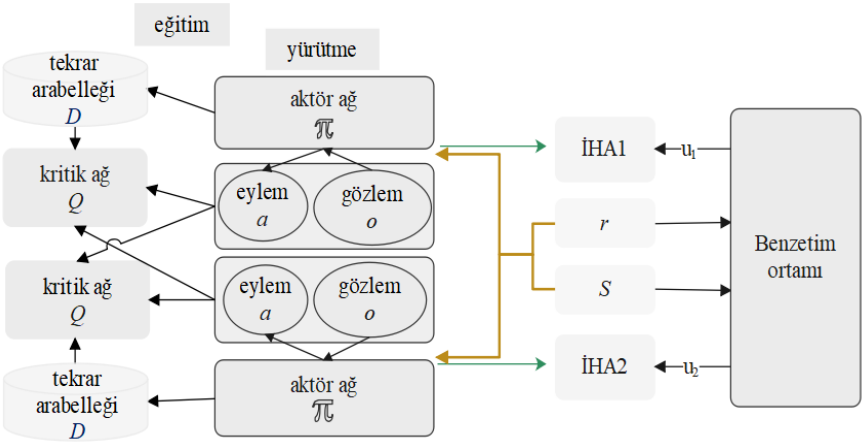
3. Deneysel Çalışmalar (Experimental Studies)

Gerçek zamanlı olarak çalışan bir benzetim mimarisi kullanmak, önerilen yöntemlerin daha az maliyetle ön doğrulama yapılabilmesi adına uygun bir yaklaşımdır [34]. Dolayısıyla önerilen yöntemin değerlendirilebilmesi amacıyla, OpenAI ekibi tarafından geliştirilen çok ajanlı aktör-eleştirmen platformunu kullanarak bir benzetim ortamı tasarlanmıştır [26]. Alan kapsama süreci için önerilen yöntem ile tasarlanan benzetim ortamı etkileşimi Şekil 4'te gösterilmiştir.

3.1. Deneysel Ayarlar (Experimental Settings)

Benzetim ortamı, geometrik merkezin orijin olduğu ve her bölümün 1 birim olarak ayarlandığı bir koordinat düzlemidir. Düzlem, ajanlardan ve bir hedef alandan oluşur. Ajanın etrafındaki daire, alan kapsamasını temsil etmek için kullanılır. Ayrıca benzetim ortamında bir ajan, dairesel bir şekil ile temsil edilir. İletişim mesafesi ve hedef alanın şekli, eğitim bölümünün başında ayarlanır. Hedef alan herhangi bir şekil olabilir. Her eğitim bölümünde, ajanların konumları ve hedef alanın konumu rastgele oluşturulur. Rasgele oluşabilecek olumlu veya olumsuz sonuçların filtrelenerek kolektif davranışın daha kolay analiz edilebilmesi amacıyla benzetim çalışmalarında en az 3 ajan kullanılmıştır. Buna ek olarak benzetim çalışmalarında ajan sayısı 8 ile sınırlandırılmıştır. Bunun sebebi, kıyaslanan yöntemlerin dağıtık öğrenme ve yürütme mekanizmasına sahip olması ve dolayısıyla ajan sayısındaki artışın, mevcut sonuçlarla benzer eğilimde sonuçlar üretmesidir. Benzetim çalışmaları Tensorflow 1.2.0 ve Python 3.7 ile Ubuntu 16.04.3 üzerinde gerçekleştirilmiştir. Eğitim için bölüm sayısı 5000, bölüme ait işlem sayısı ise 250 olarak belirlenmiştir. Yani ajanlar her bölümde 250 eylemde bulunabilir. Diğer parametreler ise şu şekilde tanımlanmıştır; çok katmanlı algılayıcıdaki birim sayısı 64, yığın boyutu 1024, indirim faktörü 0,99 ve öğrenme oranı 0.001'dir. γ_r yani r_2 işlevinde kullanılacak öğrenme oranı 1 olarak seçilmiştir. Ajanların başlangıç konumlarının rastgele belirlenmesi sebebiyle benzetim senaryoları 50 kez tekrarlanmıştır ve elde edilen metriklerin ortalama değerleri alınmıştır. Önerilen model, aynı benzetim ayarları kullanılarak DRL-EC3 ve DRL_EC3'ün geliştirilmiş versiyonu olan Dağıtık_DRL_EC3 ile karşılaştırılmış ve benzetim sonuçları değerlendirilmiştir. Karşılaştırma ve doğrulamalar için 3 konu belirlenmiştir:

- Ajan sayısı artışının kapsama üzerindeki etkisi: Sistemin kapsadığı ortalama İÇN puanıdır. Tablo 4 kullanılarak hesaplanır.
- Ajan sayısı artışının enerji kullanımı üzerindeki etkisi: Sistemin sahip olduğu ajan sayısı karşılığında kapsadığı İÇN sayısı, enerji verimliliğini ifade eder. Kapsanan İÇN'lerin normalleştirilmiş halidir; yani Tablo 4 ile elde edilen sonucun sistemdeki ajan sayısına oranıdır.
- Kapsanan İÇN'ler için adillik indisidir: İÇN kapsam puanlarına ilişkin Jain [35] adillik indisidir. N ortamdaki İÇN sayısını temsil ederken; $c_t(i)$, t anındaki i İÇN'sini kapsayan ajan sayısını ifade eder. $J = 1$ olması durumu ajanlar arasında mükemmel adilliği gösterir.

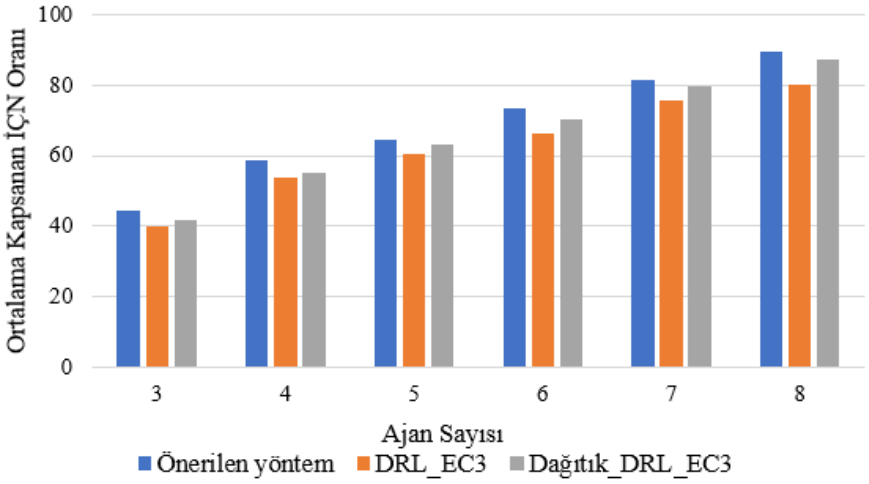


Şekil 4. Deney platformu (Experiment platform)

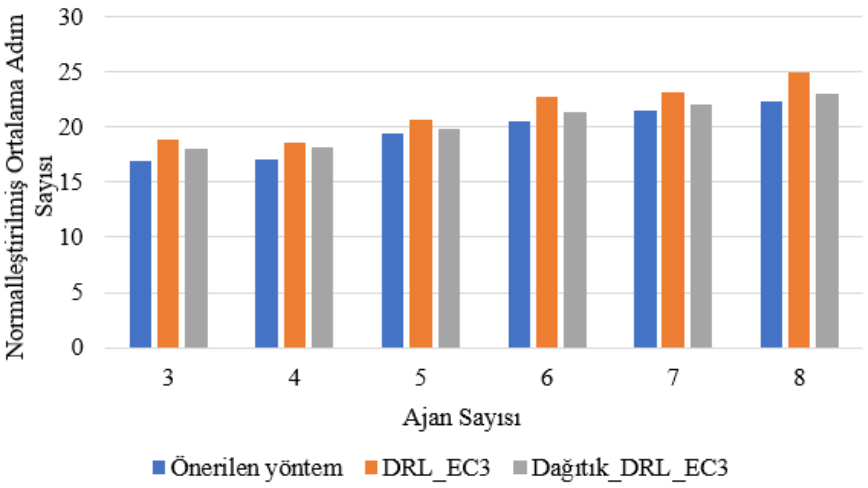
$$J = \frac{\left(\sum_{i=1}^{N} c_t(i)\right)^2}{N \sum_{i=1}^{N} c_t(i)^2} \quad (8)$$

3.2. Deneysel Sonuçlar (Experimental Results)

İlk deneysel çalışmada, ortamdaki ajan sayısı artışının kapsama üzerindeki etkisi incelenmiş ve 3 modelin performansı karşılaştırılmıştır. Karşılaştırma için kullanılacak kapsama oranları Tablo 4'te verilen yönteme göre hesaplanmıştır. Şekil 5'te görüldüğü üzere, önerilen yöntem DRL-EC3'e göre yaklaşık %8,64; Dağıtık_DRL_EC3'ye göre ise yaklaşık %3,51 daha fazla kapsama oranı elde etmiştir. Örneğin, ajan sayısı 4 olduğunda, önerilen yöntem yaklaşık %58,8 oranında kapsama yaparken; DRL-EC3 yaklaşık %53,9, Dağıtık_DRL_EC3 ise yaklaşık %55,2 oranında kapsama yapmıştır. Ajan sayısının 8 olması durumunda, önerilen yöntem tarafından elde edilen kapsama oranı yaklaşık %89,4, DRL-EC3 tarafından elde edilen %80,2 iken Dağıtık_DRL_EC3 tarafından elde edilen kapsama oranı ise %87,3 olmuştur. Diğer senaryolar için de benzer eğilim bulunmaktadır. Dolayısı ile önerilen yöntemin ulaştığı ortalama kapsanan İÇN oranı düzenli artış göstererek daha iyi sonuçlar elde edilmiştir. Ajan sayısındaki artış, İÇN kapsama sürecinde ajanların farklı desenler ile bağlantı kurmasına olanak sağlamıştır. İkinci deneysel çalışmada, 3 modelin enerji tüketimi kıyaslaması yapılmıştır. Enerji tüketimi hesaplamasında, ajanların yaptıkları eylem sayısından faydalanılmıştır. Her bölümde, sistemin elde ettiği kapsama oranı kaydedilir. Eğitimin sonunda ise kapsama puanları toplanarak eylem sayılarına bölünür. Elde edilen sonucun sistem içerisindeki ajan sayısına bölünmesi ile bir ajanın eylem başına kapsadığı İÇN bulunmuş olur. Bu sonuçtan yola çıkılarak bir İÇN kapsamak için bir ajanın başlangıç konumundan itibaren kaç adım attığı hesaplanır. Bu değer normalleştirilmiş ortalama enerji tüketimini ifade eder. Şekil 6'te görüldüğü üzere, DRL-EC3 ve Dağıtık_DRL_EC3 modelleri önerilen yönteme göre sırasıyla yaklaşık %8,7, %3,75 daha fazla enerji tüketmiştir. Örneğin, ajan sayısı 4 ve 8 iken kıyaslanan modellerin normalleştirilmiş ortalama adım sayıları şöyle gerçekleşmiştir: önerilen model 17 ve 22,3, DRL-EC3 18,55 ve 2, Dağıtık_DRL_EC3 ise 21,37 ve 22,94. Bu sonuçlara göre ajan sayısı arttıkça enerji tüketim değerlerinde büyük değişiklikler olmadığı gözlemlenmiştir. Çok ajanlı sistemlerin rekabetçi ve işbirlikçi doğası gereği ajan sayılarının politikalar üzerinde büyük bir etkisi yoktur. Bununla birlikte önerilen yöntem; daha az enerji harcayarak maksimum kapsama oranına ulaşmıştır. Bunun 2 sebebi olduğu düşünülmektedir; (i) yol planlaması için özel bir ödül stratejisi tasarlanmıştır; (ii) ödül, ajanın sadece kendi durumuna göre değil, erişebildiği tüm ajanların durumuna göre belirlenir. Son olarak, ajan sayısına göre adillik indisi açısından üç model karşılaştırılmıştır. Şekil 7'da, farklı sayıda ajanın 3 model için adillik indisi sonuçları sunulmuştur. Önerilen modelin sırasıyla DRL-



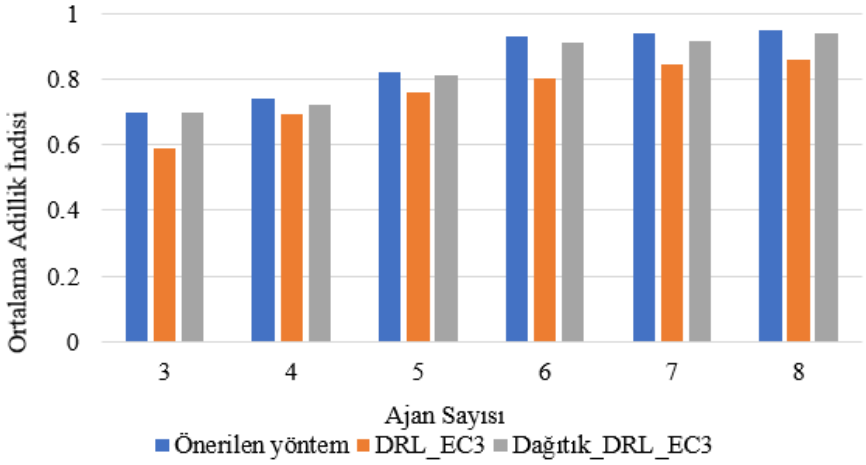
Şekil 5. Kapsanan İÇN-ajan ilişkisi (Covered POI-agent relationship)



Şekil 6. Eylem sayısı-ajan ilişkisi (Number of actions-agent relationship)

EC3 ve Dağıtık_ DRL-EC3'e göre yaklaşık %10,3 ve %1,4 oranında bir ortalama artış ile daha iyi adillik indisine sahip olduğu görülmektedir. Örneğin ajan sayısı üç olduğunda önerilen model 0,699 adillik indisine ulaşırken, DRL-EC3 0,58 adillik indisi ve Dağıtık_ DRL-EC3 0,698 adillik indisi elde etmiştir. Ajan sayısı 6 olduğu durumda önerilen model, DRL-EC3 ve Dağıtık_ DRL-EC3'ün adillik indisleri sırasıyla 0,93, 0,84 ve 0,91 olmuştur. Benzer eğilim diğer senaryolar için de gözlemlenmektedir. Ajan sayısının artması, kapsanan ızgara sayısında artışa sebep olur. Bu ilişki, adillik indisinin doğrudan iyileşmesine yol açar. Ek olarak, önerilen yöntemde daha az adım ile diğer yöntemlerin eriştiği benzer adillik indisi oranı yakalanmıştır. Bu durum, amaca yönelik olarak tasarlanan ödül stratejilerinden kaynaklanmaktadır.

Çalışmanın bu bölümünde, önerilen modelinin enerji, kapsama alanı ve adillik indisi gibi konular karşısındaki davranışları incelenmiştir. Ardından bu üç metrik kullanılarak önerilen yöntem, DRL-EC3 ve Dağıtık_ DRL-EC3 modelleri ile karşılaştırılmıştır. Benzetim sonuçları Tablo 6'da özetlenmiştir.DRL-EC3 ve Dağıtık_ DRL-EC3 yöntemlerinde bir ödül stratejisi bulunurken, bu çalışmada, amaca yönelik 3 farklı ödül stratejisi tasarlanmıştır. Kolektif ödül tasarımı ile düşük enerji tüketimi-yüksek kapsama oranı elde edilmeye çalışılmıştır. 3 farklı ödül stratejisi kullanımı, elde edilebilecek ödül boyutunu artırarak durum-eylem ikilileri arasındaki ilişkiyi belirginleştirir. Böylelikle öğrenme süresi düşerken öğrenme kalitesi artar. Bununla birlikte, hedef alandaki ızgaralar ile bağlı ajanlar arasındaki mesafenin ölçülmesi temelinde tasarlanan ödül stratejisi (Tablo 3), öğrenmeye büyük katkı sağlamıştır. Bunun iki temel sebebi vardır; (i) sistemdeki ajanların bağlanarak haberleşme mesafesi dışına çıkmadan ızgaralara konumlanmaları, (ii) ajanların, ızgaralara dağılarak kesişimi azalttıkları oranda yüksek ödül almaları. Önerilen yöntemin aksine DRL-EC3 ve Dağıtık_ DRL-EC3 yöntemlerinde hedef planlama ile ilgili bir ödül stratejisi bulunmamaktadır. Bu durum, DRL-EC3 ve Dağıtık_ DRL-EC3 yöntemlerinin önerilen yönteme oranla daha düşük sonuçlar elde etmesine sebep olduğu düşünülmektedir. Önerilen yöntemde ajanlar, bağlı ajanların bilgilerini de kullanarak ödül puanlarını en yüksek seviyeye çıkarmaya çalışmaktadır. Bu yaklaşım ajanların birbirleri ile bağlanmasını ve kolektif karar almalarını zorlamaktadır. Ayrıca önerilen yöntemde ajanlar, kendilerine en yakın mesafedeki İÇN yoğunluğu en yüksek IM'ye konumlanmaya çalıştıklarından dolayı enerji tüketimi açısından da olumlu sonuçlar elde edilmiştir. Bu



Şekil 7. Adillik indisi-ajan ilişkisi (Fairness index-agent relationship)

Tablo 6. Önerilen yöntemin benzer çalışmalarla karşılaştırılması (Comparison of the proposed method with similar studies)

Çalışma	Önerilen Yöntem	DRL_EC3	Dağıtık_DRL_EC3
Teknik	Çok ajanlı derin pekiştirmeli öğrenme	Çok ajanlı derin pekiştirmeli öğrenme	Çok ajanlı derin pekiştirmeli öğrenme
Algoritma	Geliştirilmiş ÇADDPG	DDPG	ÇADDPG
Değerlendirme Metriği Ortalama Kapsanan İÇN Oranı	3 Ajan: 44,4 4 Ajan: 58,8 5 Ajan: 64,4 6 Ajan: 73,3 7 Ajan: 81,7 8 Ajan: 89,4	3 Ajan: 39,8 4 Ajan: 53,9 5 Ajan: 60,7 6 Ajan: 66,2 7 Ajan: 75,6 8 Ajan: 80,2	3 Ajan: 41,8 4 Ajan: 55,2 5 Ajan: 63,3 6 Ajan: 70,2 7 Ajan: 79,7 8 Ajan: 87,3
Değerlendirme Metriği Normalleştirilmiş Ortalama Adım Sayısı	3 Ajan: 16,9 4 Ajan: 17 5 Ajan: 19,4 6 Ajan: 20,5 7 Ajan: 21,5 8 Ajan: 22,3	3 Ajan: 18,9 4 Ajan: 18,6 5 Ajan: 20,6 6 Ajan: 22,7 7 Ajan: 23,2 8 Ajan: 25	3 Ajan: 17,9 4 Ajan: 18,1 5 Ajan: 19,8 6 Ajan: 21,4 7 Ajan: 22 8 Ajan: 22,9
Değerlendirme Metriği Ortalama Adillik İndisi	3 Ajan: 0,69 4 Ajan: 0,74 5 Ajan: 0,82 6 Ajan: 0,93 7 Ajan: 0,94 8 Ajan: 0,95	3 Ajan: 0,59 4 Ajan: 0,69 5 Ajan: 0,76 6 Ajan: 0,80 7 Ajan: 0,84 8 Ajan: 0,86	3 Ajan: 0,69 4 Ajan: 0,72 5 Ajan: 0,81 6 Ajan: 0,91 7 Ajan: 0,92 8 Ajan: 0,94

yaklaşım ÇAS'larda, hedef atama ile yol planlama algoritmasını temsil etmektedir. Bu algoritma yardımıyla ajanların kapsama alanlarındaki kesişimi en aza indirgenerek adillik indisi yüksek sonuç elde edilmiştir.

4. Sonuçlar (Conclusions)

Bu çalışmada, düzenli-düzensiz şekilli alanlarda ÇAS kullanımı ile en yüksek İÇN kapsamayı amaçlayan bir yöntem önerilmiştir. Aktör-eleştirmen bağlamında tasarlanan yöntemde, bağlı ajanların düşük enerji tüketimi ile yüksek kapsama elde etmesi amaçlanmıştır. Deneysel sonuçlara göre önerilen yöntem dinamik ortamda kapsama görevini başarıyla tamamlayabilmiştir. Ajanlar, test yani yürütme aşamasında, kendi aktör ağlarını kullanarak grubun ortak amacına yönelik dağıtık ancak kolektif eylemler oluşturmayı başarmıştır. Önerilen yöntemin performansının değerlendirilebilmesi amacıyla 3 farklı deneysel çalışma yapılmıştır. İlk deney çalışmasında önerilen yöntem, DRL-EC3'e göre yaklaşık %8,64, Dağıtık_DRL-EC3'e göre ise %3,51 daha fazla kapsama oranı elde etmiştir. İkinci deney çalışmasında modellerin enerji tüketimleri kıyaslanmıştır. Elde edilen sonuçlara göre DRL-EC3 ve Dağıtık_DRL-EC3 önerilen yönteme göre sırasıyla yaklaşık %8,7 ve %3,75 daha fazla enerji tüketmiştir. Son deney çalışması adillik indisi üzerine yapılmıştır. Önerilen modelin sırasıyla yaklaşık %10,3 ve %1,4 ortalama artışla DRL-EC3 ve Dağıtık_DRL-EC3'e göre daha iyi adillik indisine sahip olduğu görülmüştür. Bu sonuçlara göre, dinamik alanda, değişen ajan sayısına uyum sağlanarak iletişim kısıtlamaları altında yüksek adillik indisine sahip politikalar üretilmiştir. Önerilen yöntem, yerel gözlemler kullanılarak modelsiz politika gradyanı tarzında modellenmiştir. Klasik konum tabanlı stratejilerde olduğu gibi hedef alanda ızgara kapsamaya odaklanılmıştır. Öte yandan önerilen yöntemde durum uzayı yerel gözlemlere bağlıdır; böylelikle büyüyen eylem-durum uzayı ortadan kaldırılmıştır. Bu yaklaşım, doğru politikanın daha kısa sürede üretilmesine yardımcı olmuştur. Tasarlanan ödül yapısı, ödül paylaşımına gerek kalmadan kolektif davranış üretilmesine olanak sağlamıştır. Merkezi kontrolden bağımsız yerel gözlemlere dayalı bu yaklaşım, sahip olduğu ödül stratejisi ile gerçek uygulamalarda alan kapsama çözümlenmesine yardımcı olacak niteliktedir. Gelecek çalışmada, önerilen yöntem içerisinde kullanılan kümülatif ödül işlevinin geliştirilmesi hedeflenmektedir. Her bir ödül işlevinin öğrenme üzerindeki bireysel etkisi analiz edilerek, ağırlık verme yöntemleri ile öğrenme süreci iyileştirilmeye çalışılacaktır.

Kaynaklar (References)

1. Gupta, H., Verma, O.P., Monitoring and Surveillance of Urban Road Traffic Using Low Altitude Drone Images: A Deep Learning Approach, Multimedia Tools and Applications, 81 (14), 19683–19703, 2022.
2. Lee H-R., Lee T., Multi-agent Reinforcement Learning Algorithm to Solve a Partially-observable Multi-agent Problem in Disaster Response, Eur. J. Oper. Res., 291 (1), 296-308, 2021.
3. Drew, D.S., Multi-Agent Systems for Search and Rescue Applications, Curr Robot Rep, 2 (2), 189-200, 2021.
4. Xiao J., Wang G., Zhang Y., Cheng L., A Distributed Multi-Agent Dynamic Area Coverage Algorithm Based on Reinforcement Learning, IEEE Access, 8 (1), 33511-33521, 2020.
5. Dorri A., Kanhere S. S., Jurdak R., Multi-Agent Systems: A Survey, IEEE Access, 6, 28573-28593, 2018.
6. Woolley A.W., Aggarwal I., Malone T.W., Collective Intelligence and Group Performance. Current Directions in Psychological Science, 24 (6), 420-424, 2015.
7. Gupta, S.K., Kuila, P., Jana, P.K., Genetic Algorithm Approach for K-coverage and M-connected Node Placement in Target Based Wireless Sensor Networks, Computers & Electrical Engineering, 56 (1), 544-556, 2016.
8. Njoya A.N., Ari A.A.A., Awa M.N., Titouna C., Labraoui N., Effa J.Y., Abdou W., Gueroui A., Hybrid Wireless Sensors Deployment Scheme with Connectivity and Coverage Maintaining, Wireless Personal Communications, 112 (3), 544-556, 2020.
9. Yue Y., Cao L., Luo Z., Hybrid Artificial Bee Colony Algorithm for Improving the Coverage and Connectivity of Wireless Sensor Networks, Wireless Personal Communications, 108 (3), 1719–1732, 2019.
10. Jagtap A.M., Gomathi N., Minimizing Movement for Network Connectivity in Mobile Sensor Networks: An Adaptive Approach, Cluster Computing, 22 (1), 1373–1383, 2019.
11. Shu T., Dsouza K.B., Bhargava V., Silva C., Using Geometric Centroid of Voronoi Diagram for Coverage and Lifetime Optimization in Mobile Wireless Sensor Networks, IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), Edmonton AB-Kanada, 1-5, 05-08 Mayıs, 2019.
12. Shi, W., Li, J., Xu, W., Zhou, H., Zhang, N., Zhang, S., Shen, X., Multiple Drone-cell Deployment Analyses and Optimization in Drone Assisted Radio Access Networks, IEEE Access, 6 (1), 12518-12529, 2018.
13. Mozaffari M., Saad W., Bennis M., Debbah M., Efficient Deployment of Multiple Unmanned Aerial Vehicles for Optimal Wireless Coverage, IEEE Commun. Lett., 20 (8), 1647-1650, 2016.
14. Zhang X., Duan L., Fast Deployment of UAV Networks for Optimal Wireless Coverage, IEEE Trans. Mob. Comput., 18 (3), 588-601, 2019.
15. Sun J., Masouros C., Drone Positioning for User Coverage Maximization, IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Bologna-Italya, 318-322, 09-12 Eylül, 2018.
16. Cabreira T.M., Ferreira P.R., Franco C.D., Buttazzo G.C., Grid-Based Coverage Path Planning With Minimum Energy Over Irregular-Shaped Areas With Uavs, International Conference on Unmanned Aircraft Systems (ICUAS), Atlanta GA-ABD, 758-767, 11-14 Haziran, 2019.
17. Kalantari E., Yanikomeroglu H., Yongacoglu A., On the Number and 3D Placement of Drone Base Stations in Wireless Cellular Networks, IEEE 84th Vehicular Technology Conference (VTC-Fall), Montreal QC-Kanada, 1-6, 18-21 Eylül, 2016.
18. Chiu J-H., Kuo Y-C., Sheu J-P., Hong Y-W. P., Energy-Efficient UAV Deployment and IoT Device Association in Fixed-Wing Multi-UAV Networks, IEEE Global Communications Conference, Taipei-Tayvan, 1-6, 07-11 Aralık, 2020.
19. Ganganath N., Cheng C., Tse C.K, Distributed Antiflocking Algorithms for Dynamic Coverage of Mobile Sensor Networks, IEEE Trans. Ind. Inf., 12 (5), 1795-1805, 2016.
20. Krajník T., Nitsche M., Faigl J., Vaněk P., Saska M., Přeučil L, Duckett T., Mejail M., A Practical Multirobot Localization System, Journal of Intelligent & Robotic Systems, 76 (3), 539–562, 2014.
21. Abidin H. Z., Din N.M., Yassin I.M., Omar H.A., Radzi N.A.M, Sadon S.K., Sensor Node Placement in Wireless Sensor Network Using Multi-objective Territorial Predator Scent Marking Algorithm, Arabian Journal of Science and Engineering, 39 (1), 6317–6325, 2014.
22. Liu C.H., Chen Z., Tang J., Xu J., Piao C., Energy-Efficient UAV Control for Effective and Fair Communication Coverage: A Deep Reinforcement Learning Approach, IEEE J. Sel. Areas Commun., 36 (9), 2059-2070, 2018.
23. Lillicrap T.P., Hunt J.J, Pritzel A., Heess N., Erez T., Tassa Y., Silver D., Wierstra D., Continuous Control with Deep Reinforcement Learning, 4th International Conference on Learning Representations (ICLR), San Juan-Porto Riko, 1-14, 02-04 Mayıs, 2016.
24. Liu C.H., Ma X., Gao X., Tang J., Distributed Energy-Efficient Multi-UAV Navigation for Long-Term Communication Coverage by Deep Reinforcement Learning, IEEE Trans. Mob. Comput., 19 (6), 1274-1285, 2020.
25. Aydemir F., Çetin A., Multi-agent Dynamic Area Coverage Based on Reinforcement Learning with Connected Agents, Computer Systems Science and Engineering, 45 (1), 215–230, 2023.
26. Lowe R., Wu Y., Tamar A., Harb J., Abbeel P., Mordatch I., Multiagent Actor-critic for Mixed Cooperative-competitive Environments, Advances in Neural Information Processing Systems, Long Beach CA-ABD, 6379-6390, 04-09 Aralık, 2017.
27. Keith A.J., Ahner D.K., A Survey of Decision Making and Optimization Under Uncertainty, Annals of Operations Research, 300 (2), 319-353, 2021.
28. Deng L., Yu D., Deep Learning: Methods and Applications, Foundations and Trends in Signal Processing, 7 (3-4), 197-387, 2014.
29. Song H.A., Lee S. Y., Hierarchical Representation Using NMF, International Conference on Neural Information Processing (ICONIP), Daegu-Güney Kore, 466-473, 03-07 Kasım, 2013.

30. Qie H., Shi D., Shen T., Xu X., Li Y., Wang L., Joint Optimization of Multi-UAV Target Assignment and Path Planning Based on Multi-Agent Reinforcement Learning, IEEE Access, 7 (1), 146264-146272, 2019.

31. Jianqing F., Zhaoran W., Yuchen X., Zhuoran Y., A Theoretical Analysis of Deep Q-Learning, Proceedings of the 2nd Conference on Learning for Dynamics and Control, Berkeley CA-ABD, 486-489, 11-12 Haziran, 2020.

32. Zoss B.M., Mateo D., Kuan Y.K., Toki´c G., Chamanbaz M., Goh L., Vallegra F., Bouffanais R., Yue D.K., Distributed System of Autonomous Buoys for Scalable Deployment and Monitoring of Large Waterbodies, Autonomous Robots, 42 (8), 1669-1689, 2018.

33. H¨uttenrauch M., ˇSoˇsi´c A., Neumann G., Local Communication Protocols for Learning Complex Swarm Behaviors with Deep Reinforcement Learning, 11th International Conference on Swarm Intelligence (ANTS), Roma-İtalya, 71-83, 29-31 Ekim, 2018.

34. Özçevik Y, Solmaz Ö., Baysal E., Ökten M., A real-time simulation environment architecture for autonomous vehicle design, Journal of the Faculty of Engineering and Architecture of Gazi University, 38 (3), 1867-1878, 2023.

35. Jain R.K., Chiu D.M.W., Hawe W.R., A QuantitativeMeasure of Fairness And Discrimination for Resource Allocation In Shared Computer Systems, Eastern Research Laboratory Digital Equipment Corporation, 38 (1), 1984.