

Ebabil

GNSS-denied ortamlarda sürü kamikaze İHA karar verme ve görev koordinasyonu için çoklu ajan pekiştirmeli öğrenme sistemi

Multi-Agent RL	MAPPO / CTDE	Gymnasium	GNSS-Denied	3D Replay
----------------	--------------	-----------	-------------	-----------

Proje Sahibi	Yakup Tekeli
Belge Türü	Teknik Proje Raporu
Durum	Tamamlandı
Tarih	Temmuz 2026

Proje Özeti

Ebabil, GNSS erişiminin bulunmadığı veya güvenilir olmadığı görev ortamlarında birden fazla kamikaze İHA'nın ortak hedefe yönelmesini, engellerden kaçınmasını, görev paylaşımı yapmasını ve angajman kararlarını koordine etmesini sağlayan simülasyon tabanlı bir çoklu ajan pekiştirmeli öğrenme sistemidir. Proje; özel simülasyon ortamı, hiyerarşik MAPPO/CTDE karar mimarisi, ayrıştırılmış ödül sistemi, deney izleme, değerlendirme ve 3B görev tekrar katmanlarını bütünleştirir.

1. Problem Tanımı ve Hedefler

- Merkezi GNSS bilgisinin olmadığı dinamik bir görev alanında sürü navigasyonu gerçekleştirmek.
- Her İHA için yerel gözlem kullanırken eğitim aşamasında merkezi bilgiyle koordineli politika öğrenmek.
- Çarpışma önleme, hedef yaklaşımı, formasyon kararlılığı ve görev başarısını aynı politika içinde dengelemek.
- Ödül fonksiyonu bileşenlerinin etkisini ayrı ayrı ölçmek ve farklı reward tasarımlarını karşılaştırmak.
- Eğitim sonucunu gerçek zamanlı 3B sahnede, drone rotaları ve olay durumlarıyla yeniden oynatmak.

2. Üst Düzey Sistem Mimarisi

Senaryo	Çoklu Ajan Ortamı	Hiyerarşik Politika	Reward ve Eğitim	Değerlendirme	3B Tekrar
Harita, hedef, engel, EW parametreleri →	Durum, gözlem, dinamik, termination →	MAPPO + CTDE →	Bileşen logları, checkpoint →	Başarı, çarpışma, verimlilik →	Path, olay ve görev görselleştirme

Sistem, eğitim ve görselleştirme katmanlarını ayrıştırır. Politika özel Gymnasium ortamında eğitilir; eğitim sonrası durum/eylem kayıtları hafif 3B simülasyon katmanında yeniden oynatılır. Böylece görsel motor, öğrenme sürecinden bağımsız kalır.

3. Çoklu Ajan Ortam Tasarımı

Bileşen	Teknik Açıklama
Ajan modeli	Her drone bağımsız aktör olarak temsil edilir; ortak görev bağlamı eğitim sırasında merkezi critic tarafından kullanılabilir.
Durum uzayı	Drone konumu/hızı/yönelimi, hedef görelî konumu, komşu ajan bilgisi, engel mesafeleri ve görev durumu.
Gözlem uzayı	Ajanın karar anında erişebildiği yerel ve görelî bilgiler; GNSS gerektirmeyen temsil.
Eylem uzayı	Hız/yön değişimi veya normalize edilmiş hareket komutları; görev modu seçimi için üst seviye kararlar.
Dinamikler	Zaman adımlı kinematik güncelleme, sınır kontrolü, ajan-ajan ve ajan-engel çarpışma denetimi.
Episode sonlandırma	Hedef başarısı, çarpışma, görev alanı dışına çıkma veya maksimum adım sınırı.
Deterministik deney	Seed yönetimi ile senaryo üretimi, başlangıç konumları ve değerlendirme koşulları tekrar üretilebilir.

4. Öğrenme Mimarisi

4.1 Hiyerarşik Karar Yapısı

- Üst seviye politika; seyir, kaçınma, yeniden gruplanma ve angajman gibi görev modlarını seçer.
- Alt seviye politika; seçilen moda göre sürekli veya ayrık hareket komutlarını üretir.
- Bu ayrım, uzun ufuklu görev kararları ile kısa vadeli manevra kontrolünü birbirinden ayırır.

4.2 MAPPO ve CTDE

- Aktörler, çalışma sırasında yalnızca kendi gözlemleriyle eylem üretir.
- Merkezi critic, eğitim sırasında sürünün ortak durum bilgisini kullanarak değer tahminini daha kararlı hâle getirir.
- PPO güncellemesi clipped objective, advantage normalization ve entropy düzenlemesiyle gerçekleştirilir.
- Ajanlar arasında parametre paylaşımı kullanılarak politika boyutu ve eğitim maliyeti kontrol altında tutulur.

4.3 Klasik Yöntemlerin Rolü

A*, PID/MPC ve geometrik yol planlama yöntemleri ana karar mekanizması olarak değil; karşılaştırma, başlangıç politikası, öğretmen sinyali veya güvenlik katmanı amacıyla kullanılmıştır. Bu sayede öğrenilmiş politikanın performansı klasik yöntemlerle ölçülebilir.

5. Ödül Fonksiyonu Tasarımı

Reward Bileşeni	İşlev
Hedef ilerleme	Ajanın hedefe olan görelî mesafesindeki azalma için yoğun shaping sinyali.

Reward Bileşeni	İşlev
Görev başarısı	Hedefe ulaşma veya tanımlı angajman koşulunu tamamlama için yüksek terminal ödül.
Çarpışma cezası	Engel veya diğer ajanlarla çarpışma için yüksek negatif terminal/ara ceza.
Yakınlık güvenliği	Çarpışma gerçekleşmeden önce kritik mesafeye girme durumunu cezalandıran sürekli bileşen.
Zaman maliyeti	Gereksiz bekleme ve dolaşmayı azaltmak için adım başına küçük negatif maliyet.
Yol verimliliği	Aşırı rota uzaması, keskin yön değişimi veya kararsız manevralar için düzenleyici bileşen.
Sürü koordinasyonu	Ajanların görev paylaşımı, kapsama veya uygun komşuluk davranışını teşvik eden ekip bileşeni.

5.1 Ayrıştırılmış Reward Günlüğü

Toplam episode reward tek başına başarı ölçütü olarak kullanılmamıştır. Her reward bileşeni ayrı olarak TensorBoard, CSV ve JSONL formatlarında kaydedilmiş; böylece politika davranışındaki değişimin hangi teşvikten kaynaklandığı izlenebilmiştir.

5.2 Reward Sürümleri ve Karşılaştırma

- İlerleme odaklı, güvenlik ağırlıklı ve dengeli olmak üzere birden fazla reward preset'i tanımlanmıştır.
- Her preset aynı seed ve senaryo dağılımında eğitilip ortak değerlendirme prosedürüyle karşılaştırılmıştır.
- Aktif olmayan veya senaryoda etkisiz kalan reward terimleri applicability denetimiyle işaretlenmiştir.

6. Deney ve Yazılım Altyapısı

Bileşen	Teknik Açıklama
Konfigürasyon	Senaryo, seed, reward ağırlıkları, eğitim hiperparametreleri ve değerlendirme ayarları dosya tabanlı olarak saklanır.
Deney klasörü	Her çalışmada konfigürasyon anlık görüntüsü, run metadata, model checkpoint, log ve değerlendirme raporu üretilir.
Test altyapısı	Environment sözleşmeleri, reward bileşenleri, logging, evaluation ve eğitim giriş noktaları için otomatik testler.
Leaderboard	SQLite tabanlı deney kayıt sistemi; kompozit puanla reward sürümleri ve politikalar sıralanır.
Çapraz platform	MacOS üzerinde geliştirme/smoke test, Linux RTX 3060 sisteminde GPU eğitim iş akışı.

7. Değerlendirme Metodolojisi

Metrik	Tanım
Görev başarı oranı	Tanımlanan hedef veya angajman koşulunu tamamlayan episode yüzdesi.
Çarpışma oranı	Ajan-ajan ve ajan-engel çarpışmalarının episode bazında oranı.
Yol verimliliği	Gerçek rota uzunluğunun referans/geometrik minimum rotaya oranı.
Tamamlama süresi	Başarılı episode'ların ortalama adım veya simülasyon zamanı.
Kontrol kararlılığı	Aşırı yön/hız değişimleri, salınım ve gereksiz manevra ölçümleri.
Sürü performansı	Tek ajan başarısından bağımsız olarak tüm sürünün ortak görevi tamamlama kalitesi.

Leaderboard kompozit puanı, yalnızca ortalama reward'a dayanmaz. Başarı, güvenlik, yol verimliliği ve kararlılık birlikte değerlendirilir. Bu yaklaşım, reward hacking sonucu yüksek puan alan ancak görevi güvenilir biçimde tamamlamayan politikaların öne çıkmasını engeller.

8. 3B Görev Tekrar ve Görselleştirme

- Her drone için zamana bağlı konum ve yönelim verisi kayıt altına alınır.
- Drone rotaları sahne üzerinde kalıcı çizgilerle gösterilir.
- Kaçınma manevrası, yeniden gruplanma ve angajmana geçiş gibi karar olayları zaman damgasıyla işaretlenir.
- Kamera, serbest dolaşım ve takip modlarında sürü davranışını incelemeyi sağlar.
- Görselleştirme motoru RL eğitiminden bağımsızdır ve kayıtlı episode verisini yeniden oynatır.

9. Tamamlanan Teknik Çıktılar

- Özel çoklu ajan navigasyon ve görev ortamı.
- Hiyerarşik MAPPO/CTDE eğitim ve çıkarım mimarisi.
- Birden fazla reward preset'i ve bileşen bazlı logging sistemi.
- Seed tabanlı tekrarlanabilir deney yönetimi ve otomatik test altyapısı.
- Başarı, güvenlik, verimlilik ve kararlılık odaklı değerlendirme modülleri.
- SQLite deney leaderboard'u ve politika karşılaştırma aracı.
- Drone path'leri ve görev olaylarını gösteren bağımsız 3B tekrar katmanı.
- MacOS geliştirme ve Linux GPU eğitim sistemlerini birleştiren çapraz platform iş akışı.

10. Sonuç

Ebabil, çoklu İHA görev koordinasyonunu yalnızca bir eğitim algoritması olarak değil; ortam, reward tasarımı, gözlemlenebilirlik, tekrarlanabilirlik, değerlendirme ve görsel doğrulama bileşenleriyle birlikte ele alan tamamlanmış bir araştırma ve prototipleme sistemidir. CTDE tabanlı hiyerarşik öğrenme yaklaşımı, ajanların çalışma sırasında merkezi konum bilgisine bağımlı olmadan sürü davranışı üretmesini sağlar.