1. GİRİŞ

İnsanların kendi başlarına kolaylıkla yapabildiği bazı davranışları, verdiği kararları veya en basitinden daha çocukluktan itibaren gördüğü her nesneyi bir sınıfa koyabilme becerisini anlayıp bunu gerçekleyen bir sistem oluşturmak hiç de kolay bir iş değildir. Çoğu zaman farkında bile olmadan saniyeler için de verdiğimiz kararları bazen nasıl verdiğimizden, neyi düşünerek ya da neyi göz ardı ederek bunu yaptığımızdan bile haberdar değiliz. Ancak bilinçli olarak farkında olmasak da her davranışın ardında mutlaka bir plan ve bir fayda gözetme söz konusudur. Hayatımız bunun doğrultusunda devam etmektedir. Biliş bilimin çalışma konusu da aslında tam olarak budur. Psikolojik ve bilişsel olayların ve bunun sonucunda da insanda meydana gelen algı, hafıza, karar verme ve dikkat gibi süreçlerin nasıl gerçekleştiğini, bu zihinsel süreçlerin altındaki biyolojik temelleri açıklamaya çalışır [1]. Bu tür sistemleri oluşturmakta da kimi zaman yapay sinir ağı yapıları kullanılmaktadır.

Her canlı için var olan ödül mekanizması aslında bir nöron aktivasyonundan ibarettir. Bu nöron aktivasyonunun oluşmasında etkin olan sinir sistemi alt yapıları ve mekanizmalar incelenmiş [2, 3] ve bu mekanizmaların pekiştirmeli öğrenmede etkin bir yöntem olan zamansal fark (Temporal Difference (TD)) ile ifade edilebileceği fark edilmiştir [2, 3, 4, 5]. Pekiştirmeli öğrenme yapay sinir ağları çerçevesinde kullanılarak bir sistem için "öğrenme" kavramının geliştirilmesinde etkin bir hal almıştır. Artık bir canlı için saniyeler içerisinde verilen kararlar ya da sergilenen davranışlar, benzer şekilde bir mühendislik sisteminin işlevi olarak yer almıştır. Ancak bir durum karşısında bir davranışı öğrenmiş olmak bir canlı ya da bir sistem için yeterli değildir. Burada önemli olan, ki problemlerin çoğunda aynı durum söz konusudur, bir dizi işlem sonunda istenilen noktaya ulaşmaktır. Zaten birçok problem içerisinde bir den fazla aşama yani durum söz konusudur ve sergilenen her davranış bir başka durumunun nedenidir. Tek bir aşama ile sonuca ulaşmak çoğu karmaşık sistem için mümkün değildir. Bu nedenle tasarlanan öğrenme kuralları ve modellemeler bir ardışıl problem içerisinde

kullanılmaktadır. Her bir durumda çözüm üretebilen ya da her bir duruma cevap verebilen bir sistem için de yapılması gereken zaten budur.

Bu bitirme çalışmasında ilk olarak ele alınan, pekiştirmeli öğrenmenin ve dolayısıyla da ödül mekanizmasının bir makina öğrenmesi yöntemi olarak kullanılmasına dayalıdır. Bu aşamada ele alınan problem, bir arabanın belli sınırlar içerisinde, üzerine menteşelenmiş çubuğun da belli açılar içerisinde kalarak hareket etmesini sağlayan araba-çubuk (cart-pole) problemidir. Bu problemin çözümünde kullandığımız yöntem, daha önce yapılmış çalışmalardan [5] farklı olarak, kutu yöntemi yerine durum değişkenlerinin doğrudan tasarlanan ağın girişlerini oluşturmasıdır. Bunun için kullanılacak olan diferansiyel denklemler Euler metodu ile çözülmüş ve benzetim için gerekli olan MATLAB kodu en baştan oluşturulmuştur.. Bu şekilde ağın girişindeki, kuantalama yapılarak elde edilmiş fazla sayıdaki durum yani giriş yerine sadece durum değişkenlerinin yer aldığı girişlerden oluşan daha basit ve daha anlaşılır bir yapı tasarlanmıştır. ~~Yazdığım~~ Yazılan kodun benzetiminde ise, farklı ilk koşullar ve durumlar göz önüne alınarak çok defa denemeler yapılmış ve sonuçlar değerlendirilip olumlu ve olumsuz taraflar ayrıntısıyla açıklanmıştır. Daha sonrasında, canlı vücudundaki dopamin aktivasyonunu ödül mekanizması ile birleştirerek elde edilen "TD" yönteminden faydalanarak, karar verme ve ardışıl öğrenme işlemleri gerçekleştirilmiştir.

Karar verebilmenin önemi, bir sistem ya da çevre içerisinde değişen ortam koşullarına uyum sağlayabilme açısından önemlidir ve bunun sonucunda alınan kararlar çevrede bir takım değişikliklere ve yeni durumların oluşmasına neden olacaktır. Bunun yanında yine gerçekte var olan sistemlerde olduğu gibi, her bir duruma ilişkin bir karar alıp bir davranışta bulunmak da tek başına anlamlı değildir. Tıpkı alınan kararların daha öncekilerle ilişkilendirilmesi gibi öğrenmede ve davranışlarda da bir ilişki olmalıdır. Bu noktada da ardışıl öğrenme devreye girecektir. Her bir durumda sergilenen davranışlar, ortamda bir değişikliğe yol açacak ve bunun sonucunda oluşan yeni durumda yine doğru bir davranışta bulunabilmek söz konusu olacaktır. Dolayısıyla her bir davranış, bir sonraki durumu ve davranışı etkilemektedir veya tersi şekilde, bir davranış sergilenecekken daha öncekilerin de tekrar göz önüne alınması gerekecektir. Bitirme çalışmasında da ödül işaretinden faydalanarak ardışıl öğrenme testi gerçekleştirilecektir. [6]. Burada da daha önce yapılmış bu çalışmadan farklı olarak, durum sayısı ve kullanılan

örüntülerin boyutları arttırılmış, son aşamada ise ART yapısı da [7] koda dahil edilerek MATLAB kodu yeniden düzenlenmiş ve ardışıl öğrenme testi bir de bu şekilde gerçekleştirilmiştir.

Burada artık karar verici olarak bir birey değil de bir mekanizma ve içinde bulunduğu bir ortam söz konusudur. Aslında yapılan iş sadece bir birey ya da aktör ve çevre ilişkisine dayanmaktadır. Bir birey için, zihninden geçen onlarca düşünce ya da bir canlı için vücutta gerçekleşen onlarca ya da binlerce biyolojik olay bu anlamda göz önüne alınmalıdır. Bu çalışmada da, bir karar verme sürecinde nöronların aktivasyonlarından yola çıkarak, bir mühendislik sisteminde benzer işleri yapmayı planlıyoruz. Bir sınıflandırma problemi içerisinde seçilecek benzerlik kriterinin değeri yani benzerlik derecesi, bir nevi karar verme yeteneğiyle ve değişen ortam ve sistem koşullarına göre belirlenecektir. Bunu gerçeklemek için iki farklı yapı bir arada ele alınacaktır. Değişen ortamı değerlendirme pekiştirmeli öğrenme ile gerçeklenirken, sınıflandırma süreci ART (uyarlanabilir yankılaşım teorisi) ile gerçeklenecektir [7]. Bu iki yapının birleştirileceği bu aşamada ise ilgili kodlar uygun şekilde düzenlenerek birleştirilmiş ve daha fazla boyutta testler gerçekleştirilebilecek şekilde de yeniden değiştirilerek güncellenmiştir.

Kısacası, yapay sinir ağları canlılara dair karar mekanizmasını ve öğrenme biçiminin modellenmesinde kullanılan bir yapı olduğundan geniş bir uygulama alanına sahiptir ve bunun sonucunda da meydana gelmiş birçok yöntemden oluşmaktadır. Bizim problemimizde ise en çok kullanılan yöntemlerden biri olan pekiştirmeli öğrenme kullanılacaktır. Burada ele alınan öğrenme modelinin insan ya da canlı metabolizmasındaki ve psikolojideki karşılığı, dayandığı temeller ikinci bölümde açıklanacaktır. Üçüncü bölümde, bir kontrol problemi olan araba-çubuk problemi çözülerek, kurulan düzeneğin istenen sınırlar içerisinde kalması sağlanmaya çalışılacaktır. Dördüncü bölümde de, dopamin maddesinin yani ele alınan modeldeki karşılığı olan ödül işaretinin ardışıl öğrenmede ne şekilde kullanıldığını açıklamak üzere ardışıl öğrenme testi gerçekleştirilecektir. Bunun yanında benzer bir ardışıl öğrenme testinin, bir başka yapay sinir ağı yapısı olan ART da kullanılarak ne şekilde gerçekleştirilebileceği de beşinci bölümde açıklanarak sonuçlar değerlendirilecektir. Sonuç bölümünde ise, bu çalışmada ele alınan öğrenme modelinin ve yöntemlerinin olumlu ve eksik tarafları tartışılacaktır.

2. AMACA YÖNELİK DAVRANIŞ

2.1 Tanım

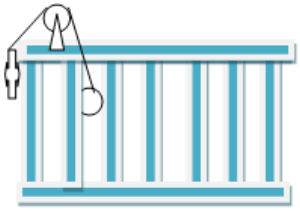
Canlılar bilinmeyenlerle dolu dünyada ve dolayısıyla da bulunduğu her ortamda, karşılaştığı her duruma karşılık bir tepki oluştururlar. Her durum karşısında davranışta bulnur ve bunun sonucunda da farklı durumlarla karşılaşır. Bu davranışların neye göre oluştuğu ve nasıl sergilendiği psikolojinin ilgi alanı içinde yer almaktadır ve biliş bilimde de bu sorulara yanıt aranmaktadır.

Canlıların davranışları bu şekilde incelenmek istenildiğinde ilk olarak "klasik koşullanma" üzerinde durulmuştur. Burada canlılar çevresel faktörlerden dolayı oluşan uyaranlara göre bir davranış sergilemektedir. Örneğin, Pavlov' un yaptığı deneylerde olduğu gibi, deneğin bulunduğu ortama uygulanan bir düdük sinyalinin ardından köpeğe besin verilmesi olayı birkaç defa tekrarlandığında, köpek artık her düdük sesini duyduğunda ayağa kalkar ve ağzını açıp dilini çıkararak besin bekliyor duruma geçer. Köpeğin bu şekilde davranışı ve olaylar karşısındaki tepkisi sadece klasik şartlanma ile açıklanmaktadır [8]. Ancak bu şekilde bir koşullanma bizim ele almak ve incelemek istediğimiz davranışlara denk düşmemektedir çünkü burada davranışta bulunan canlı, sergilediği davranış ile ortamda herhangi bir değişikliğe yol açmamaktadır. Canlı tamamen bir uyaran sonucunda tek bir davranışı sergilemeye odaklanmış yani şartlanmıştır. Oysa bizim incelemk istediğimiz ve sonraki modellemelerimizde kullanacağımız davranış biçimi, ortamı değiştirerek ve ortamdan alınan sonuçları göz önüne alarak davranışları sergilemektir. Yani, gelen uyaranlara karşılık olarak belirli davranışlar sergiledikten sonra, elde edilen sonuçlara göre ve daha sonrasında gelecek olan farklı uyaranlara karşı yeni ve doğru bir davranışta bulunabilmeyi sağlamaktır. Bu noktada da "amaca yönelik davranış (goal directed behavior)" devreye girmektedir. Bu şekilde davranışların nedenlerini açıklamaya ve psikolojik anlamda bunu bazı anlamlı dayanaklarla sunmaya yönelik çok sayıda çalışma yer almaktadır [8].

Amaca yönelik davranış denildiğinde 3 temel unsur göz önüne alınmalıdır [8];

* Bir amaç doğrultusundaki davranış
* Davranış sonunda olması beklenen bir sonuç
* Bu davranış ve sonuç arasındaki ilişki

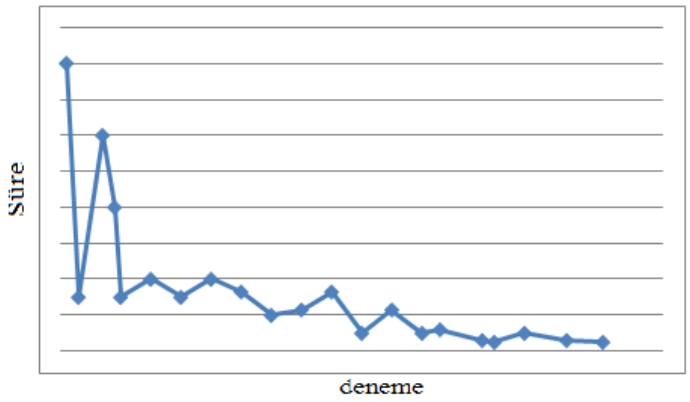
Bulunulan durum altında bir davranış sergilenir ve bunun sonucunda ortamda bir değişim meydana gelir ve davranış bir sonuç doğurur. Daha sonra bu sonuç ve davranış arasında bir ilişkilendirme yapılarak, eğer istenilen sonuçlar elde ediliyorsa benzer davranışlar sergilenmeye devam edilir. Eğer çok çalışmak başarı getiriyorsa, aynı şekilde çalışılmaya devam edilir. Bu konuyla ilgili olarak, kediler üzerinde yapılan bir çalışma şu şekildedir [8];



Şekil 2.1. Kedinin amaca yönelik davranışını incelemek üzere kurulmuş halkalı kafes düzeneği [8].

Şekil 2.1' de görüldüğü gibi kapalı kutu içerisinde bir kedi bulunmakta, içeriye sarkıtılan halkayı çektiğinde ise kurulan düzenek sayesinde kapı açılmakta ve kedi serbest kalmaktadır. Bunla ilgili olarak yapılan denemelerde kedinin ilk defa serbest kalması yaklaşık olarak 160 saniye sürmüştür. Daha sonraki denemelerde ise kedi tekrar aynı kafese konulduğunda kafesten kurtulma süresi git gide azalmış ve son denemelere doğru 6 saniye civarına kadar inmiştir.

Bu araştırma sonunda varılan sonuç ise, kedi aslında sistemin çalışma mekanizmasını ya da kafesten nasıl kurutulacağını kavramıyor. Yaptığı şey sadece halka ile kafesten kurtulmayı ilişkilendirmesidir. Artık onun için, halkayı çekmek bir davranış ve bu davranışı sergilemesinin nedeni de kafesten kurtulma sonucuna erişmektir. Bu şekilde davranışlar aslında elde edilen ya da edilecek sonuçlar doğrultusunda oluşmaktadır.



Şekil 2.2. Kedinin kafesten çıkma süresinin zaman göre değişim grafiği. İlk başlarda 160 saniye civarında olan bu süre, yapılan denemeler sonunda zamanla azalarak ve kedinin davranışlarıyla sonuçları ilişkilendirmesi sonucu 6 saniyeye kadar inmektedir. Kurulan düzenekten her kurtulmada sergilenen davranış ve onun sonucu arasındaki ilişki daha da kuvvetleniyor ve böylece kafesten çıkmak için sarf edilen süre azalıyor [8].

Şekil 2.2' den de görüldüğü gibi davranışlar ve verilen kararlar elde edilen sonuçlar doğrultusunda şekillenmektedir. Eğer bir karar sonucunda olumlu sonuçlarla karşılaşmak isteniyorsa elde edilen sonuçlar doğrultusunda davranış sergileyip, bu davranışın sonucunda gelecekte oluşabilecek durum ve sonuçlar hakkında doğru tahminlerde bulunabiliyor olunmalıdır. Bu şekilde bir davranış-ödül ilişkisi kurulabilir ve farklı durumlarla karşılaşıldığında dahi olumlu sonuçlar bir davranış sergilenebilir. Ancak bu şekilde bir yaklaşım her zaman iyi bir sonuç doğuracak anlamına gelmez. Bu nedenle, beklenen durumlarla karşılaşılmadığında yani seçilen davranışın ödülle sonuçlanmadığı durumlarda da yine bu elde edilen sonuç ve beklenti arasındaki farktan ya da hatadan yararlanarak davranış-sonuç ilişkileri güncellenmeli, sonraki durumlardaki tahminler de buna göre yapılmalıdır.

Canlıların davranış prensiplerinden biri olan bu mekanizma psikolojide pekiştirmeli öğrenme olarak adlandırılır. Bu prensipten yararlanarak mühendislik sistemlerinde de "akıllı" davranan yapılar gerçeklenmeye çalışılmaktadır.

2.2 Modelleme

Bir canlıya ait karar mekanizmasının sayısal ifade ve denklemlerle elde edilmesi veya bir sistemin başarıya ulaşmasını sağlayacak bir davranış serisinin bu şekilde gerçekleştirilmesi oldukça önemli bir problemdir. Bu noktada yapılan çalışmalarda canlının davranışına yön veren içsel ödül mekanizmasındaki dopamin nöronlarının etkinliği ön plana çıkmaktadır [2]. Bu nöronlar, aksonlarını beyindeki motivasyon, ödül ve amaca yönelik davranış ile ilgili kısımlara göndererek bir etki meydana getirmeye neden olurlar.

Canlının içinde bulunduğu şartları ve karşılaştığı sonuçları yorumlaması, bünyesinde dopamin nöronlarının aktivasyonu ile açıklanmaktadır. Bu şekilde karşılaşılan sonucun, daha önce de bahsettiğimiz şekilde verdiğimiz kararlar veya yaptığımız tahminler ile ilişkisi ve uyumluluğu çok önemlidir. Tahminlerimiz çevreyle olan ilişkilerin modellenmesinde ve geliştirilmesinde önemli bir etkendir. Beklenenden daha iyi bir sonuç ile karşılaşmak ya da bir başka deyişle tahmin edilemeyen bir sonuç ile karşılaşmak, canlı vücudunda dopamin nöronlarının aktivasyonunu arttırır. Bu şekilde meydana gelen ödül, pozitif yönde zorlayıcı olarak rol oynar ve bu da sergilenen davranışın yapılma sıklığının artmasına ve bir süre sonra da alışkanlık haline gelmesine neden olur [2]. Aksi bir durumda ise bu nöronlarının aktivasyonu zayıflar, hatta negatif bir yönde bir sonuç doğurur. Bu ikisinin ortasında, eğer zaten tahmin edilen bir durumla karşılaşılmış ise, herhangi bir tepki ile karşılaşılmaz yani dopamin nöronlarının aktivasyonu sabit kalmış olur.

Ayrıca da her bir uyaranı ya da bir dinamik sistem için düşünürsek her bir sistem girişini tek başına bir davranışın nedeni olarak düşünmek eksik ele almak olur. Aslında her bir uyaran, daha önce sınanalar doğrultusunda oluşturulmuş bir fonksiyonun bileşenlerinden ibarettir. Zaten bir davranışı sadece bir uyaranın yanıtı olarak düşünecek olursak bu klasik şartlanmanın ötesine geçemez. Oysa burada üzerinde çalıştığımız öğrenme kavramı, içinde bulunulan durum ve koşullar neticesinde sergilenecek davranışın ve oluşturacağı sonucun, tam olarak tahmin edilemiyor ve bilinemiyor olmasından doğmaktadır. Klasik şartlanmada ses unsurunda sonra sürekli olarak besin verilmesi bir şartlanma meydana getirmektedir. Bir zaman sonra bu sesin yanında bir de ışık ya da koku vs. verilmesi ve bunun ardından besin sunulması, canlıdaki bu şartlanmada herhangi bir değişikliğe neden

olmaz. Burada ki ışık veya koku şeklinde uygulanan ikinci unsur canlı için hiçbir anlam ifade etmemektedir. İşte tam olarak ayırt edilmesi gereken nokta burasıdır. Amaç her türlü uyarı ya da faktörü hesaba katarak ileriye dair bir öğrenme meydana getirmektir. Bu noktada ödül de, tahmin edilen ve gerçekleşen sonuç arasındaki farktan dolayı oluşan negatif veya pozitif bir değerdir. Bu da bir durumun ya da sonucun, öğrenilenler sonucunda "iyiliğini" kavramaya ve değerlendirmeye yarar [2].

Tüm bu anlatılanlardan yola çıkarak, sistemlerin ya da canlıların çevreyle ilişkisini olumlu yönde ilerletmek amacıyla, davranışlar-sonuçlar doğrultusunda ve aynı vücuttaki dopamin nöronlarının aktivasyonlarını göz önüne alarak geliştirilen yöntem, "Temporal Difference" (TD) olarak adlandırılmıştır [2]. Burada esas olan, beklenen bir durum yani tahminden farklı bir sonuç ile karşılaşıldığında meydana gelen hatayı gözeterek öğrenme biçimi yaratmaktır [9]. Tahmin edilemeyen ya da beklenmeyen şeyler olduğunda aktif olan, tersi durumda negatif etki yaratan ve beklentiler doğrultusunda bir netice olduğunda da sabit kalan dopamin maddesi bu şekilde TD yöntemi ile modellenebilmiştir [9].

Bu modelleme içerisinde, canlıların kendi içerisinde sahip oldukları dopamin aktivasyonları ile mühendislik sistemleri içerisinde kullanılacak bu TD yöntemi arasında bir ilişkilendirme yapıldığında ilk olarak uyaranlar üzerinde durulabilinir. Burada canlıları için sürekli olarak değinilen uyartılar bir sistem yani çevre içerisinde var olan bir durum değişkenini, ele alınan modelde ve dolayısıyla da tasarlanan ağ yapısında ise bir girişi ifade etmektedir (x(t)). Bu durumların sayısı istenilen hedef doğrultusunda hareket edildiğinde, birden fazla sayıda oluşabilir ki aslında beklenen de budur. Aksi durumda tek bir durum için tek bir davranış sergileneceğinden bir öğrenmeden söz edilemez.

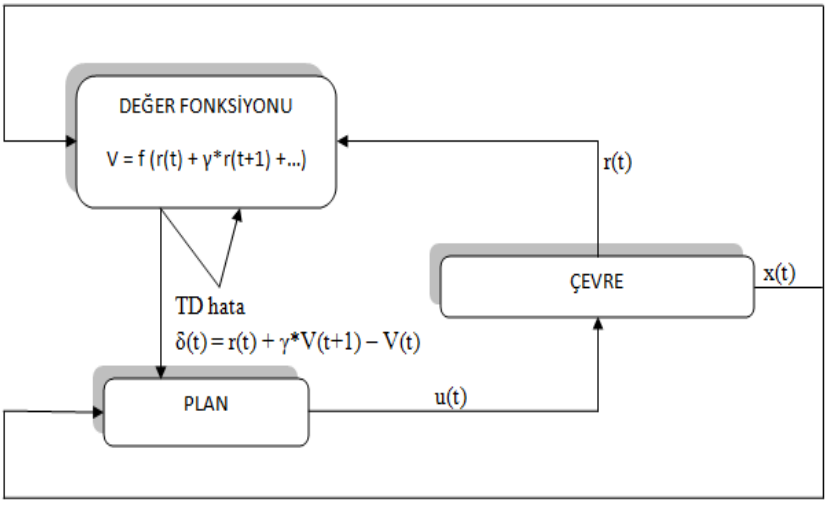
Her bir durum kendine ait bir ağırlığa (w(t)) sahiptir. Bunun anlamı, içinde bulunulan durumdan yola çıkarak bir davranış tahmininde bulunulacağından, bu durumları kendi içerisinde birbirlerinden üstün ya da daha az etkili kılacak şekilde bir sınıflandırmaya tabi tutmaktır. Bu aşamadan sonra sergilenen davranış, çevre tarafından yeni bir durum oluşmasının yanı sıra bir ödül değeri ile cevap bulacaktır. Bu gelen değerler işte, beklenen değerlerden yola çıkarak, bize her bir durum veya durum-davranış çifti için bir değer fonksiyonu (V(t)) belirlememizi sağlar.

Örneğin; 3 farklı renkte (mavi, yeşil, kırmızı) ışık ile aydınlatılmış bir ortam olduğunu varsayalım. Her bir ortamda bulunan besin miktarının da farklı olduğu (mavi > yeşil > kırmızı) farz edilsin [2]. Bir canlı için ya da genel olarak herhangi bir ortamda ya da bir sistemde amaç, ödülleri en fazla yapmak olduğuna göre, burada bir canlının mavi ışık ile aydınlatılan bölgeye ya da onun yakınındaki bir yere gelmesi, bulunduğu durumun değer fonksiyonunun da diğerlerine göre daha büyük olması anlamına gelmektedir. Benzer şekilde, mavi ile aydınlatılan bölgeden yani çok fazla besinin olduğu bölgeden uzaklaşılması halinde de, içinde bulunulan durumun değer fonksiyonu bir öncekine göre daha düşük olacaktır. Bu şekilde bir çıkarım, davranış seçimi ve öğrenmede çok önemli bir rol oynamaktadır.

Yapılan tahminler ve gerçekleşen cevaplar arasında bir fark yani hata meydana geldiğinde, sistemden gelen bu sonuçlar yorumlanarak ağırlıkları ve durum fonksiyonlarını günceller. Böylece belirlenen yeni bir davranış ile hatanın en aza indirilmesi amaçlanır. Bu yapılan iyileştirmenin son aşaması ise, artık sistemden gelen ödül ile tahmin edilen ödüllerin neredeyse birbirine eşit ya da çok yakın değerde olmasıdır. Bunun sonucunda bir iyileştirmeye ihtiyaç duyulmaz ve aynı davranış sergilenmeye devam edilir. Aynı bir canlıda olduğu gibi.

Eğer bir davranışın ya da verilen bir kararın sonunda beklenen, bir başka deyişle tahmin edilen sonuç oluşuyorsa, artık bu durum-davranış çifti öğrenilmiş olunur ve farklı zamanlarda da sergilenmeye devam edilir. Aksi bir durumda ise beklenen ödül işareti gelmemiş olacağından dolayı, artık o anki durum için farklı bir davranış sergilenecek ve oluşacak durumlara göre de benzer değerlendirmeler yapılacaktır.

Ele aldığımız “TD” yöntemini en iyi şekilde açıklayan bir blok diyagram şekil 2.3’ te yer almaktadır.

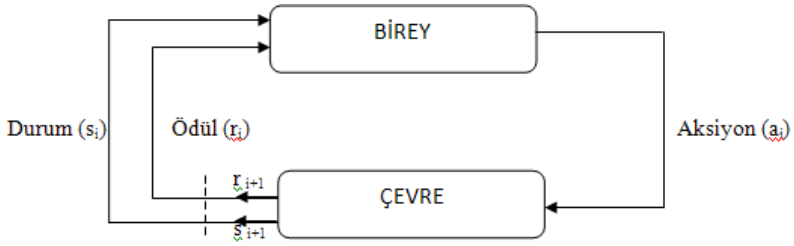


Şekil 2.3 TD algoritması [10]. Dopamin nöronlarının aktivasyonundan faydalanarak elde edilen karar verme mekanizmasının mühendislik sistemler için modellenmiş şeklidir. Her bir durum için oluşturulan değer fonksiyonları, davranış planı içerisinden bir davranışın (u(t)) seçilmesini sağlar ve bunun sonucunda meydana gelen ödül (r(t)) ve durum (x(t)), bu değer fonksiyonlarının güncellenmesini ve yeni bir davranış tahmininde bulunulmasını sağlar. Bu şekilde hata değeri en aza indirilmeye çalışılır.

3. PEKİŞTİRMELİ ÖĞRENME

Pekiştirmeli öğrenme en basit haliyle bir birey (agent) ve çevre (environment) ilişkisidir aslında. Birey ve çevre arasındaki ilişkiden yola çıkarak belli bir amaca ulaşılmaya çalışılır. Bireyin seçtiği davranışa (action) çevreden bir cevap gelir yani çevre bireye yeni durumlar sunar ve bunun sonucunda da birey yeni bir davranış belirler. Bu davranışının belirlenmesinde en etkili faktör ise yaptığı işin ne kadar doğru ve olumlu olduğunu gösteren ödüllerdir (reward). Bu ödül işareti, geliştirilen modellerde skaler bir büyüklük olarak alınır ve yine çevre tarafından bireye gönderilir. Bireyin bu sistemde amacı da zaten, aldığı bu ödüllerin sayısını ya da bir başka deyişle toplamını belli bir zaman aralığı sonunda en fazla yapmaktır. Bu değerler ve dolayısıyla pekiştirmeli öğrenme yapısı hiçbir zaman bireye yapmak istediğini nasıl yapması gerektiğini söylemez. Sadece davranışı sonucunda ne durumda olduğunu ve bunun kendisi için iyi olup olmadığını söyler [10]. O yüzden de pekiştirmeli öğrenme bir eğiticili öğrenme yapısı değildir ve kesin doğrular yoktur. Tıpkı gerçek hayatta hiçbir durumda neyin bizim için gerçekten doğru ve faydalı olacağının belli olmadığı ve bilinemediği gibi.

Öğrenmek, ödüllerin tahmin edilemezlik derecesine büyük ölçüde bağlıdır [9]. Bu yüzden içinde bulunulan koşullar altında başarıya ulaşmak için ne yapmak gerektiğini tahmin edebildiğimiz sürece öğrenme tamamlanmış demektir. Uzayan süreçler göz önüne alındığında ise, öğrenme de aynı şekilde zorlaşacağı aşikârdır. Çünkü karmaşıklığın fazla olduğu bir ortam ya da sistemle yüz yüze olmak, aynı şekilde uzun bir süreçten meydana gelen bir problemin içerisinde yer almak, çok fazla durum ve durum değişkenini hesaba katmayı gerektireceğinden, öğrenme içerisinde var olan bilinmeyen sayısı artacak ve öğrenme zorlaşmış olacaktır.



Şekil 3.1. Birey - Çevre ilişkisi [11]. En genel şekilde pekiştirmeli öğrenme yapısının ifadesidir. a_i davranışı birey tarafından seçildiğinde bu davranış bireyin etkide bulunduğu ortamda bir s_i durumuna karşılık düşer ve r_i ödül işareti ortam tarafından üretilir.

Şekil 3.1 ile ifade edilen yapıda, bireyin gelecek durumlara göre belirleyeceği yeni davranışların olasılıklarını içeren bir davranış planı (policy), π_t, mevcuttur. Bu davranış planı içerisinde, her bir duruma karşılık yeni bir davranış seçimi dolayısıyla her bir "t" anı için bir durum-davranış çifti mevcuttur (π_t (s,a) | s_t = s => a_t = a). Bu nokta pekiştirmeli öğrenme algoritması ise, bireyin geçmiş tecrübelerine göre davranış planını nasıl değiştireceğini açıklar.

Bu kavramların daha iyi anlaşılması açısından bioreaktörler ile kimyasal üretme örneğini verecek olursak [11]; burada, faydalı kimyasalların üretildiği bir sistemin ya da mekanizmanın sıcaklık ayarının yapılması istendiğini düşünelim. Bu sistemde, davranış olarak gösterebileceğimiz etkenler içerisinin sıcaklığını ayarlamada kullanılacak olan ısıtıcı ya da motorların etkin ya da devre dışı edilmesi gerektiğini söyleyecek komutlar ya da işaretlerdir. Durumlar da, sistem içerisinde yer alan sıcaklık ölçerlerin, alıngaçlar ya da benzeri aletlerin gösterdiği değerler olacaktır. Bunların sonucunda çevre tarafında oluşturulacak ödül değeri ise, an ve an ölçümü yapılan faydalı kimyasal miktarı ya da oranının ölçülmesi olacaktır. Bu miktar istenilen şekilde ya da aksi durumda devam ediyorsa ona göre bir davranış üretilecek ve bunun sonucunda da farklı bir duruma geçilecektir.

Bir birey ve çevre arasındaki ilişkinin etkili olduğu faktörlerin bir sonraki durumlar ve politikanın değiştirilmesi olduğunu gördük. Bu demek oluyor ki, bireyin karar mekanizması aslında birçok karmaşık yapıyı içerisinde barındırmaktadır. İşte bu yüzden, bu yapının belli bir karara varıp davranışta bulunabilmesi için yeni bir kavram, "değer fonksiyonu (value function)" devreye girer. Değer fonksiyonu,

herhangi bir durumda iken seçilecek herhangi bir davranışının neden olacağı sonucun bizim için ne kadar faydalı olacağına göre her bir duruma ya da durum-davranış çiftine atanan bir değerdir. Bu fonksiyon; eğer ki bunu durum için yapıyorsak $V^{\pi}(s)$, ”durum değer fonksiyonu”, durum-davranış çifti için yapıyorsak $Q^{\pi}(s,a)$, “davranış değer fonksiyonu” olarak gösterilir.

Amaca ulaşmak için elde edilen ödüllerin miktarı en fazla olmalıdır. Bu maksimum değer bir beklenen değerse;

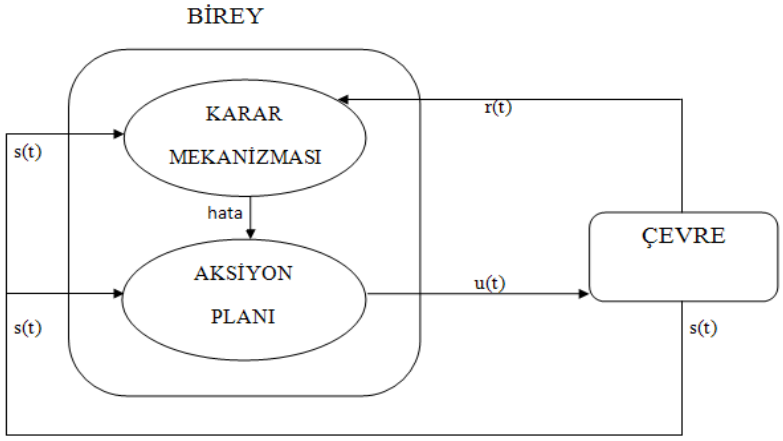
$$R^t = r_{t+1} + \gamma\, r_{t+2} + \gamma^2 r_{t+3} + \ldots\ldots = \sum \gamma^k\, r_{t+k+1} \qquad (3.1)$$

belirlediğimiz davranış planı doğrultusunda içinde bulunduğumuz durumun ya da durum-davranış çiftlerinin değer fonksiyonları şu şekilde ifade edilir;

$$V^{\pi}(s) = E_{\pi}\{ R_t \mid s_t = s \} \qquad (3.2)$$

$$Q^{\pi}(s,a) = E_{\pi}\{ R_t \mid s_t = s, a_t = a \} \qquad (3.3)$$

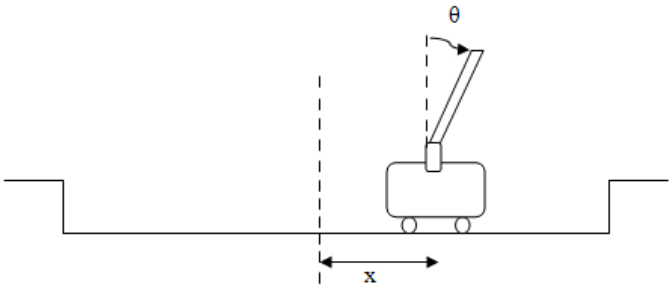
Bunlarla birlikte pekiştirmeli öğrenme yapısında bireyi temsil eden yapı içerisinde, aslında bir karar mekanizması ve bunun sonucunda da değişebilen bir davranış planı olduğu görülür. Tüm bunların ilişkisi de şekil 3.2’ de yer almaktadır.



Şekil 3.2. Pekiştirmeli öğrenmenin genel şeması. Çevreden gelen r, ödül işareti ve s, durum bilgisine göre karar mekanizması tarafından bir hata analizi yapılır ve bunun sonucunda seçilecek davranışlarda ve dolayısıyla da davranış planında bir değişiklik yapılır. Sonunda bir u, davranışı seçilir ve çevreye uygulanır.

3.1 Pekiştirmeli Öğrenme ile Araba-çubuk problemi

Pekiştirmeli öğrenme kuralının kullanıldığı zor ve klasik problemlerden biri araba çubuk problemidir [12]. Burada bir boyutlu bir platform üzerinde bir araba ve ona menteşelemiş dik duran bir çubuk ve arabaya etkiyen bir F kuvveti bulunmaktadır. Bu kuvvetin değeri +10 N ya da -10 N olmak üzere iki değer almaktadır. İstenen durum ise, arabanın x ekseni boyunca referans noktasına uzaklığının, platformun merkezinden 2,4 m ile -2,4 m aralığında, üzerindeki çubuğunda düşey eksenle açısının ise -12° ile +12° arasında kalmasıdır [5].



Şekil 3.3. Araba - Çubuk Problemi [5]

Şekil 3.3' te gösterilen problemde, her bir t anında ağa ait olan karar mekanizması sistemden bir ödül işareti almaktadır. Eğer arabanın konumu x, ya da çubuğun düşey eksenle olan açısı θ, 'dan bir tanesi, daha önce belirttiğimiz şekilde istenilen sınırlar içinde hareket etmez ise bu ödül işaretinin değeri başarısızlık ifade edecek şekilde negatif olmaktadır. Ağdan beklenen ise bu ödül işaretini mümkün olduğunca negatif olamayacak şekilde tutmak yani sistemi başarısızlığa uğratmamaktır. Burada yer alan dinamik sistemin de davranışında tek etkili faktörün uygulanacak kuvvet olduğu düşünülürse, tasarlanacak yapay sinir ağının önemi de bu kuvvetin hangi yönde olacağının karar verilemesinde ortaya çıkmaktadır. Dolayısıyla ortam ya da çevre ifademiz araba ve çubuk sisteminin içinde yer aldığı dinamik sistem ve davranış olarak söz edeceğimiz etken ise uygulayacağımız kuvvetin yönü olacaktır.

Bu ağ yapısı bir pekiştirmeli öğrenme yapısı olduğundan her hangi bir hafıza özelliği yoktur. Ağın özelliği gelen şu anki cevaba göre yeni bir şeye karar vermektir.

Yapı içerisinde iki tür yapay sinir ağı yapısı bileşeni bulunmaktadır. Bunlar; sistemin girişine gelen değişkenlere ve sistemin ürettiği çıkışa yani ödül işaretine bağlı olarak, sisteme dair yeni bir tahmini ödül işareti oluşturan "Uyarlanabilir Eleştiri Bileşeni (ACE)" ve buradan aldığı işarete göre doğru yönde bir kuvvet uygulayarak sistemi süren "Uyarlanabilir Karar Bileşeni (ASE)" 'dir [5]. ASE, tasarlanan ağ içerisinde kontrol davranışlarına karar veren bir eleman görevindedir bunu yapmasını sağlayan işareti ACE' den alır. Yani direkt olarak sistemden gelen hata işaretine maruz kalmaz. Bunun sonucunda da uzun süreli başarıyı sağlayacak olan davranış dizisini elde etmeye çalışacaktır. ACE ise sitemden gelen hata işaretini güncelleme ve bunu ASE' ye iletme görevindedir.

Bu problem içerisinde yer alan dinamik sisteme ilişkin durum değişkenleri ve durum-uzayı denklemleri ise şu şekildedir;

Durum değişkenleri;

x : arabanın platform üzerindeki konumu, ± 2.4 m

θ : araba üzerindeki çubuğun açısı, $\pm 12°$

$\dot{x}$: arabanın hızı, $\pm\infty$ m/s

$\dot{\theta}$: araba üzerindeki çubuğun açısal hızı, $\pm\infty°/s$

Bu problem için uygulanan pekiştirmeli öğrenme yapısı içerisinde uygulanabilecek iki tür yöntemden söz edebiliriz. İlki kutu (box) yöntemidir [5].

İkinci bir yöntem olarak bu bitirme çalışmasında uygulayacağımız yöntemde ise sistemin çıkışları direkt olarak ağın girişine verilir. Bunu yapmaktaki amacımız; dört değişken üzerinden ağı tasarlayıp, sadece, bizim için önemli olan konum ve açı bilgisine göre bir ödül işareti oluşturmak ve buna göre uygun kuvvet yönünü belirlemektir. Onun haricinde diğer iki değişken için istenilen sınırlar sonsuz aralığında olduğundan değerlendirmeye katılmasına ihtiyaç yoktur. Ayrıca, kutu yönteminde olduğu gibi ağın girişlerine ait ağırlıklardan söz etmek mümkündür. Burada belirlenecek olan "ödül (r(t))" işareti de, bu iki değişken istenilen sınırlar içerisinde ise "1", istenilen sınırlar içerisinde değilse de "-1" değerini alır.

$$r(t) = \begin{cases} 1, & -2{,}4\text{m} < x < 2{,}4m \;\; ve \;\; -12° < \theta < 12° \\ -1, & \text{diğer durumlarda} \end{cases} \quad (3.4)$$

Bu hata işaretinin etkisi kutu yöntemindeki gibi sadece belli durumlar üzerinde değil, ağın tamamı üzerindedir.

Dinamik bir sistem olan araba-çubuk sistemine ilişkin durum-uzayı denklemeleri ise şu şekildedir;

$$\ddot{\theta}t = \frac{g \sin \theta t + \cos \theta t \left[\frac{-F - m\, l\, \dot{\theta}t^2 \sin \theta t + \mu c\, sgn(\dot{x}t)}{mc + m}\right] - \frac{\mu p\, \dot{\theta}t}{m\, l}}{l\left[\frac{4}{3} - \frac{m \cos^2 \theta t}{mc + m}\right]} \quad (3.5)$$

$$\ddot{x}t = \frac{F + m\, l\left[\dot{\theta}t + \sin \theta t - \ddot{\theta}t \cos \theta t\right] - \mu c\, sgn(\dot{x}t)}{mc + m} \quad (3.6)$$

$$m1 = mc + m \quad (3.7)$$

$$f1 = m\, l\, \dot{\theta}t^2 \sin \theta t - \mu c\, sgn\,(\dot{x}t) \quad (3.8)$$

$$f2 = g \sin \theta t + \cos \theta t \left[\frac{-F - f1}{m1}\right] - \frac{\mu p\, \dot{\theta}t}{m\, l} \quad (3.9)$$

olmak üzere;

$$\ddot{\theta}t = \frac{f2}{l\left[\frac{4}{3} - \frac{m \cos^2 \theta t}{m1}\right]} \quad (3.10)$$

$$\ddot{x}t = \frac{F + (f1 - m\, l\, f2 \cos \theta t)}{m1} \quad (3.11)$$

$\theta t = \theta 1$, $\dot{\theta}t = \theta 2$ ve $xt = x1$, $\dot{x}t = x2$ olmak üzere (10) ve (11)' teki diferansiyel denklemlerin ve diğer durum değişkenlerine ait denklemlerin açık Euler kuralı ile yazılmasının ardından aşağıdaki sonuçlar elde edilir;

$$\theta 1(k+1) = \theta 1(k) + \beta\,[\theta 2(k)] \quad (3.12)$$

$$\theta 2(k+1) = \theta 2(k) + \beta \left[\frac{f2(k)}{l\left[\frac{4}{3} - \frac{m \cos^2 \theta 1(k)}{m1}\right]}\right] \quad (3.13)$$

$$x1(k+1) = x1(k) + \beta\,[x2(k)] \tag{3.14}$$

$$x2(k+1) = x2(k) + \beta\left[\frac{F(k) + (f1(k) - m\,l\,f2(k)\cos\theta1(k))}{m1}\right] \tag{3.15}$$

Sisteme ilişkin parametreler:

g = - 9.8 m/s^2 , yerçekimi ivmesi

m_c = 1 kg , arabanın kütlesi

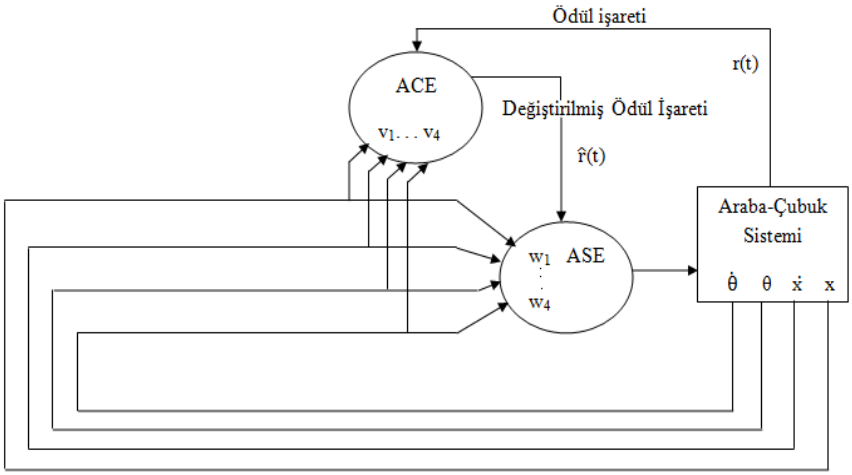
m = 0.1 kg , çubuğun kütlesi

l = 0.5 m , çubuğun uzunluğu

μ_c = 0.0005 , araba ve yol arasındaki sürtünme katsayısı

μ_p = 0.000002 , araba ve çubuk arasındaki sürtünme katsayısı

F(t) = ± 10 N , arabanın merkezinden sisteme uygulanan kuvvet



Şekil 3.4. Araba – Çubuk probleminin çözümünde kullanılan ağ yapısı [5]

Bu probleminin çözümünde kullandığımız şekil 3.4' teki ağ yapısında, ACE ve ASE bileşenlerinin karar mekanizmasında gerçekleştirdiği işlemler aynı olmasına karşın, [5]' teki kutu sistemindeki işleyişten farklı olarak burada ağın

girişinde bir dekoder yoktur. Çünkü her bir değişkenin belli aralıklarda kuantalanarak ağın girişinden uygulanmasının yerine direkt olarak bu değişkenlerin ağa uygulanması ön görülmüştür.

3.2 Pekiştirmeli Öğrenmenin Gerçeklenmesi

Daha önce belirtiliği gibi bu problemde amaçlanan araba ve çubuğun istenilen sınırlar içinde kalmasını sağlamak yani sisteme doğru yönde kuvvet uygulamaktır. Bunu sağlamak amacıyla ilk olarak ACE, sitemin çıkışından gelip ağırlıklarla çarpılan değerleri alıp, her bir davranış sonucunda sistemden gelen hata sinyaline göre sonraki adımda oluşacak hataya ilişkin bir tahminde bulunur. Daha sonra bu tahmini ASE' ye iletir. Bu tahmin, ASE' de davranış belirleme konusunda var olan belirsizliği az da olsa azaltmaya yarayacaktır. Bu *p(t)* tahmin işareti sistemin durumunu belirleyen bilginin fonksiyonudur ve şu şekilde ifade edilir;

$$p(t) = \sum_{i=1}^{n} vi(t)\, xi(t) \qquad \text{ya da} \tag{3.16}$$

$$p(t) = [\, v1 \quad v2 \cdots \cdots \cdots vn \,] \begin{bmatrix} x1 \\ . \\ . \\ xn \end{bmatrix} \tag{3.17}$$

Bu bitirme ödevinde bu problemin çözümünde kullanılan ağ yapısında bu tahmin işareti bir sigmoid fonksiyondan geçirilerek belli bir adım sonra sonsuz değerlere ulaşması engellenmiştir. Onun haricinde burada, değişken sayısı 4 olduğu için; n=4' tür. Bu tahmin işareti ve sistemden gelen r(t) hata işaretine göre ACE tarafından bir $\hat{r}(t)$ güncellenmiş hata işareti oluşturulur ve ASE' ye gönderilir. Yani bu değer Adaptif Kritik Elemanın çıkışıdır ve 3.18 denklemi ile ifade edilir;

$$\hat{r}(t) = r(t) + \gamma\, p(t) - p(t\text{-}1) \tag{3.18}$$

Burada γ , $0 \leq \gamma < 1$ olmak üzere, Adaptif Kritik Elemanın yapmış olduğu hata işaretine dair tahminin daha önceki tahminleriyle benzerliğini belirlemeye yarayan bir sabittir ve bu problem için 0.95 olarak belirlenmiştir. [γ p(t) – p(t-1)] terimi minimum yapılarak, güncellenmiş hata işaretinin sabit kalması ve bunun da sistemden gelen ödül değerine eşit olması istenir. Yapılan bu tahminlerin doğruluğunu arttırmak için ise Adaptif Kritik Eleman' a ait ağırlılar güncellenir. Bu

güncelleme, hata işaretine dair yapılan tahminin yakınsamasını da sağlayacaktır. Güncelleme işlemi de 3.19' de verilen denklemle ifade edilir;

$$v(t+1) = v(t) + \beta\,[\, r(t) + \gamma\, p(t) - p(t\text{-}1)\,]\,\bar{x}(t) \qquad (3.19)$$

β, v' nin değişme oranını belirleyen bir sabit olmak üzere, değeri 0 - 1 arasında değişmektedir. Bu problem için değeri 0.5 alınmıştır. Bu denklemde yer alan $\bar{x}(t)$ terimi; ACE' nin ürettiği işaretten bağımsız olarak giriş işaretinin seçilebilirliğini gösterir**.** Bu terimin güncellenmesi ise 3.20' de verilen denklem ile yapılmaktadır.

$$\bar{x}(t+1) = \lambda\,\bar{x}(t) + (1 - \lambda)\,x(t) \qquad (3.20)$$

Burada λ, bozulma oranını ifade etmektedir ve değeri, $0 \leq \lambda < 1$ olmaktadır. Bu problem için değeri ise, 0.8 alınmıştır.

ASE içerisinde gerçekleştirilen işlemlere bakacak olursak, ilk olarak ağa ait olan çıkış değeri yani sisteme uygulanacak kuvvet belirlenir.

$$y = f\,[w^{T}(t) + x(t) + n(t)\,]\ , \qquad n(t) : \text{gürültü} \qquad (3.21)$$

$$y = f(x) = \begin{cases} 1\,, & x < 0 \\ -1\,, & x \geq 0 \end{cases} \qquad (3.22)$$

ASE için de ağırlıkların güncellenmesi en önemli işlemlerden biridir. Eğer çıkışlar istenildiği şekilde elde ediliyorsa o anki her bir giriş değişkeni için ağırlıklarda artma meydana gelir. Bunun amacı, sistemin bu koşulları sağlayan durumda kalmasını sağlamaktır. Aksi koşullarda ise, şu an ki durumdan farklı bir duruma gidilmesi gerektiği anlamı çıkar ve buna neden girişler için var olan ağırlıklar azaltılmalı demektir. Bu nedenle de ağırlıklarda bir azaltma meydana getirilir. ASE için ağırlıkların güncellenmesi işlemi de 3.23 denklemi ile yapılmaktadır.

$$w(t+1) = w(t) + \alpha\,\hat{r}(t)\,e(t) \qquad (3.23)$$

Burada α, ağırlıkların değişme oranının göstermektedir. Denklemde yer *e(t)* ise, sistem için uygunluk derecesini göstermektedir. Bu terim, seçilen davranışların oluşturduğu ve buna bağlı olarak sistemden gelecek hata işareti arasında bağlantı kurmaya yarar. Böylece, hangi davranışlarla en uzun süre başarıya

ulaşılacağı hakkında bir fikir ve bunu ifade eden bir katsayı elde ediliyor. Kısacası gelen sonucu seçilen davranışla ilişkilendirme fonksiyonudur. Bu fonksiyonun ifadesi ise 3.24' teki denklemde yer almaktadır.

$$e(t+1) = \delta\, e(t) + (1-\delta)\, y(t)\, x(t) \tag{3.24}$$

Burada δ, bozulma sabiti olarak ifade edilebilir ve $0 \leq \delta < 1$ olmaktadır.

Bu güncelleme ve hesapların ardından ASE sisteme etki edecek kuvvetin yönüne karar verir ve bu da ASE' nin çıkış değeridir. ASE için çıkış ifadesi de 3.25 denklemi ile belirlenir.

$$F = \begin{cases} F\ ,\ y >= 0\ , aynı\ yönde\ kuvvet\ uygulanmaya\ devam\ edilir \\ -F\ ,\ y < 0\ , bir\ önceki\ duruma\ göre\ ters\ yönde\ kuvvet\ uygulanır \end{cases} \tag{3.25}$$

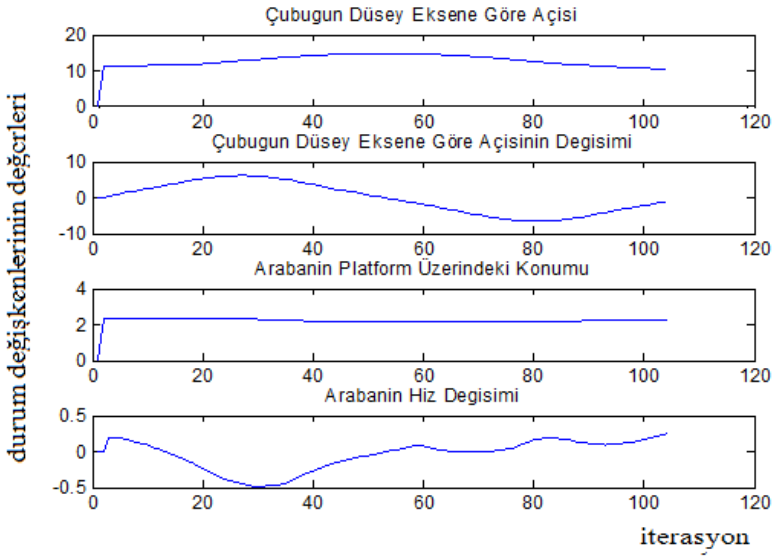
3.3 Benzetim Sonuçları ve Değerlendirmeler

Başlangıç koşulları olarak, konum ve açının istenilen sınırlara yakın noktalar seçilmesiyle araba – çubuk problemi için tasarlanan pekiştirmeli öğrenme ağ yapısının benzetimi gerçekleştirilmeye başlanmıştır. Bunun için, kendim en baştan oluşturduğum MATLAB kodu ile ilgili M-file dosyaları da bitirme çalışmamın Ek'ler bölümünde yer almaktadır. Burada beklenilen sonuç, eğer sisteme uygun şekilde kuvvet uygulanıyorsa, karar mekanizmasında oluşturulan değiştirilmiş ödül işaretinin sistem tarafında gönderilen ödül işareti ile aynı olmasıdır. Yani (3.17) denklemi göz önüne alınırsa yapay sinir ağı bileşeninin her bir adımda yaptığı tahmin işaretinin bir önceki tahmin işaretiyle tutarlı olması ve böylece sıfıra giden ikinci tarafın ardından $(\gamma\, p(t) - p(t-1))$, $\hat{r}(t) = r(t)$ eşitliğini sağlanmasıdır.

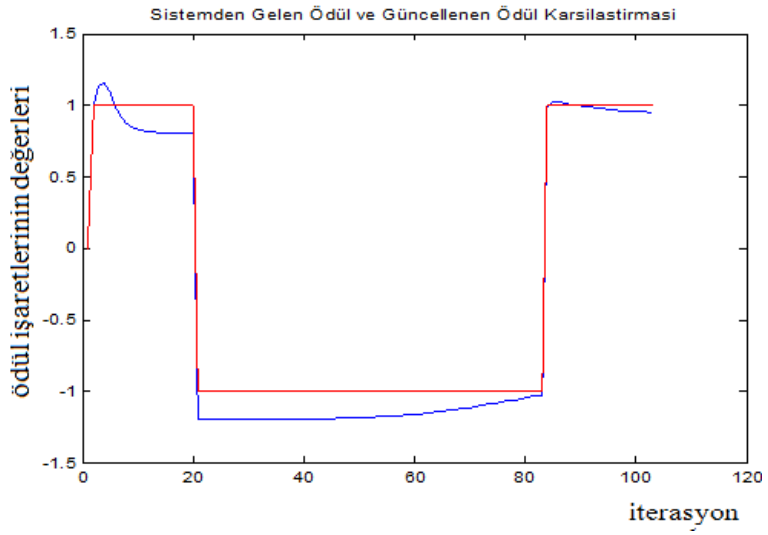
Sınır değerler, ilk koşullar olarak seçildiğinde;

```
Q1(2)= 11;          'Çubuğun Düşey Eksene Göre Açısı'
Q2(2)= 0;           'Çubuğun Açısının Değişim Hızı'
x1(2)= 2.3;         'Arabanın Referans Noktasına Göre Konumu'
x2(2)= 0;           'Arabanın Konumunun Değişim Hızı'
```

arabanın platform içerisindeki konumu, ivmelenmesi, açının konumu ve açının değişim grafikleri sırasıyla şekil 3.5' te olduğu gibi elde edilmiştir.



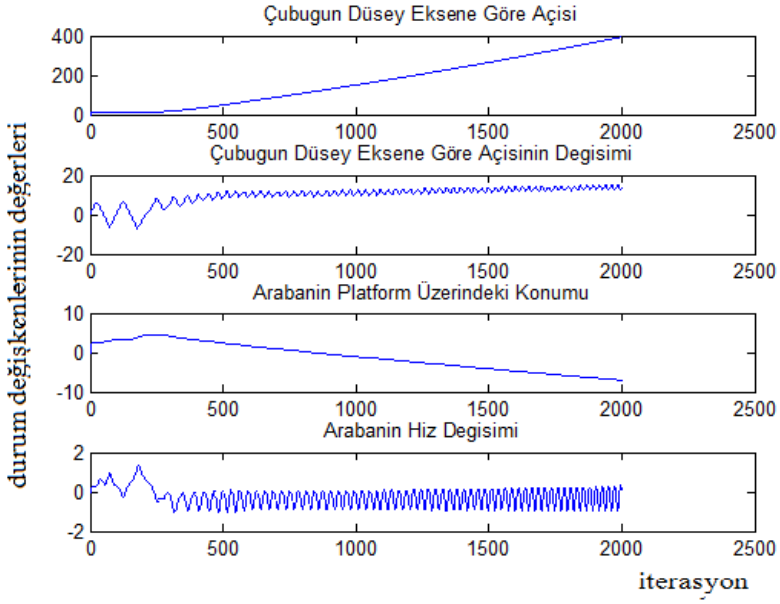
Şekil 3.5. Araba-çubuk problemi için, ilk koşullar istenilen sınırlara yakın değerlerde iken yapılan denemelerde başarılı olunan durumlar için elde edilen sonuçlar. Tasarlanan ağ yapısı çoğunlukla dinamik sistemi bu sınırlar içerisinde tutmayı başarmıştır.



Şekil 3.6. Sınır koşullara yakın başlangıç değerleri için başarıya ulaşıldığında sistem tarafından gönderilen ödül işareti (kırmızı) ile güncellenen yani yapay sinir ağı elemanı tarafından tahmin edilen ödül işaretinin (mavi) değişimi.

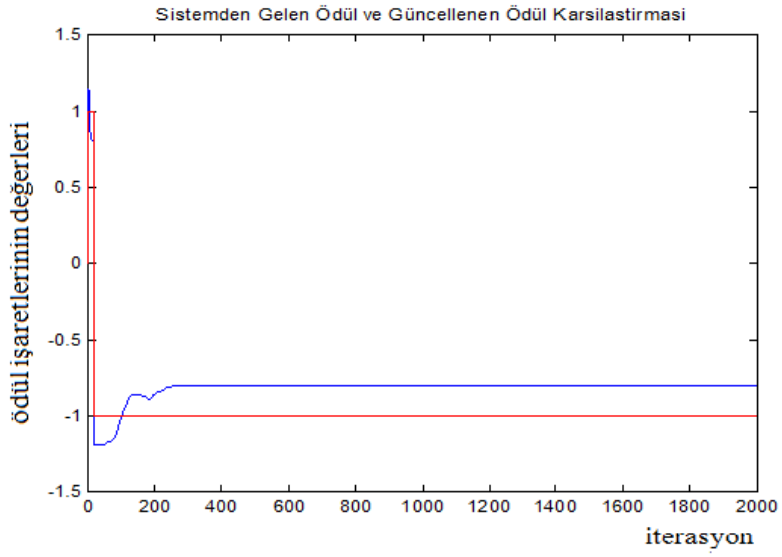
Şekil 3.6' da görüldüğü gibi, güncellenen işaret belli bir süre sonra gelen ödül işareti ile yakın hatta eşit değerler almış bu sayede sistemin başarılı olması sağlanmıştır.

Bu şekilde seçilen başlangıç koşulları için yaptığımız 20 deneme boyunca ortalama 12 – 13 kez, istenildiği şekilde araba ve çubuğun istenildiği şekilde belirlenen sınırlar için kalması sağlanmıştır (Şekil 3.5 ve Şekil 3.6). Bunu yaparken belirlediğimiz durdurma kriteri ise, tasarladığımız ağ yapısı sistemden 40 defa pozitif bir ödül işareti alıyor olmasıdır. Bunun gerçekleştirildiği durumlarda ağ yapımız sistemi kontrol edebilir hale gelmiştir diyebiliriz. Onun dışındaki durumlarda ise araba ve çubuk istenilen sınırlar dışına çıkmıştır yani ağ yapımız başarısız olmuştur. Bu durumu ifade eden bir sonuç için grafikler şu şekildedir;



Şekil 3.7. Araba-çubuk problemi için, ilk koşullar istenilen sınırlara yakın değerlerde iken yapılan denemelerde başarısız olunan durumlar için sonuçlar.

Şekil 3.7' de görüldüğü gibi, tasarlanan ağ yapısı böyle durumlarda sistemi başarıya ulaştırmakta yetersiz kalmış, istenilen sınırlar dışına çıkılmıştır.



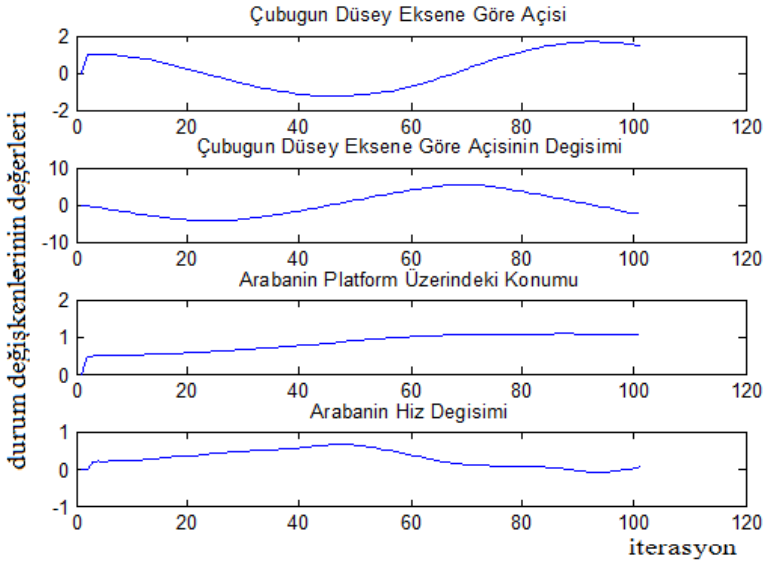
Şekil 3.8. Sınır koşullara yakın başlangıç değerleri için başarısız olunan durumlarda sistem tarafından gönderilen ödül işareti (kırmızı) ile güncellenen ödül işaretinin (mavi) değişimi.

Sistem istenildiği sınırlar dışarısına çıktığında ödül işareti negatif değer almış ve tekrar bu sınırlar içerisine girilemediği için de bu değerinde sabit kalmıştır (şekil 3.8). Bu nedenle de güncellenen işaret, sistemden gelen ödül işaretine yakınsayamamıştır.

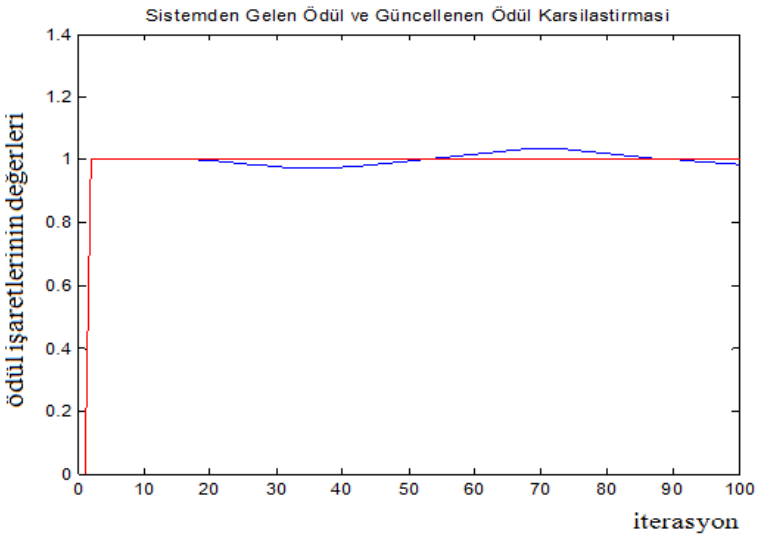
Daha sonra sistemin başlangıç koşullarını referans noktasına yakın noktalarda seçerek denemeler yaptığımızda,

```
Q1(2)= 1;
Q2(2)= 0;
x1(2)= 0.5;
x2(2)= 0;
```

Şekil 3.9' da görüldüğü gibi, 20 deneme sonunda, dinamik sistem tüm denemelerde istenilen sınırlar içerisinde kalmıştır. Yani pekiştirmeli öğrenme ağ yapımız bu başlangıç koşulları dâhilinde sistemi kontrol etmekte başarılı olmuştur. Burada durdurma kriteri olarak sistemden 100 defa pozitif değerde ödül işareti alınması istenmiştir. Bunun nedeni de, eğer arabanın konumu ya da çubuğun açısı sınır değerlere yaklaşacak olursa, sonraki durumlarının ne olacağını görebilmektir.



Şekil 3.9. Araba-çubuk problemi için, ilk koşullar referans noktasına yakın değerlerde iken yapılan denemelerde elde edilen sonuçlar. Tasarlanan ağ yapısı böyle durumlarda sistemi başarıya ulaştırmış her denemede sistem istenilen sınırlar içinde kalmıştır.



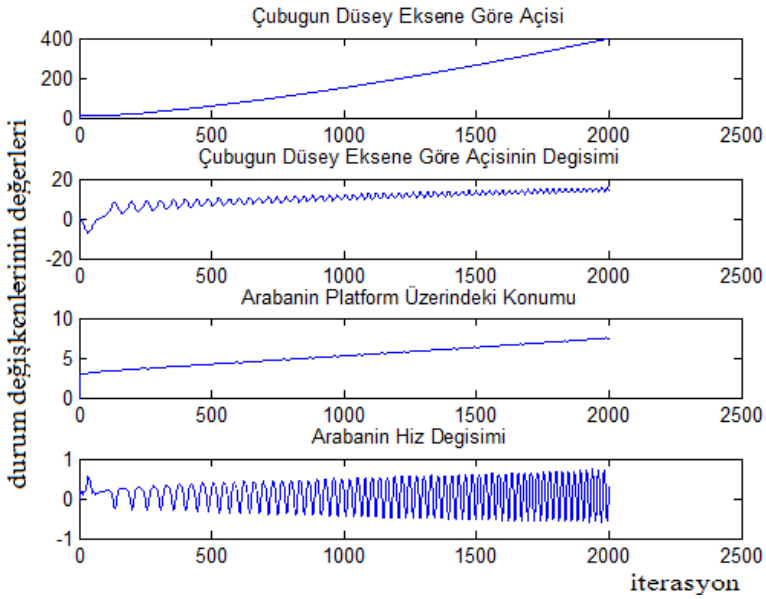
Şekil 3.10. Referans noktasına yakın başlangıç değerleri için yapılan denemelerde, sistem tarafından gönderilen ödül işareti (kırmızı) ile güncellenen ödül işaretinin (mavi) değişimi.

Şekil 3.10' a baktığımızda ise kolaylıkla görülüyor ki, güncellenen işaret sürekli olarak gelen ödül işareti ile yakın hatta eşit değerler almış bu sayede sistemin başarılı olması sağlanmıştır.

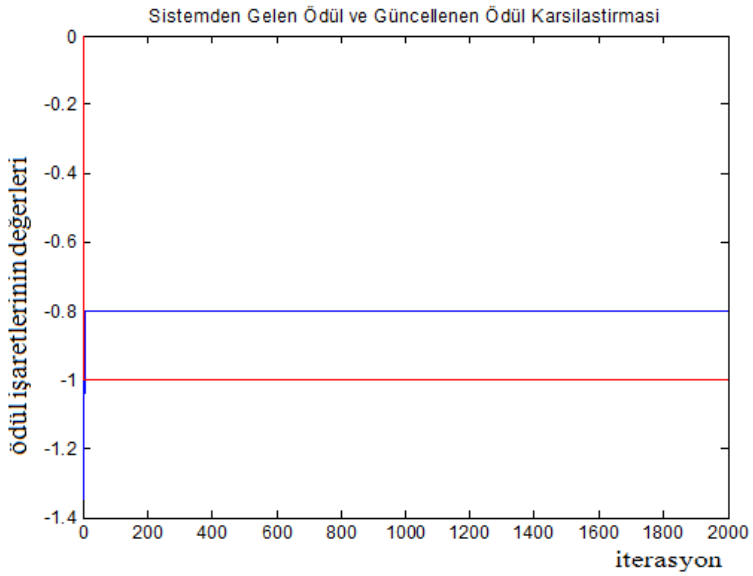
Sistemin başlangıç koşulları istenilen sınırlar dışında seçilerek;

```
Q1(2)= 15;
Q2(2)= 0;
x1(2)= 2.9;
x2(2)= 0;
```

tasarlanan ağ yapısı sisteme uygulandığında elde edilen sonuçlarda, ilk baştan itibaren oluşan negatif ödül işaretinin benzetim boyunca aynı kaldığı ve dolayısıyla da testin başarısız olduğu görülmektedir**.** Sonuçlar, şekil 3.11 ve şekil 3.12' de görülmektedir.



Şekil 3.11. Araba-çubuk problemi için, ilk koşullar sınır değerlerin dışarısında iken yapılan denemelerde elde edilen sonuçlar. Tasarlanan ağ yapısı böyle durumlarda sistemi başarıya ulaştırmakta yetersiz kalmış, sınırlar dışarısında başlayan sistem salınımını arttırarak sonsuz değerlere doğru gitmektedir.



Şekil 3.12. Sınır koşullar dışında seçilen başlangıç değerleri için başarısız olunan durumlarda sistem tarafından gönderilen ödül işareti (kırmızı) ile güncellenen ödül işaretinin (mavi) değişimi.

Şekil 3.11 ve şekil 3.12' de görülen son yaptığımız denemelerde, araba çubuk düzeneği belirlenmiş sınırlar dışındayken ödül işareti negatif değerler almış fakat bunun sonucunda uygulanan kuvvet ile istenen sınırlar içerisine giremeyerek kararsız bir şekilde sınırlardan tamamen uzaklaşmıştır.

Elde edilen bu sonuçları genel olarak değerlendirecek olursak, çoğu durumda istenildiği şekilde aracın konumu ve çubuğun açısı belirli sınırlar içerisinde kalmıştır. Ancak dinamik sistemin belli bir anda sınırlar dışına çıkmasıyla kontrol edilebilirlik ortadan kaybolmakta ve ağ yapımız başarısız olmaktadır. Ya da bir başka şekilde sistem direk olarak sınırlar dışarısında başlatmak da ağın kontrol işini başaramamasına ve sistemi tekrar sınırlar içine alamamasına neden olmaktadır. Yine de sınırlar içerisinde başlayan çoğu durumda ve referans noktasına yakın değerlerden başlayan tüm denemelerde araba ve çubuk beklenildiği ve sistemden istenildiği şekilde hareket etmiştir. Böylece kullandığımız yapay sinir ağı bileşenleri ve onların oluşturdukları karar mekanizması ile birlikte araba – çubuk sistemi kontrol edilebilir hale gelmiştir denilebilir.

4. ARDIŞIL ÖĞRENME

Biz farkında olsak da olmasak da öğrenme diye tabir ettiğimiz birçok davranış belli bir sürecin ardından meydana gelmektedir. Sergilenen her bir davranış veya gerçekleşen bir takım olayların ardından gösterdiğimiz tepki ve aldığımız kararlar aslında bu sürecin bir ürünüdür. Her bir adımda ortaya çıkan davranışlar bu süre boyunca bir davranış dizisi oluşturur ve birbirini takip eder. Bunun sonucunda da öğrenme gerçekleşmiş demektir. Bizim şu an için yaptıklarımız ve aldığımız kararlar, bir sonraki adımda karşılaşacağımız durumlara ve dolayısıyla da daha sonra yapacaklarımıza doğrudan etki etmektedir. Biz de bu şekilde edindiğimiz tecrübelerin bir sonucu olarak, herhangi bir aşamaya gelmek için, belli aralıklarla ve sırasıyla neler yapmamız gerektiğini öğrenmiş oluruz.

Bir mühendislik probleminin çözümünde ya da bir dinamik sistemin kontrolünde de elde edilmek istenen, herhangi bir durum karşısında ya da bir dinamik sistem için her hangi bir değişkende değişme meydana geldiğinde olası davranışa ya da müdahaleye karar verebilmek değildir. Asıl önemli olan nokta, bir dizi davranışı ve ya değişen durumlara göre istenen davranışları sergileyerek başarıya ulaşmak yani bu çalışma boyunca söylenildiği şekilde ödülü elde etmektir. Planlama gibi süreçlerin temelinde de bu ardışıl davranışları oluşturmak vardır.

4.1. Pekiştirmeli Öğrenmenin Ardışıl Öğrenme Testi İçerisinde Uygulanması

Daha önceden değindiğimiz, dopamin maddesinin öğrenme üzerindeki etkisi bu kez ardışıl öğrenme için modellenecektir. Bunun sonucunda her bir adımda farklı koşullar sağlanacak, farklı davranışlar beklenecek ve belli bir ardışıl süreç tamamlandıktan sonra ödüle varılacaktır [9]. Bu şekilde elde edilen öğrenme şekli, uzun süreli ve birden fazla adımlı öğrenme problemlerinde de önemlidir. Bu amaç doğrultusunda, ele alınan bir problemi şimdiye dek yapılan modellemeler ve tasarlanan ağ yapısı ile çözümlemek istersek şu şekilde bir uygulama yapmak yerinde olacaktır.

Bir ağa farklı durumlar sunularak elde edilen yanıtlar değerlendirilecek ve bir dizi başarılı davranış sonucunda ödüle ulaşılacaktır. Burada ele aldığımız problem ve çözümde kullanacağımız model ile üçüncü bölümdeki araba çubuk problemi için kullandığımız model ve karar mekanizmaları aynı olmasına rağmen hata işaretinin kullanılmasında farklılık vardır. Araba çubuk probleminde, sistemden gelen ödül işaretine göre doğrudan bir kuvvet düzeneğe etki ederken, burada ise bir takım harfler ağa bir sıra halinde sunulacak ve bu harflere karşı bir takım rakamlar atanarak doğru biçimde ve ardışıl olarak eşleştirilmesi sağlanacaktır. Yapılan iş yine üçüncü bölümde olduğu gibi bir karar mekanizmasının işleyişi ve dolayısıyla da bir davranış seçme işlemidir. Buradaki farklılık, seçme işlemi sonrasında meydana gelen hata işareti, ikinci bölümde sıkça değindiğimiz dopamin aktivasyonunun bir modeli olarak yer almaktadır ve dinamik yapı içerisindeki parametreleri doğrudan etkileyerek harf seçimini değiştirebilmektedir. Bu problem için oluşturulan ve karar verme işlemini gerçekleştiren dinamik yapı ise, beyindeki davranış seçme işleminde etkin olan Basal Ganglia, Talamus ve Korteks döngüsünden esinlenerek önerilen bir yapıdır [6]. Beyinde dopamin maddesinin etkisiyle, bu yapıların değişen fonksiyonları ve davranışları, bizim modelimizde ise, hata terimiyle güncellenen parametreler olarak yer almaktadır. Burada ki hata teriminin güncellenmesi ise, araba-çubuk probleminde olduğu gibi üçüncü bölümde yer alan 3.18 denklemi ile yapılmaktadır.

Ardışıl öğrenme olayı gerçekleştiğinde ağ, belli durumları belirli davranışlarla eşleştirebiliyor olacaktır. Konuyla ilgili olarak yapılan çalışmada, ağın A, B, C durumlarına karşılık 1, 2, 3 davranışlarını sergilemesi beklenmiş ve bunun sağlanmasının ardında da ödül işareti pozitif olmuştur [6]. Bizim burada yapacağımız ise, aynı çalışamadan faydalanarak ancak boyutları yani ardışıl öğrenmedeki aşama sayısını dörde çıkarak, benzer şekilde bir ardışıl öğrenme testi gerçekleştirmektir.

“A” > “1” => “B” > “2” => “C” > “3” => “D” > “4” => ödül

Ardışıl öğrenmede izlenen yöntem, öğrenmeye son aşamadan başlanarak geriye doğru gidilmesidir [7]. Yani bir ağa ilk olarak diziyi oluşturan son durum olan “D” sunulur ve bununla “4” davranışının eşleştirilmesi beklenir. Burada yapılan, bir davranışın seçilerek buna bir değer atılmasıdır. İlk olarak gelen “D” harfi bir durumu temsil ediyorken, bunun karşılığında seçilen davranışa da “4” değeri atanmaktadır.

“4” değeri bizim için bir beklenen değerdir ve ağ bunu seçtiği sürece başarılı olarak ödül işaretini alacaktır. Bu da yapılan seçimin sonraki zamanlarda da aynı durum için tekrarlanmasını sağlayacaktır. Tıpkı araba-çubuk probleminde gelen ödül işareti doğrultusunda aynı yönde bir kuvvet uygulanmasına devam edilmesi gibi. Bu da tamamen beyindeki dopamin etkisini bize göstermektedir. Olumlu sonuçlanan davranışlar benzer durumlarda tekrarlanır ve belli bir sürecin ardından da artık öğrenilmiş olur [8]. Bu problemde de eşleştirmenin sağlandığı anda ağa bir ödül işareti verilir ve tekrar başa dönülür. Öğrenmenin gerçekleştiğini varsaymak için ise, bu eşleştirmenin istenen sayı kadar doğru yapılması gerekmektedir. Bu koşul sağlandığında artık öğrenme sonuçlanmış demektir.

“D” > “4” => ödül

Bu aşama için öğrenme tamamlandığında bir sonraki adıma geçilir ya da aslında bir adım geri gidilerek, bir başka durumun karşılığı olacak şekilde “C” harfi ağa sunulur. Ağ bir önceki adımda “4” davranışını seçmeyi öğrendiği ve parametreleri, ki bunlar ağ içerisinde yer alan ağırlıklardır, bu amaç doğrulusunda güncelleştirdiği için, bu yeni durum karşısında ilk olarak yanlış bir davranış sergileyerek, “4” ile eşleştirme yapacak ve bunun ardında da ödül bekleyecektir. Ancak bu istenen durum değildir. Yine aynı araba-çubuk probleminde olduğu gibi bir durum söz konusudur. Orada, bir yönde uygulanan kuvvet, ta ki araba ya da çubuktan biri istenen sınırlar dışına çıkıp da negatif bir ödül işareti gelene kadar devam ediyordu. Negatif ödül işareti ile de kuvvet, bu kez ters yönde düzeneğe uygulanmıştı. Burada da genel olarak benzer bir durum söz konusudur. Gelen “C” harfine karşılık, ağın seçtiği davranışı “3” ile eşleştirmesini bekleriz ancak bu olmaz. Beklenen “3” rakamı ile seçilen “4” rakamı arasındaki bu fark ödül getirmediği gibi bir hata terimi olarak da deltayı (δ) doğurur. Bu hata işareti ilk anda çok büyük değerdedir çünkü seçilen değer ile beklenti arasındaki farklılık çok fazladır. Bu hata işaretine göre dinamik yapı içerisindeki parametreler güncellenerek aktörün başka bir seçimde bulunması sağlanır. Ne zaman ki seçilen davranış, “3” numarası ile eşleştirilir, o zaman ağ ödül işaretini alır. Bu doğru seçim tekrarlandıkça da artık hata işareti de azalmaya başlar ve bu aşama için de öğrenme tamamlanmış olur. Artık ağ “C” ile “3” ‘ ü de eşleştirmiş olacaktır. Bu öğrenmenin gerçekleşmesinin ardından ağa ödül yerine bu sefer “D” verilir ve zaten daha önceden öğrenilmiş olan bu eşleşme sonucunda ödüle ulaşılmış olunur.

“D” > “4” => “C” > “3” => ödül

Görüldüğü gibi burada yer alan delta terimi ile ödül işareti arasında açık bir ilişki söz konusudur. Sergilenen davranış karşılığında ödül işareti gelmezse, delta hata terimi büyük değerlerde olmakta ve bunun sonucunda da ağ ağırlıklar güncellenerek farklı davranışlar sergilenmekte yani harf-sayı eşleştirmeleri yapılmaktadır. Doğru eşleşme sonucunda da gelen ödül işareti delta teriminin sıfıra doğru yakınmasına neden olur ve bu da ağırlıkları ayını değerinde tutarak bu eşleşmenin öğrenilmesini sağlar. Böylece delta terimi davranışların belirlenmesinde kullanılarak nihayetinde ödül işaretinin elde edilmesinde çok önemli bir rol oynamaktadır.

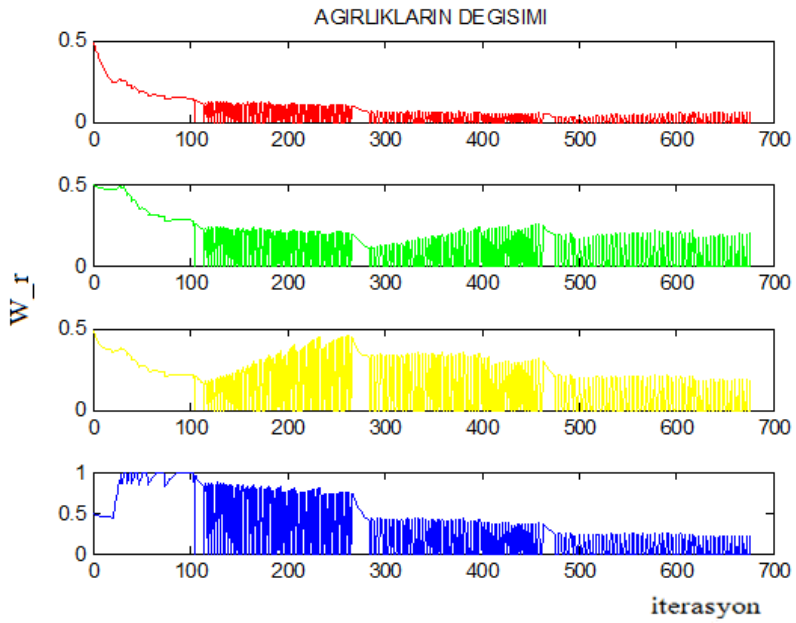
Bu şekilde öğrenme son iki aşamada da tamamlandıktan sonra yine aynı şekilde “B” ve “A” durumu içinde aynı şeyler yapıldıktan sonra öğrenme tamamlanmış olur.

“D” > “4” => “C” > “3” => “B” > “2” => “A” > “1” => ödül

4.2. Ardışıl Öğrenme Testinin Sonuçları ve Değerlendirmeler

Ardışıl öğrenme için, genel yapı itibariyle [6] çalışmasındaki MATLAB kodu kullanılmıştır. Ancak orada yapılan çalışmada üç boyutlu veriler kullanılmış ve üç aşamalı öğrenme yapılmıştır. Bu çalışmada ise, aynı kod üzerinde yapılan değişiklik ve düzenlemelerle dört aşamalı öğrenme sağlanmış ve verilerin de boyutları dörde çıkarılmıştır. Kodun bu düzenlenmiş hali tezin sonundaki Ek’ ler bölümünde yer almaktadır.

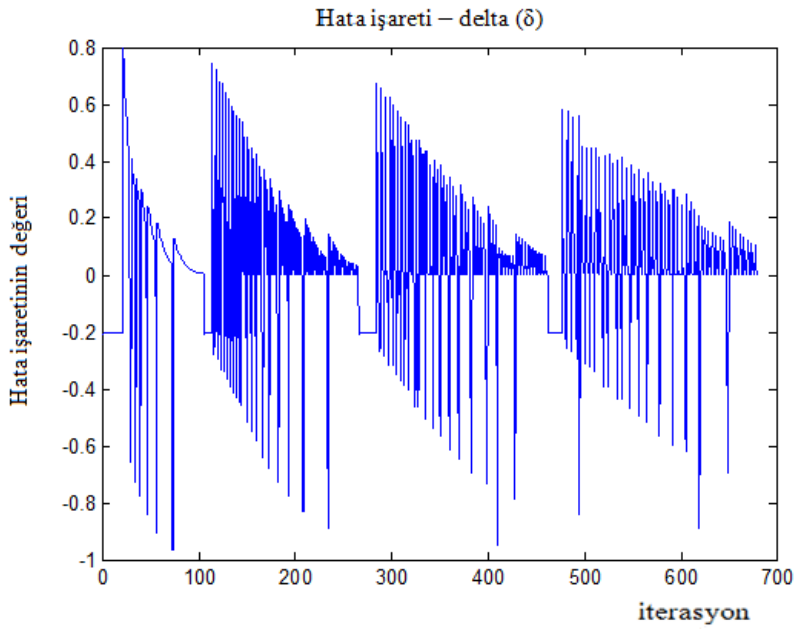
Bu bitirme çalışmasında gerçekleştirdiğimiz testlerde elde ettiğimiz sonuçlar şekil 4.1 ve şekil 4.2’ de yer almaktadır.



Şekil 4.1. Ardışıl öğrenme testi sırasında ağırlıklarda meydana gelen değişmeler.

Davranış seçimi için [6]'da önerilen korteks- bazal çekirdekler ve talamus arasındaki ilişkileri modelleyen dinamik sistemin davranışı bu sistemdeki w_r parametresine bağlı olarak değişmektedir [6]. Bu bitirme çalışmasında ki amaçda bu parametredeki değişimin boyut büyüdüğünde de gözlemlemektir ve Şekil 4.1'de dört boyutlu olan bu parametrenin her bir bileşenindeki değişim verilmiştir.

Her bir aşama da öğrenme gerçekleştikçe ağırlıklar salınım yapmaktan yani rastgele değerler almaktan uzaklaşarak birbirine yakın değerleri takip etmektedir. Yeni bir durumla karşılaşıldığında ise ani ve büyük değişimler meydana gelmektedir. Böylelikle farklı durumlarla karşılaşıldığında istenen davranışı sergilemek mümkün hale gelmektedir. Hatanın büyük olması hızlı değişimlere neden olurken, hata azaldıkça değişimlerde meydana gelen büyüklük de aynı şekilde azalmaktadır.



Şekil 4.2. Ardışıl öğrenme testi sırasında hata (δ) işaretinde meydana gelen değişmeler.

Her bir adımda, öğrenme gerçekleşmeden önce büyük değerlikli hatalar, zamanla sıfıra yakınsar. Daha önce de bahsettiğimiz gibi bu hata yani delta işareti, seçilen davranışın eşleştirilmesinin beklenildiği rakam ile eşleştirildiği rakam arasındaki farkı göstermektedir. Bir durumla ilk kez karşılaşıldığında, seçilen davranış beklenenden oldukça farklı olduğu için bu hata değeri büyüktür. Sonrasında bu hata değerinden faydalanarak dinamik yapı içindeki parametreler güncellendikçe, hata değerinde de azalma meydana gelir. Bunun nedeni sistemin sergilediği yeni davranış seçimleri sonucunda ödül işareti almasıdır. Ödül işareti doğru bir davranış seçildiğini gösterir ve bu yüzden hata işareti - delta zamanla azalma eğilimindedir. Böylece ağırlıkların bu değerde sabit kalması sağlanarak o anki durum davranış eşleştirmenin öğrenilmesi sağlanır.

Bir sonraki adıma geçildiğinde yani farklı bir ortam koşulunda ise hata işareti tekrar yükselme gösterir. Çünkü bu aşamada da istenen davranış sergilenmediği için ödül işareti gelmemiştir ve bu nedenle hata terimi büyüktür. Mesela yine daha önce değindiğimiz gibi, “D” durumuna karşılık seçilen davranış “4” rakamı ile eşleştirilmişken, bir sonraki aşamada yani “C” durumuna karşılık yine

“4” rakamına karşı davranışın seçilmesi büyük bir hatanın oluşmasına neden olur. Çünkü beklenen ve seçilen davranışlar tamamen farklıdır. İşte bu hata teriminden faydalanarak, dinamik sisteme ait ağırlıkların üçüncü bölümde anlatıldığı şekilde güncellemesi sonucunda farklı bir davranışın seçilmesi ve dolayısıyla da farklı bir rakamla eşleştirmenin yapılması sağlanmış olur. Ne zaman ödül doğru bir davranış ile bir sonraki aşamaya ve dolayısıyla da en sonunda ödüle ulaşılır, o zaman artık bu yeni durum için de öğrenme gerçekleşmiş demektir. Bu andan itibaren de hata işareti delta, azalarak sıfıra yakınsar.

Bizim ele aldığımız problem burada 4 aşamadan meydana geldiği için, 4 farklı bölgede deltanın sıfıra doğru yakınsadığı görülmektedir. Sonuç olarak, elde ettiğimiz sonuçlar beklentilerimiz doğrultusunda çıkmış ve delta işareti ile ödül işareti arasındaki açık bir şekilde görülebilmiştir.

5. ART YAPISININ ARDIŞIL ÖĞRENME YAPISI İÇERİSİNDE KULLANILMASI

Ardışıl öğrenmenin gerçekleştirilmesi farklı şekillerde ve farklı problemlerde de ele alınabilmektedir. Bu açıdan ART yapısının da içinde yer aldığı bir başka problemden söz edilebilir. Bu da dağınık şekilde var olan harflerden anlamlı bir kelime oluşturma problemidir.

ART yapısı, kendi içerisinde var olan kısa ve uzun süreli bellekler sayesinde örüntüleri ve kendisine gelen her türlü veriyi saklama özelliğine sahiptir [7]. Ancak bunları belli bir sıraya göre sıralamaya veya seçmeye yönelik bir işleve sahip değildir. Örneğin ele alınacak problem açısından baktığımızda, herhangi bir kelimenin oluşturulması ya da belli verilerin bir araya getirilerek sunulması ART içerisinde birbirinden bağımsız olarak yapılmaktadır. Yani bir kelime ancak, doğru harfler doğru sırayla çağırıldığında oluşturulabilmektedir. İşte bu noktada, ART yapısının yanında pekiştirmeli öğrenmeden de faydalanmak, ardışıl bir öğrenmenin gerçeklenmesinde ön plana çıkmaktadır.

Bu çalışmada da yapacağımız şey, ART tarafından seçilen harflerle anlamlı bir kelime oluşturabilmektir. Çünkü baktığımızda bir kelimenin oluşturulması belli durumların, ki burada durumlar harfler ile temsil edilmektedirler, belli bir sıra ile birbirini takip ediyor olmasına dayanmaktadır. Bu da daha önce gerçekleştirdiğimiz ardışıl öğrenmeye benzemektedir. Bu bölümde ele alacağımız problemde artık, dördüncü bölümde yaptığımız gibi durumları ve rakamları kodun kendi içerisinde belirtmiyoruz. Bunun için dışarıdan herhangi bir kullanıcı ile ilk olarak oluşturulacak kelimenin harfleri alfabedeki sıralarıyla ART tarafından bellekten çağırılıyor. Sonra da, dağınık halde gelen bu harflerin kelime içerisindeki doğru sıraları girilerek de ardışıl öğrenme kodunda doğru kelimenin oluşturulması sağlanıyor. Böylece ardışıl öğrenme farklı bir şekilde ve daha anlaşılır bir problemin çözümünde kullanılıyor olacaktır.

5.1. ART ve Pekiştirmeli Öğrenmenin Ardışıl Öğrenme Testi İçerisinde Uygulanması

ART ve pekiştirmeli öğrenmenin bir arada çalışmasının düşünüldüğü durum için ele alınan bir problemde anlamlı bir kelime oluşturulması söz konusudur. Bu aşamada, dışarıdan gelen kullanıcı, ilk olarak ART için yazılan kod sonucu oluşturulan ve bellekte saklanan harflerden kendi oluşturacağı kelimenin harflerini rastgele bir sırayla çağırır. Bunun nasıl yapıldığı ve çağırma işleminin sırasında da pekiştirmeli öğrenmenin kullanılış biçimi [7] çalışmasında yer almaktadır. Bu şekilde bellekten çağırılan harfler daha önce belirttiğimiz gibi dağınık bir biçimde bu çalışmada kullanılan MATLAB koduna iletilir.

Ardışıl öğrenme için kullanılan koda gelen bu örüntüler, anlamlı bir kelimenin oluşmasını sağlayacak olan harflerdir ancak doğru sıra ile gönderilmemiştir. Bunun haricinde de bu harfler, bir önceki bölümde yer alan “A”, “B”, “C” ve “D” durumlarına karşılık düşmektedir. Bundan sonra yapılacak iş, bu durumlara karşı olarak seçilecek davranışlara bir rakam eşleştirmesinde bulunmaktır. Yani yine önceki bölümde yer alan “1”, ”2”, ”3” ve ”4” rakamlarına karşılık burada ne gelecek? Bunun için de yine kullanıcıdan, seçtiği dört harfin (çünkü yazılan kod dört aşamalı öğrenme yapmakta yani dört boyutta çalışmaktadır) kelime içerisindeki doğru sıraları alınır. Böylece eşleştirilecek rakamlar da elde edilmiş olunur ve öğrenme işlemine geçilir. Öğrenme için yapılacak olan ise hata işaretinden faydalanarak doğru kelimeyi oluşturabilmektir Yani her bir adımda aktör, ilk sırada olması gereken harfi doğru sırası ile eşleştirdikten sonra, gelen ödüle göre sonraki aşamalarda da seçimlerini gerçekleştirecektir. Bu şekilde harfler gerçek yerleriyle eşleştirilir ve anlamlı bir kelime oluşturulur. Aksi durumda eğer ki hala istenen kelime oluşmamışsa, ardışıl öğrenme başarıyla gerçekleştirilememiş demektir.

5.2. Test Sonuçları ve Değerlendirmeler

ART ve ardışıl öğrenmeyi sağlayacak yapıları uygun şekilde birleştirdiğimizde ilk olarak oluşturulacak kelimenin harfleri ART kodu ile bellekten çağırılır [7]. Bunun sonucunda gelen veriler şekil 5.1’ de yer almaktadır.

Şekil 5.1. ART kodu ile dağınık biçimde oluşturulan harfler

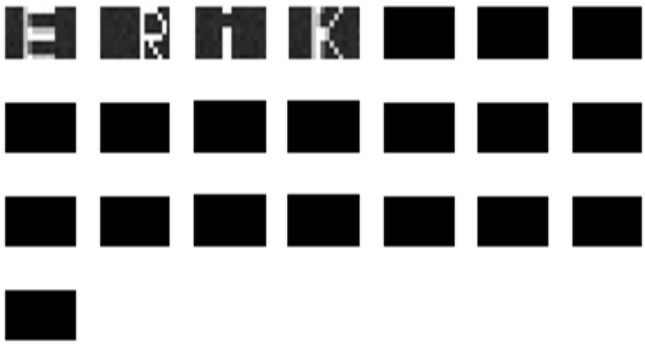
Burada oluşturmak istediğimiz "ERİK" kelimesinin harfleri istenenden farklı olarak, "RKİE" şeklinde elde edilmiştir. Bu nedenle yapılması gereken gelen harfleri kelime içerisindeki doğru sıraları ile eşleştirerek ardışıl öğrenmeyi gerçekleştirmektir. Bu amaçlar şu şekilde bir eşleştirme yapılacaktır;

R = > 2
K = > 4
İ = > 3
E = > 1

Bunun sonucunda ise kod içerisinde gerçekleştirilecek olan öğrenme yöntemi;

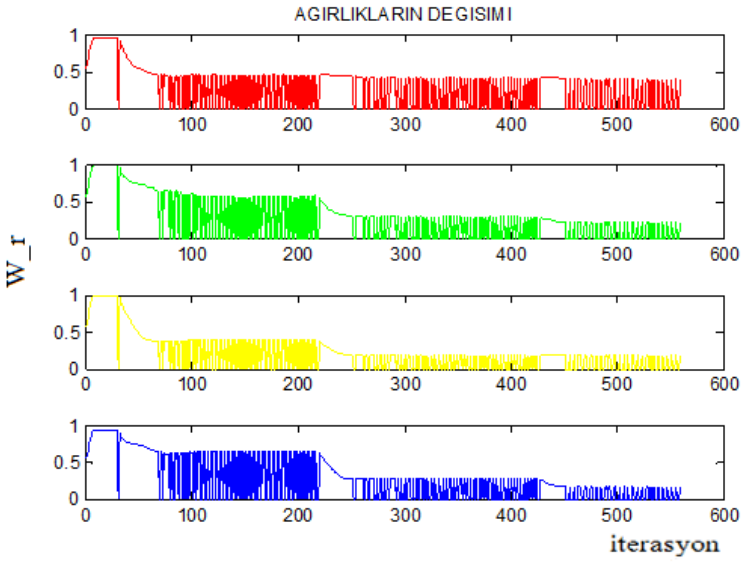
"R" > "2" => "K" > "4" => "İ" > "3" => "E" > "1" => ödül

Her bir aşamada aktör tarafından yapılan eşleştirmeler doğru sonuçlanırsa ödül elde edilmektedir ve bu eşleştirme 30 defa doğru olarak yapıldığında öğrenme gerçekleşmiş sayılmaktadır. Bunu şekilde kod çalıştırıldığında elde edilen sonuçlar şu şekildedir;

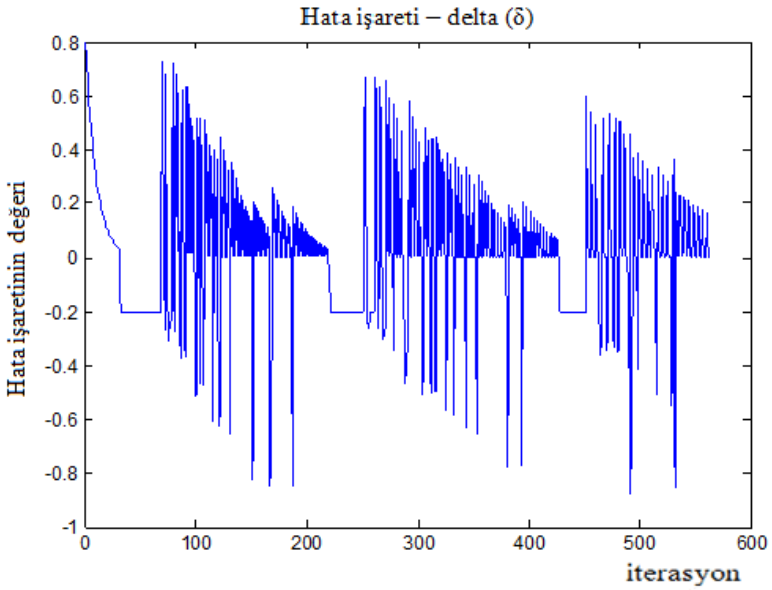


Şekil 5.2. ART kodu ile dağınık biçimde oluşturulan harflerin ardışıl öğrenme ile doğru biçimde sıralanması.

Şekil 5.2' de görüldüğü gibi ardışıl öğrenme sonucunda sıralama işlemi başarıyla gerçekleştirilip harflerin doğru sıralanması sağlanmıştır. Bu öğrenme süreci boyunca ağırlıklarda ve hata işaretinde meydana gelen değişmeler ise şekil 5.3 ve şekil 5.4' te yer almaktadır.



Şekil 5.3. ART ve pekiştirmeli öğrenmenin bir arada gerçekleştirildiği ardışıl öğrenme testi sırasında ağırlıklarda meydana gelen değişmeler.



Şekil 5.4. ART ve pekiştirmeli öğrenmenin bir arada gerçekleştirildiği ardışıl öğrenme testi sırasında hata işaretinde (δ) meydana gelen değişmeler.

Daha önce de belirttiğimiz gibi hata işareti (δ) öğrenme gerçekleştiği sürece sıfıra doğru yakınsamaktadır. Ardışıl öğrenme sırasında yeni bir durumla karşılaşıldığında ise yani bir sonraki harfe geçildiğinde hata tekrar maksimum değerine çıkmak da doğru seçimler yapılıp ödül elde edildikçe yine azalma göstermektedir.

6. SONUÇ VE ÖNERİLER

Bu çalışmada bizim yapmaya ve açıklamaya çalıştığımız, insanlarda öğrenmenin nasıl gerçekleştiğini ve neye dayanarak ya da neleri düşünerek bir davranışta bulunduklarıdır. Bunu yaparken özellikle dopamine maddesinin aktivasyonundan faydalanarak elde edilen modeller ve yöntemleri ele aldık. Bunlardan bizim düşünce sistemimize en yakın olanı TD algoritmasıdır. Bu şekilde bir öğrenme kuralı belirleyip farklı problemlerde uygulamak mümkün olmaktadır. Yaptığımız her davranışın arkasında bir ödül ya da ceza mekanizması mı var yoksa bizim kararlarımız her zaman alışkanlıklardan mı ibaret? İşte tüm bu soruların cevaplarını sunmaya çalışan pekiştirmeli öğrenme sayesinde bir takım problemleri ele alarak çözmeye çalıştık.

İlk aşamada ele aldığımız problem bu konuda sıkça karşılaşılan araba-çubuk problemidir. Bizim yapay sinir ağı ile yapmak istediğimiz ise bu noktada, bizdeki karar verme mekanizmasını modelleyerek dinamik bir sistemin kontrolünün sağlayabilmektir. Bu problemin çözümünde durum değişkenleri (araba ve çubuğun konumları, hızları vs.) bizim ağ yapımız içerisinde direk olarak girişleri meydana getirmektedir. Bu şekilde kod yazılarak sonuçlar elde edilmiştir. Ancak bu problem farklı şekillerde ve farklı değişkenleri giriş olarak alarak da çözülebilmektedir. Bizim uyguladığımız yöntem ile birçok durumda sistem başarıya ulaşarak istenilen sınırlar içerisinde kalmayı başarmıştır. Burada yer alan karar mekanizması tıpkı insanlarda olduğu gibi, oluşan durumu ve çevreden gelen ödül ya da cezaları göz önüne alarak bir davranışta bulunmakta, bu da sisteme herhangi bir yönde kuvvet olarak etki etmektedir.

Dördüncü ve beşinci bölümde ise yapmaya çalıştığımız, TD modelini ele alarak ardışıl olarak tanımlanmış bilişsel süreçlere ilişkin problemleri çözmeye çalıştık. Bunu gerçekleştirmek için ele aldığımız problem ilk olarak bir eşleştirme problemiydi. A, B, C ve D harflerini sırasıyla 1, 2, 3 ve 4 numaralarıyla eşleştirerek ardışıl öğrenmeyi sağladık. Bu seçimler ağ içerisinde aktör tarafından yapılmakta ve doğru seçimler sonunda da ağa ödül verilmektedir. Daha önceden de belirttiğimiz

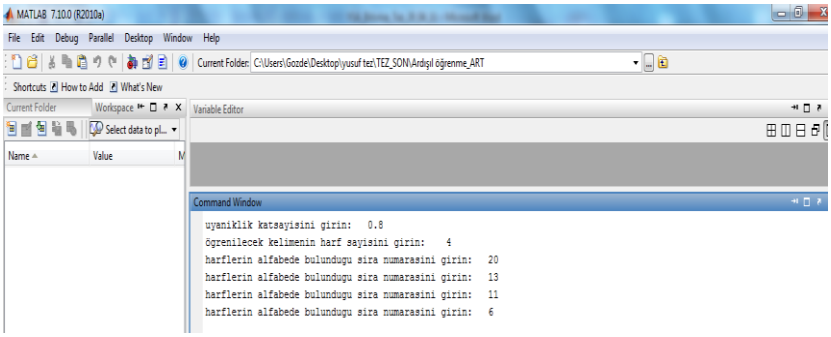
gibi, yapılan yanlış seçimler ise bir hata teriminin oluşmasına neden olmaktadır. Bu hata değerine göre de dinamik yapı içerisindeki parametreler güncellenir ve farklı bir seçimde bulunabilme sağlanmış olur. Ödül işareti geldikçe de bu hata teriminin düştüğü gözlemlenmektedir. Bu aynı ikinci bölümde yer alan dopamin maddesinin aktivasyonuna benzemektedir. Olumlu sonuçlarla karşılaşılan davranışlar dopamin maddesinin etkisi ile bir öğrenme meydana getirir ve bu davranışlar benzer durumlarda da sergilenmeye devam eder. Zaten bizim tasarladığımız yapay sinir ağında da, bu yüzden hata işareti, dopamin maddesinin bir modeli olarak yer almaktadır. Benzer şekilde ele aldığımız bir diğer problemde ise, bu kez ART yapısından da faydalanarak harflerin doğru sıralanıp bir kelime elde edilmesi sağlanmıştır. Burada seçimler yine aktör tarafında yapılarken, kullanılan örüntüler ART sayesinde oluşturulmuştur. Böylece pekiştirmeli öğrenmenin farklı kullanım biçimlerine ve amaçlarına değinilmiştir.

Sonuç olarak, bu bitirme çalışmada yaptığımız, insandaki karar mekanizmasının ve bu mekanizma içinde nöronların işleyişini modellemektir. Bunun sonucunda da ardışıl bir takım işlerin nasıl öğrenildiğini ve gerçekleştirildiğini açıklamaktır. Çünkü hayatımızın büyük çoğunluğunda, her şey başka şeylerin bir devamı ya da tamamlayıcısı olarak vuku bulur ve bu şekilde bir dizi davranış sonucu olaylar sonuçlanır. Belli bir zamandan sonra ise, belki farkında bile olmasak da bazı şeyleri öğrenmiş ve sırasıyla uyguluyor oluruz.

7. EKLER

EK.1

Bitirme çalışmasının beşinci bölümünde, ART ve pekiştirmeli öğrenmeden faydalanarak çözülen ardışıl öğrenme problemi için ilgili kod ("Ardisil_ögrenme_ART > Reinforcer_complete_ART"), tez sonundaki CD'den temin edilerek MATLAB programında koşturulduğunda ilk olarak şekil 6.1' deki ekran "Command Window" penceresinde görülür.



Şekil 6.1. "Reinforcer_complete_ART" kodu koşturulduğunda ilk çıkan ekran

Burada kodun ART kısmına ait olan parametreler ve ele alınacak problem için gerekli bilgiler girilecektir. Örneğin, bizim çalışmada ele aldığımız problem için;

Uyaniklik katsayisini girin: 0.8

Ogrenilecek kelimenin harf sayisini girin: 4

Harflerin alfabede bulundugu sira numarasini girin: 20

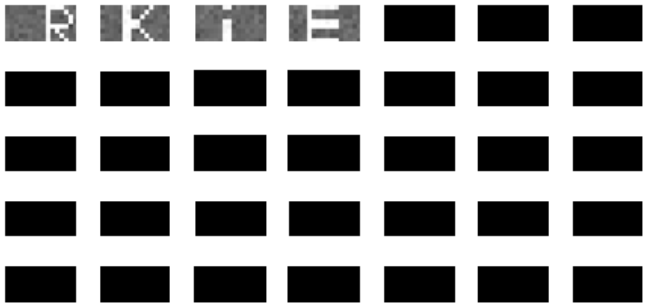
Harflerin alfabede bulundugu sira numarasini girin: 13

Harflerin alfabede bulundugu sira numarasini girin: 11

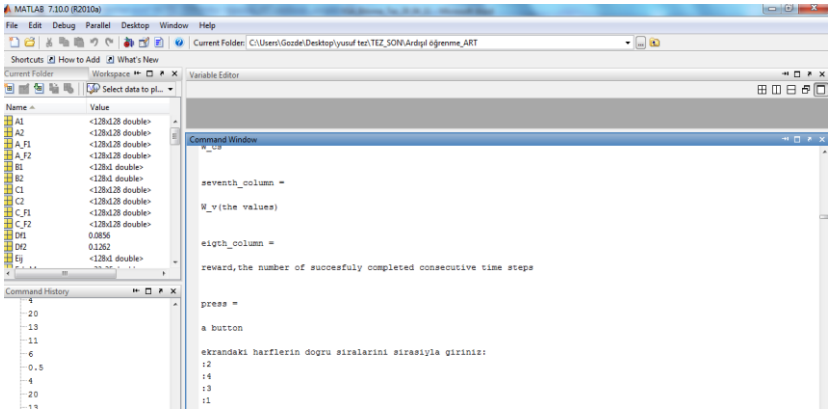
Harflerin alfabede bulundugu sira numarasini girin: 6

Burada harflerin sıraları girilirken “Ğ” olmadığı için 28 harf üzerinden sıranın değerlendirileceğine dikkat edilmelidir.

Sonrasında, şekil 6.2’ de görüldüğü gibi girilen harflerin oluştuğu bir figür penceresi oluşacak ve kodun ardışıl öğrenmeyi sağlayan bölümü kendiliğinden aktif olacaktır ve bu aşamada da şekil 6.3 deki ekran “Command Window” penceresinde belirecektir.



Şekil 6.2. Girilen harflerin, girildiği sıra ile ekranda gösterilmesi



Şekil 6.3. Ardışıl öğrenme kısmının çalışması sonucu gelen ekran. Burada harflerin gerçek sıraları girilecektir.

Şekil 6.2 de çıkan “RKİE” harflerine bakacak olursak; anlamlı bir kelimenin oluşabilmesi için, şekil 6.3’ te oluşan ekranda aşağıdaki sayılar girilecektir;

ekrandaki harflerin dogru siralarini sirasiyla giriniz:
:2
:4
:3
:1

Çünkü oluşturulması istenen “ERİK” kelimesine göre, “R” harfi 2. sırada, “K” harfi 4. sırada, “İ” harfi 3. sırada ve “E” harfi ise 1. Sırada olmalıdır.

Bunun sonucunda da harfler doğru sıralarıyla öğrenilip ekranda yeniden çizdirilecektir.

KAYNAKLAR:

[1] S. Metin, " Doğrusal olmayan sistem yaklaşımı ile duygusal davranışlarda etkin sinir sistemi alt yapılarının incelenmesi " , Doktora Tezi Önerisi, Aralık 2009

[2] W. Schultz, P. Dayan and P.R. Montague "A Neural Substrate of Prediction and Reward", SCIENCE ~~Vol~~Cilt,. 275 14 ~~March~~ Mart 1997

[3] P. Dayan "Reinforcement Learning", C.R. Gallistel (Editör), Steven's Handbook of Experimental Psychology, New York, Wiley, 2001.

[4] R.E. Suri, J. Bargas and M.A. Arbib "Modeling Functions of Striatal Dopamine Modulation in Learning and Planing" Neuroscience ~~Vol~~Cilt. 103, No. 1, sf.~~pp~~. 65 – 85, 2001

[5] A.G. Barto, R.S. Sutton and C.W. Anderson "Neuronlike Adaptive Elements That Can Solve Difficult Learning Control Problems", IEEE Trans. Syst., Man., Cybern., ~~vol~~Cilt. SMC-13, No:5, ~~September/October~~Eylül/Ekim 1983

[6] N. S. Şengör, Ö. Karabacak and U. Steinmetz, " A Computational Model of Cortico-Striato-Thalamic Circuits in Goal-Directed Behaviour", ICANN 2008, LNCS 5164, ~~pp~~sf. 328–337, 2008.

[7] C. Yücelgen, " Uyarlanabilir yankılaşım kuramı ile öznitelik belirleme" , Bitirme Çalışması, İ.T.Ü., Mayıs 2011

[8] M. Domjan "The Principle of Learning and Behavior", ~~Fifth Edition~~5. Basım, Thomson Wadsworth, 2003

[9] R.E. Suri, W. Schultz "Learning of Sequential Movements By Neural Network Model With Dopamine-like Reinforcement Signal", Springer-Verlag, 30 ~~April~~ Nisan 1998

[10] H. Dağhan "Modeling Reinforcement Learning at Basal Ganglia", Bitirme Çalışması, İ.T.Ü., 2004

[11] R.S. Sutton, A.G. Barto, "Reinforcement Learning", (2nd ~~printing~~basım), A Bradford Book, The MIT Press, 1998

[12] M. Coşkun "Uyarlamalı Öğrenme Yöntemi İle Ardışıl Hareket Testinin Modellenmesi", Bitirme Çalışması, İ.T.Ü., 2005

Formatted: Not Superscript/ Subscript

ÖZGEÇMİŞ:

Yusuf KUYUMCU, 1988 İstanbul doğumludur. İlköğrenimini Gaziosmanpaşa ilköğretim okulunda, Orta öğrenimini ise, Pertevniyal Anadolu Lisesinde tamamladı. Lisans eğitimine 2006 yılında İstanbul Teknik Üniversitesi Elektronik Mühendisliğine başlamıştır.