



Kompleks Sistemler ve Veri Bilimi

Complex Systems and Data Science

Uzay ÇETİN^a

Araştırma Makalesi/Research Article

Başvuru/Received: 30.11.2020; Kabul/Accepted: 29.12.2020

ÖZ

Kompleks Sistemler, kendi kendini örgütleyebilen ve değişen şartlara uyum gösterebilen organik sistemlerdir. Kompleks bir sistem, nispeten basit kuralları takip eden çok sayıda etmenin birbiriyle olan yoğun etkileşimi neticesinde ortaya çıkar. Burada asıl olan etmenler değil, onlar arasındaki ilişkililerdir. Kompleks sistem bilimsel bakış açısında yarattığı paradigma değişiminin, çok daha fazlasını derin öğrenme modelleri mühendislik problemlerinin çözümünde yaratmıştır. Katmanlı bir yapıya sahip olan veriye aç bu modellerde, ilk katmanlar en basit öznitelikleri otomatik bir biçimde çıkarırken, sonraki katmanlar hiyerarşik bir biçimde basitten daha karmaşığa doğru üst seviye öznitelikleri çıkarabilme yeteneğine sahiptir. Özel derin öğrenme algoritmaları zamansal ve uzamsal ilişkileri yakalayabilmektedir. Bu çalışmanın amacı veri bilimi, yapay öğrenme ve karmaşık sistemler yaklaşımlarının evreni anlamak için bize birbirini tamamlayan bakış açıları kazandıracağını vurgulamaktır.

Anahtar kelimeler: kompleks sistemler, veri bilimi, derin öğrenme, anomali tespiti

ABSTRACT

Complex Systems are organic systems that can self-organize themselves and adapt to changing conditions. A complex system results from a huge amount of interactions of many agents following simple rules. It is not the agents that are essential, but the relationships between them. Deep learning models created a much greater paradigm change in solving engineering problems than the one created by complex systems in science. Deep learning models are composed of a layered structure. The first layers automatically learn the simplest features, while the next layers have the ability to extract high-level features in a hierarchical manner from simple to more complex. Special deep learning algorithms can even capture temporal and spatial relationships. The purpose of this article is to emphasize the fact that data science, machine learning and complex systems will provide us a complementary perspective to study our universe.

Keywords: complex systems, data science, deep learning, anomaly detection

^a Bilgi Üniversitesi, Bilgisayar Mühendisliği Bölümü, uzay.cetin@bilgi.edu.tr, ORCID: 0000-0002-0784-253X

1. Giriş

Bilgisayar mühendisliğinin temelinde yatan algoritmik yaklaşıma göre, tarif edebildiğimiz her şeyi programlayabiliriz. Peki ya tarif edemediğimiz şeyler? Canlılığı nasıl tanımlarsınız? Bilinç nedir? Kompleks sistem bilim dalının temellerinde yatan ana fikir, “*bütün, parçaların toplamından daha fazladır*” ilkesidir. Bu parçalarda bulunmayan bir özelliğin, bütünde *zuhur* edebileceğini ifade eder. Örneğin tek başına bir karınca zeki değildir, ama bir bütün halinde koloni en kısa yol problemini çözecek zekâ belirtisini gösterir. Yapay öğrenmenin başarısının arkasında ise “*bol miktardaki örneklerden hatayı minimize eden hipotezi bulmak*” yatar. Bu sayede, bir sürecin nasıl oluştuğuna dair elimizde algoritma olmasa bile, elimizdeki örnek çıktı verileri ile tahminimiz arasındaki hatayı minimize eden parametreleri bularak, tarifini bilemediğimiz fakat kolayca yapabildiğimiz yüz tanıma, ses tanıma gibi işleri bilgisayarlara yaptırabiliriz. Bu çalışma, veri bilimi, yapay zekâ ve kompleks sistemler bilim dallarının bir arada ele alınması gerektiğini ve yapay zekaya dair çalışmaların geleceğinde, kompleks sistem yaklaşımının büyük bir rol üstleneceğini iddia eder. Denetimsiz bir yapay sinir ağı çeşidi olan kendi kendine örgütlenebilen haritalar özelinde etmen temelli modellemeye olan benzerliğine vurgu yapılmış ve anomali tespiti gibi çeşitli alanlardaki uygulamaları incelenmiştir. Bu çalışmada kompleks sistemler ve karmaşık sistemler aynı anlama gelecek şekilde kullanılmıştır.

Yapay zekaya dair ilk çalışmalar başladığında, yapay zekâ ile ilgili çalışmaların insan zekasının ne olduğuna dair bizleri aydınlatacağını sanıyorduk. Bugün gelinen noktada, araştırmacılar bu idealden büyük ölçüde vaz geçmiş görünüyorlar. Doğadan esinlenen, insan gibi düşünen ve insan gibi davranan modeller yerine rasyonel düşünen ve rasyonel davranan modellere doğru bir eğilim olduğunu görmekteyiz (Russell & Norvig, 2009). Zor, karmaşık fakat evrenin işleyişi hakkında ipucu veren yaklaşımlar yerine, evrenin işleyişinden bağımsız, basit ama işe yarayan indirgemeci yaklaşımlar ağırlıklı olarak tercih edilmektedir. Araştırmacılar olarak, bir an önce sonuç elde etmeye fazlaca odaklanmanın doğal sonucu olarak bu noktadayız. Ama evreni anlamak istiyorsak, farklı bir bakış açısı geliştirmek zorundayız.

1.1. Evrene Bakış Açısı

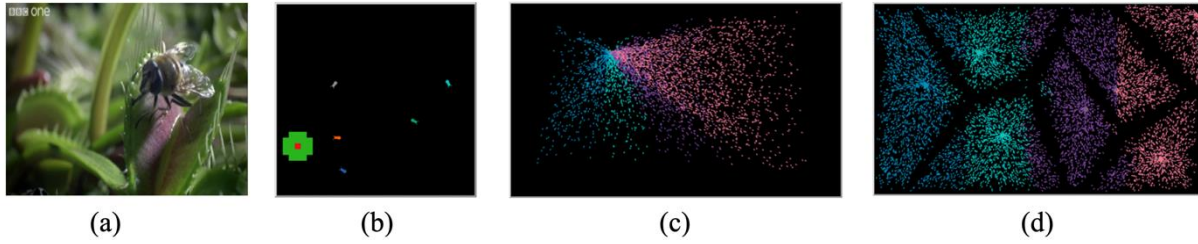
Seth Lloyd, 2006 yılında basılmış *Programming the Universe* adlı kitabında şöyle söylüyor: “Evrendeki her atom, her parçacık bilgi kaydeder. Bu parçacıklar arasındaki her bir çarpışma, ne kadar küçük olursa olsun, meydana gelen her bir etkileşim sistematik bir biçimde o bilginin işlenmesidir.” Buradaki iki kritik nokta, *bilginin kaydı* ve *bilginin işlenmesidir*. Bilindiği üzere bunlar modern programlamanın temel unsurlarıdır. Demek ki, sadece bilgisayarlar bilgi depolayan ve o bilgiyi işleyen aygıtlar değildir. *Evrenin kendisi de baştan sona bilgi işleyen devasa bir sistemdir*. Bu bakış açısı bizi doğal olarak, *düşüncenin* de bir bilgi sayım işlemi olduğu sonucuna götürür. Düşüncenin bir bilgi sayım işlemi olduğuna dair inancımız olmasaydı, yapay zekâ çalışmaları belki de hiç başlamazdı. Daha fazla ilerlemeden, doğadaki bilgi-sayım bir örnek verelim. Şekil 1 (a)’da görülen bitki Venus-flytrap adıyla bilinir (Attenborough, 2009).

Bu bitki, sinekleri cezbeden bir salgı üretir ve içinde bulunan tetikleyici tele bir sinek 20 saniye içinde iki kez dokunduğunda birden kapanır. Diğer bir ifadeyle bu bitki, beyni olmamasına rağmen, bilgi sayım yapar. Şekil 1 (b)’de, bu bitkinin işleyişini taklit eden bir simülasyonu görmekteyiz. Bu simülasyonu NetLogo dili ile yazdık. Sistemin nasıl işlediğini, yani algoritmasını bildiğimiz için, bitki ve sineğin simülasyonunu rahatça yapabiliyoruz, bakınız Şekil 2.

Simülasyonun kuralları,

- Böcekler en yakın cezbedici noktaya gider.
- Eğer böcek bitkinin merkezindeki tele dokunursa,
 - Bitki kapanır ve böceği sindirir!

Şekil 1. (a) Venus-Flytrap Bitkisi (b) Bitkinin ve Birkaç Böceğin Basit Simülasyonu (c) Böcek Sayısı Arttırılarak Yapılan Simülasyon (d) Hem Böcek Hem Bitki Sayısı Arttırılarak Yapılan Simülasyon.



Dikkat ederseniz burada, basitleştirmelere gittik. Modellemenin başarısı, önemli olan noktaları seçmek kadar önemsiz olanları da göz ardı etmeye dayanır. Gerçek o kadar karmaşıktır ki, seçici bir şekilde basitleştirmeye gitmek zorunludur. Bilgisayar temelli simülasyon, bize farklı senaryoları denemek için muhteşem bir laboratuvar imkânı sunar. Böcek sayısını arttırsak ne olur acaba? Şekil 1 (c)'de görüldüğü gibi simülasyonda, bitki karadeliği andırır hale gelir ve bütün böcekleri yutar. Peki hem böcek sayısını arttırır hem de bitki sayısını arttırsak ne olur? Şekil 1 (d)'de görüldüğü gibi, böcekler tıpkı yapay öğrenmede sıkça kullanılan K-Means kümeleme algoritmasını andıracak bir biçimde en yakın bitki merkezlerine doğru yönelir (Jain, 2010). Belirli bir süre için, merkez noktası farklı bitkiler olan kümelenmeler ortaya çıkar. Oldukça basit bir modelin, ufak değişikliklerle bu kadar zengin sonuçlar doğurması şaşırtıcı derecede güzeldir. Yeri gelmişken belirtmekte fayda var, teleskop astronomiyi, mikroskop biyoloji temellerinden değiştirdi. Bilgisayar ve simülasyon ise tüm bilimlere kökünden değiştirmektedir. Alternatif senaryoları deneme gücü, simülasyonun en önemli artlarından biridir.

2. Doğal Zekanın Simülasyonu Olarak Yapay Zekâ

John McCarthy 1956'da Dartmouth Üniversitesinde 10 kişilik bir ekip ile 2 aylık bir çalıştay düzenledi. Bu çalıştayda ilk kez yapay zekâ adı kullanılmış ve yapay zekâ bilimi resmen doğmuştur. Burs başvurusunda, McCarthy yapay zekâyı şöyle tanımlamıştır: “Bu araştırma, zekâ gerektiren öğrenmenin tüm yönleri ile birlikte bir makine tarafından *simüle* edilebileceği varsayımından yola çıkar.”

Daha yapay zekanın, dünyaya ilk adımlarını atışında bile simülasyon kendini hissettirir. Amaç insan zekasının bir benzerini makinelerde ortaya çıkarmaktır. Yapay zekâ alanındaki ilk çalışma, şaşırtıcı bir biçimde bugün de yapay zekâ alanını neredeyse tamamıyla etkisi altına almış olan, yapay sinir ağlarına dair bir çalışmadır. 1943 yılında McCulloch ve Pitts, beyin yapısından esinlenerek, ilk yapay sinir hücresi modelinin temellerini atmışlardır. Ne var ki, 1969 yılında Minsky ve Papert, Perceptrons adlı kitaplarında bu sinir ağı modelinin sadece çok basit bazı problemler dışında kullanışlı olmadığını göstermiştir. Rumelhart ve arkadaşlarının 1980'lerde Nature dergisinde yayınlanan makalelerinde ortaya koydukları, neredeyse tamamen türevde zincir kuralına dayanan, geri-yayılım algoritmasına kadar yapay sinir ağı modelleri rafa kaldırılmıştır (Rumelhart vd., 1986). Bugün ise, rüzgâr tamamen tersine dönmüş ve yapay sinir ağlarındaki katman sayısı arttırılarak elde edilen derin öğrenme yöntemi, yapay zekâ dünyasını adeta kasıp kavurmaya başlamıştır. Bugün artık yapay öğrenme ve derin öğrenmeyi eş anlamlı kullanan araştırmacılara rastlamak mümkündür.

Yapay sinir ağları, katmanlı bir yapıya sahiptir. Girdi katmanındaki nöron sayısı, gözlem verisindeki özellik sayısına eşittir. Çıktı katmanındaki nöron sayısı ise, çıktı verisinin boyutu kadardır. Ara katmanlar da ise, gözlem verileri hakkında soyutlama yapabilen nöronlar bulunur. Ağın mimarisinde belirli tasarım kuralları sabittir. Örneğin aynı katman içindeki nöronlar birbiriyle bağlantı kuramaz ve ileri katmanlardaki nöronlardan önceki katmanlardaki nöronlara (genelde) geriye doğru yönlü bağlantı kurulmaz. Bu kurallar, geri-yayılım algoritmasındaki

hatanın türev alınarak geriye doğru yayılmasını kolaylaştırmak içindir. Yapay sinir ağları da tıpkı insanlar gibi tecrübe olarak nitelendirebileceğimiz örnek verilerden öğrenir.

Başlangıç,

- Nöronlar arasındaki bağlantı ağırlık değerleri modelin parametreleridir. Gelişigüzel parametrelerle başlanır.

Her örnek girdi-çıkı çifti için,

- İleri yayılım
 - Her nöron önceki katmandan gelen girdilerin ağırlıklı toplamını alır.
 - Bu toplam, doğrusal olmayan bir aktivasyon fonksiyonuna (sigmoid vb.) sokularak ileri katmandaki nöronlara iletilir. Böylece veri içindeki doğrusal olmayan karmaşık ilişkiler yakalanabilir.
- Hata hesabı
 - İleri yayılım sona erdiğinde, çıktı katmanında üretilen değer ile gerçek değer arasındaki fark türevi alınabilir konveks bir fonksiyona verilir. Fonksiyonun çıktı değeri hata olarak belirlenir.
- Geri Yayılım
 - Nöronlarda gerçekleştirilen aktivasyon ve ağırlıklı-toplam işlemlerinin ayrıca bağlantı ağırlıklarının genel hataya olan katkıları türevde zincir kuralı ile, çıktı katmanından girdi katmanına geriye doğru, tek tek hesaplanır.
 - Hatayı minimize eden en iyi parametreleri bulmak için eğimle (gradyan) iniş algoritması temel alan bir optimizasyon algoritması kullanılır. En basit haliyle, parametreler türevin tersi yönde, türevle orantılı olarak, güncellenerek hatayı minimize eden parametre değerleri hesaplanır.

Yukarıdaki işlemler, hata belirli bir değerin altında kalıncaya dek tekrar edilir. Yapay sinir ağları, ilk kez gören araştırmacılara zor görünebilir fakat temelde yapılan şey, matris çarpımları ve türev alma gibi işlemlerin tekrarına dayanır. Çoğu deneyimli araştırmacı ise, bu kadar basit bir modelin, problem çözmede bu kadar başarılı olabilmemesine şaşırıldıklarını itiraf etmektedirler. Bu başarının arkasında yatanı anlamak için, karmaşık sistemler bakış açısının oldukça faydalı olacağını düşünüyoruz. Sistem düşüncesine geçmeden önce, eğimle (gradyan) iniş algoritmasının ya da geri yayılım algoritmasının başarısındaki simülasyonun rolünü vurgulamak istiyoruz. Bu algoritmalar, analitik çözümün doğrudan bulunamadığı durumda kullanılan yinelemeli algoritmalarlardır. Her yinelemede, ağırlıklar türevin rehberliğinde güncellenir. Dolayısıyla bu algoritmalar gücünü matematiğin alanına giren kalkülüs ve simülasyonun alanına giren yinelemenin iş birliğinden alır.

2.1. Derin Öğrenme Devrimi

Yapay Sinir Ağları katmanlı bir yapıya sahiptir. Katman sayısı fazla olduğu derin bir mimari kullanan yapay sinir ağları ile yapılan yapay öğrenme işlemine derin öğrenme adı verilir. Yapay Öğrenme kavramı, bilgisayarların bir problemi çözmekte örnek veri ya da deneyimi kullanacak biçimde programlanması olarak tanımlanmıştır (Alpaydın, 2011). Herbert Simon'a göre bir problemi çözmek büyük ölçüde problemi doğru temsil etmek ile ilgilidir. Simon'a göre problem doğru temsil edilebilirse, çözüm de görünür hale gelmiş olur. Yapay Öğrenme her ne kadar deneyim yani veri üzerinden örüntü ve kuralları bulabilse de problemin doğru temsili için insana ihtiyaç duyar. Yapay öğrenmenin bir alt kolu olan Derin Öğrenme ise farklı olarak, problemin doğru temsili kendisi bulabilir. Bu yapay öğrenme çalışmalarında çok büyük bir dönüşüme neden olmuştur. Derin öğrenme öncesinde doğal dil işleme çalışan bir araştırmacının, görüntü işleme çalışması pek mümkün değildi. Öznitelik çıkarımı ya da doğru temsil problemi araştırmacılar açısından bir tür bariyer işlevi görüyordu. Çünkü bir disipline özel bilginin transferi oldukça zahmetliydi. Derin öğrenme bu bariyeri ortadan kaldırdı.

Derin öğrenmedeki katmanlı yapıda, ilk katmanlar en basit öznelikleri otomatik bir biçimde çıkarırken, sonraki katmanlar hiyerarşik bir biçimde basitten daha karmaşığa doğru üst seviye öznelikleri çıkarabiliyor. Özel bir yapay sinir ağı olan, Evrişimsel Sinir Ağları (CNN) ile yapılan görüntü işlemi örneğinde bu durum en net şekilde görülmektedir. Örneğin yüz tanıma probleminde, ilk katmanlar kenar, çizgi gibi basit şekilleri algılayarak sonraki katmanlar göz, burun, dudak gibi formları algılayabiliyor ve en sondaki katman bunları birleşimi olan yüzü seçebiliyor.

Evrişimsel Sinir Ağları (CNN) piksellerden oluşan bir resimdeki, piksellerin komşuları ile ilgili alansal ilişkilerini otomatik olarak yakalama üzerine kuruludur. Özellikle doğal İşleme alanında kullanılan, Yinelemeli Sinir Ağları (RNN)'ler ise bir kelimenin ne olduğunu, öncesinde gelen kelimelerden çıkarır. Genel olarak CNN'ler alansal ya da mekânsal komşuluk ilişkisini kullanırken, RNN'ler zamansal sekans ilişkisini kullanır. Bir sonraki bölümde daha detaylı değineceğimiz Kompleks Sistem yaklaşımının temelinde parçalardan ziyade, parçalar arasındaki ilişkiler yatar. Dolayısıyla Kompleks Sistem ve Yapay Öğrenme tahmin edilenin çok ötesinde ortaklıklara sahiptir.

2.2. Verinin Önemi

Daha önceleri karmaşık ve zor problemleri yapay öğrenme ile bilgisayarlara çözdürmek için yepyeni algoritmik yaklaşımlara ihtiyaç duyduğumuza inanıyorduk. Gelişmeler gösterdi ki yepyeni bir yaklaşımdan ziyade, büyük miktarlardaki veriyi halihazırda elimizde olan modellere doğru şekilde dahil etmek başarımızı oldukça arttırabiliyor (Çetin, 2020). Özellikle Derin Öğrenme modelleri veriye aç modellerdir. Ne kadar veri artarsa, başarı da o kadar artar. İşte tam da bu nedenle, yeni çağın petrolü veridir. Veri, sinyal artı gürültü olarak düşünülmelidir. Ham veri içindeki değerli ve anlamlı bilgiye sinyal, geri kalana ise gürültü deriz. Önemli olan, veri işleme ile veriyi gürültüden temizleyerek değerli sinyali elde edebilmektir (Çetin & Gündoğmuş, 2019). Bugünün en modern derin öğrenme mimarileri bile öncelikli olarak ön eğitim dediğimiz aşamada daha fazla eğitim verisine ihtiyaç duyar. Fakat ancak ince ayar ile adlandırılan ikinci aşamada, eğitilmiş model az veri ile yeni bir görevi tamamlayabilir hale gelir. Ön eğitim aşamasındakinden sonra gelen bu ince ayar aşaması, bir nevi algoritmalar arası öğrenim aktarımıdır.

3. Sistem Düşüncesi

En basit tanımıyla, birbiriyle ilişkide olan elemanlar topluluğuna sistem denir. Sistemleri doğrusal sistemler ve doğrusal olmayan sistemler diye ikiye ayırmak mümkündür. Doğrusal bir şekilde bir araya gelmiş parçaların oluşturduğu sistemlere Türkçe'de dizge diyebiliriz. Fakat, doğrusal olmayan yani karmaşık ya da kompleks bir biçimde bir araya gelmiş parçalar bütününe *birge* demek gerektiğini düşünüyoruz.

Birge: Belirli bir amaca ulaşmak için, *özel bir biçimde*, bir araya gelen çok sayıdaki bileşenin oluşturduğu bütüne birge denir.

Zuhur: Sistemi oluşturan parçaların karşılıklı etkileşimi ile ortaya çıkan artı güç/iş yapabilme yeteneği.

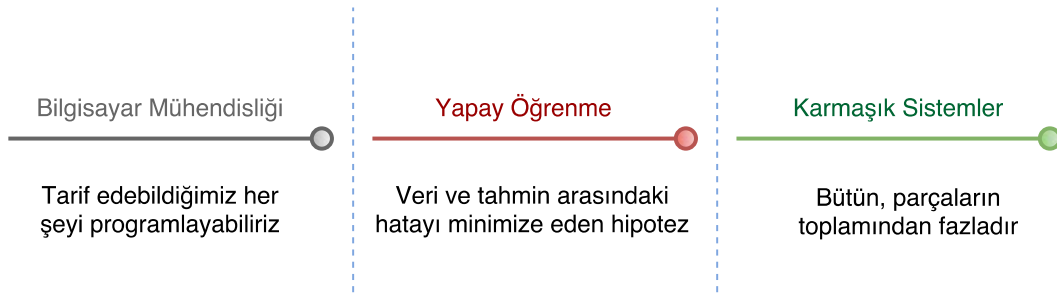
Doğrusal sistemler, $f(ax + by) = af(x) + bf(y)$ şeklinde ifade edilebilir. Doğrusal bir sistemde, bütünü çözmek ile bütünü oluşturan parçaları tek tek çözmek aynı şeydir. Dolayısıyla doğrusal sistemler anlaşılması kolay, tahmin edilebilir sistemlerdir. Girdideki ufak bir değişim, çıktıda da ufak bir değişime neden olur. Tüm bir kalkülüs bu gözleme dayanır. Ne var ki, doğrusal olmayan bir sistemde girdideki değişim, çıktıda beklenmedik değişimlere yol açabilir. Espriyle karışık bir iki örnek vermek gerekirse, bir kediyi köşeye sıkıştırırsanız kaçmaya çalışacaktır. Daha çok üzerine giderseniz, daha çok kaçacaktır. Fakat bir noktadan sonra, eğer sıkıştırmaya devam ederseniz, kaçmak yerine size saldırır. Kedileri tanımayanlar için bu çok şaşırtıcı olabilir. Şimdi de bir hükümdarın halktan aldığı vergileri arttırdığı durumu düşünelim.

Vergiler ilk birkaç kez arttığında halk perişan olup ağlaşır. Ama bir noktadan sonra, vergiler artmaya devam ederse halk oynamaya başlar. Bu davranış değişikliği, girdideki değişimin çıktıda orantılı bir değişime yol açtığı doğrusal sistemlerden beklenmeyen bir durumdur. Aynı şeyi yapıp farklı sonuçlar beklemek, doğrusal olmayan sistemlerin karakteridir. Kontrolü zor olan bu sistemlerden bilim bugüne kadar uzak durmayı tercih etmiştir.

Birge'yi parçalara ayırırsanız sistemin birliği bozulur ve bütüne ait işlev yok olur. Dizgeyi parçalara ayırmak ise işlev kaybına neden olmaz. Batı uygarlığının bir ürünü olan pozitif bilimin temeli, analitik düşüncedir. Bu bakış açısına göre, bütün kendini oluşturan daha basit parçalara bakarak anlaşılabilir. Bu yaklaşım, çoğu zaman işe yarasa da parçalar arasındaki ilişkiyi yok saydığı anda gerçek hayattaki birçok sistemi anlama da yetersiz kalır. Örneğin, atomlar hakkında bilinebilecek her şeyi bilmek, atomlardan meydana gelen tek hücreli bir canlıyı anlamaya yeter mi? Başka bir evrenden gelip ilk defa hidrojen ve oksijen görüyor olsaydık, onların bileşiminden ortaya çıkan suyun akışkanlık özelliğine sahip olacağını tahmin edebilir miydik? Akışkanlık, canlılık, bilinç ve benzeri zuhur eden birçok özellik, birge'yi oluşturan parçalarda bulunmaz ama bütünde ortaya çıkar. İlkokuldan üniversiteye tüm bir eğitim sistemi dünyanın karmaşıklığını görmezden gelen indirgemeci bir anlayışa göre dizayn edilmiş durumdadır. Bunun sebebi doğrusal olmayan sistemleri çalışmada matematiğin yetersiz kalmasıdır. Gerçek hayatın karmaşıklığı ile baş edebilecek yeni bir bilimsel araca ihtiyacımız vardır. Bu araç bilgisayarın artan gücünü kullanan etmen temelli modelleme ve simülasyondur. Bu yeni bilimsel yaklaşım genel olarak karmaşık sistem yaklaşımı olarak adlandırılmaktadır (Çetin & Bingöl, 2015).

Kompleks Sistemler, kendi kendini örgütleyebilen ve değişen şartlara uyum gösterebilen organik sistemlerdir. İnternette ekonomiye, beyinden hayvan sürülerine kadar birçok farklı sistem, merkezi bir planlayıcısı olmadan, nispeten basit kuralları takip eden çok sayıda işlemci birim'in birbiriyle olan yoğun etkileşimi ve geribildirimleri neticesinde ortaya çıkar.

Şekil 2. Bilgisayım, Yapay Zekâ ve Karmaşık Sistemler



Karmaşık sistemler yaklaşımının özellikle sosyal bilimlerde gittikçe artan bir şekilde kullanıldığına tanık oluyoruz (Çetin & Bingöl, 2014). Fakat bu yaklaşımın, karmaşıklığı katlanarak artan dünyamızda giderek bütün bilim dallarına yayılacağına inanıyoruz. Özellikle yapay sinir ağları, genetik algoritmalar, sürü optimizasyonu gibi temel yapay zekâ algoritmalarının etmen-temelli modellemedeki kökleri ve karmaşık sistemler ile bağlantılarının daha derinden irdelenmesi gerektiğini düşünüyoruz. Örneğin, yayınladığımız bir çalışmada, o şekilde programlanmalarına rağmen evrimsel süreçte nesiller zararlı ajanlara karşı belleklerini arttırmaları gerektiğini kendileri buldular (Çetin & Bingöl, 2016). Bahsi geçen çalışmada, mahkûm ikilemi oynayan etmenler, evrimsel sürecin sonunda kendiliğinden adeta bir bağışıklık mekanizması geliştirdi. Ve bu zuhur eden bir mekanizma olarak ortaya çıktı. Kullandığımız genetik algoritma, bu sonucun ortaya çıkmasında büyük bir rol oynadı. Üzerinde çalıştığımız yeni bir çalışmada, parametre optimizasyonunu geri yayılım algoritması yerine genetik algoritmalar ile yapan ve bu sayede ağın yapısını kendisi dizayn edebilen hibrit bir yapay sinir ağı üzerine çalışıyoruz. Ağın sabit katmanlı yapısına dokunulmadan, bu tür çalışmalar yapıldığını biliyoruz (Lehman & Miikkulainen, 2013). Ağın kendi kendisini organize ettiği yapay sinir ağlarına bir sonraki bölümde değineceğiz. Şekil 2'de görülmekte

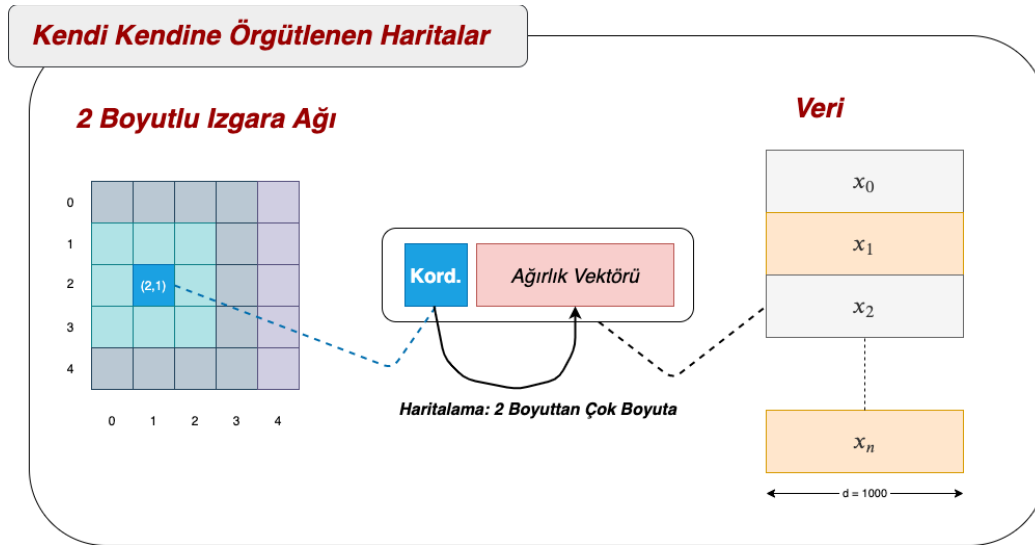
olan bilgi sayım, yapay öğrenme ve karmaşık sistemler yaklaşımlarının evreni anlamak için bize birbirini tamamlayan bakış açıları kazandıracığını düşünüyoruz.

Yapay zekâ, yapay öğrenmeyi yani istatistiksel makine öğrenmesini de kapsayan oldukça geniş bir konu. Kompleks sistemler ile bağlantı kurulabileceğimiz en heyecanlandırıcı konulardan biri kolektif zekâ algoritmalarıdır. Bu algoritmalar genel olarak doğadan esinlenen algoritmalar oluyor. Mesela karınca kolonisi algoritması yardımıyla en kısa yol problemini çözülmektedir. Genetik algoritmalar ile planlama ve optimizasyon problemlerini çözülebilmektedir. Bir sonraki bölümde detaylandıracağımız Kendi Kendine Örgütlenen Haritalar yardımıyla verinin boyutunu azaltmak ve anomali tespit etmek mümkündür. Kolektif zekâ ile, basit etmenler birlikte çalışarak zor problemleri kendiliğinden çözme yetisi gösterebilirler. Büyük bir problem, beklenmedik şekilde küçük ajanların birlikte çalışması ile çözülebilmektedir. Kompleks sistem, yapay zekâ ve veri biliminin kesişiminde bulunan bu doğadan esinlenen algoritmalara daha çok önem vermek gerektiğini düşünüyoruz.

4. Kendi Kendine Örgütlenen Haritalar (SOM)

Kendi Kendine Örgütlenen Haritalar beyinin yaşamsal çalışma yapılarından biri olan işlevsel yakınlık düzenlenişini taklit eden bir algoritmadır. Vücudumuzda birbirine yakın olan organlarla ilgilenen beynimizdeki nöronal yapılar da birbirine yakındır. Vücudumuzdaki uzamsal yakınlık, beynimizde de karşılık bulur (Schultz, 2001). Örneğin ayak ve ayak tabanı ile ilgili sinir hücreleri beyinde uzamsal olarak birbirine yakın bulunan hüreselleridir. Teuvo Kohonen 1981 yılında, bu yakınlık ilişkisini modelleyen bir yaklaşım olan Kendi Kendine Örgütlenen Haritalar'ı geliştirdi ve bilim dünyası ile paylaştı.

Şekil 3. Kendi Kendine Örgütlenen Haritalar



Kendi Kendine Örgütlenen Haritalar, türev olarak hata düzeltme yaklaşımı yerine Kompleks Sistem araştırmacılarının aşına olduğu biçimde, birbiri ile ilişkili nöronlar arasında yarışmacı bir öğrenme yaklaşımını benimser. Bu öğrenme modeli tamamen veri odaklı denetimsiz bir öğrenme modelidir. Denetimsiz öğrenme, çıktı verisi olmadan sadece girdi verisinin elimizde bulunduğu daha zor bir problemidir.

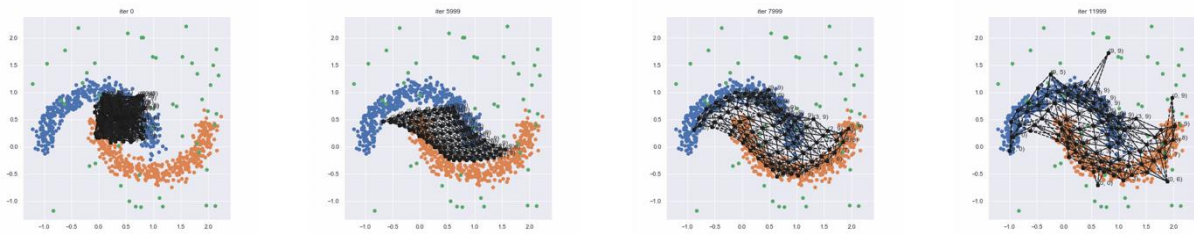
Nöronları, etmen temelli modellemedeki etmenler gibi düşünebiliriz (Çetin & Bingöl, 2015). Şekil 3'te görüldüğü gibi, nöronlar genellikle ızgara tipi bir ağ yapısı içinde bulunur. Köşe ve kenardaki nöronlar hariç, her nöronun 8 komşusu bulunur. Nöronların ızgara ağında kendi koordinatları ve komşuları bulunmaktadır. Aynı zamanda her nöronun haritalandığı çok boyutlu bir ağırlık vektörü daha bulunur. Dolayısıyla, her nöronun 2 boyutlu nöronal bir koordinatı ve eşlendiği çok boyutlu bir ağırlık vektörü bulunur. Nöron bu ağırlık vektöründen sorumludur. Beynimizdeki gerçek nöronların, vücudumuzdaki gerçek organlardan sorumlu

olmasını taklit eden bir modelleme yapılmıştır. Kendi Kendine Örgütlenen Haritalar algoritmasının genel şeması aşağıdaki gibidir.

- Ağırlık vektörlerine başlangıç değerlerini ver
- Döngü içinde
 - Rastgele bir veri noktası x_i seç
 - Bütün nöronlarla mesafesini ölç
 - (En kısa mesafeye sahip) Kazanan nöronu belirle
 - Kazanan nöron ve komşularının ağırlık vektörlerini x_i 'ye daha yakın olacak şekilde güncelle

Yukarıda tarif edilen algoritma yardımı ile, başlangıçta nöronların sorumlu olduğu ağırlık değerleri rastgele atanır. Tamamen veri odaklı olan bu algoritmada, rastgele bir veri noktası x_i seçilir. Ve bu veri noktasına en yakın olan ağırlık değerine sahip olan nöron belirlenir. Bu nöron kazanan nörondur. Kazanan nöron ve onun ızgara tipi ağdaki komşu nöronlarının ağırlık değeri, bu veri noktasına daha yakın hale gelecek şekilde güncellenir. Kazanan nöron ve komşuları çok boyutlu veri uzayında adeta veri noktası x_i 'ye doğru çekilir.

Şekil 4. Veri Topografyası



Her farklı (mavi, turuncu ve yeşil renkteki) veri noktası, (siyah renkteki) nöronların ağırlık değerlerini kendi üzerine doğru biraz daha çeker. Nöron ağırlıkları, adeta konulduğu kabın şeklini alan bir su gibi, çok boyutlu verinin gizli topografyasına uyum sağlar (Şekil 4).

Bu tekrar eden döngüdeki amaç, beynimizdeki yerel bir bölgede bulunan nöronların benzer işlev görme özelliğini taklit etmektir. Kendi Kendine Örgütlenen Haritalar'da beynimizi simgeleyen ızgara tipi ağdaki komşu nöronlar, benzer ağırlıklara sahip olarak, benzer işlev görme özelliğine sahip olur. Böylece ızgara tipi ağda, yerel kümelenmeler kendiliğinden oluşur. Her bir küme, verideki bir örüntü grubuna denk gelir. Dolayısıyla çok boyutlu verinin topografyası olduğu gibi iki boyutlu ızgara tipi nöronal ağa yansıtılır. Detaylarına girmeden kabaca anlattığımız bu yaklaşım, boyut indirme, görselleştirme, kümeleme, anomali tespiti gibi birçok alanda kullanılmaktadır.

5. Anomali Tespiti

Anomaliler, düzenliliği bozma amacıyla üretilmiş bir bakıma anti-örüntülerdir. Beklenene anlamlı bir biçimde aykırı değerlere sahiptirler. Her gürültü bir anomalidir, ama her anomali bir gürültü değildir. Gürültü yanlışlıkla kaza sonucu üretilmiştir. Anomali ise belirli bir amaca yönelik bilinçli olarak üretilmiş de olabilir. Normal veriler ve anomaliler arasındaki en temel fark niceliktir. Anormal veriler oldukça seyrekken, normal veriler ezici çoğunluğa sahiptir.

Yapay öğrenme kavramı, bilgisayarların bir problemi çözmekte örnek veri ya da deneyimi kullanacak biçimde programlanması olarak tanımlanmıştır (Alpaydın, 2011). Yapay öğrenme algoritmaları genellikle, veri içindeki düzenlilikleri ve örüntüleri yakalamaya odaklanır. Peki düzen yıkıcıları, çizginin dışındakileri, aykırı değerleri yani anomalileri nasıl yakalanabilir? Bu iş için, K-Sigma (Çetin & Taşgın, 2020), *Isolation Forest* (Liu vd., 2008) gibi çok çeşitli yöntemler mevcuttur. Biz bu çalışmada Kendi Kendine Örgütlenen Haritalar'ın anomali tespiti için nasıl kullanılabileceğine kısa bir bakış yapmak istiyoruz.

Şekil 4'te görüldüğü gibi, 2 tane ters yarım ay biçiminde olan doğrusal olmayan bir verinin topografyasına, siyah renkle gösterilen nöronal ızgara ağının iterasyonlar neticesinde nasıl serildiğini görmekteyiz. Bu aşamadan sonra, her bir nöronun ağırlık değerleri ızgara tipi ağdaki komşularının ağırlık değerleri ile karşılaştırılır. Komşu nöronlar arası ortalama mesafesi en yüksek olan nöronlar, potansiyel anomali merkezleridir. Çünkü algoritma bu mesafeyi veri odaklı bir biçimde küçültme üzerine kuruludur. Şekil 4'teki en sağdaki görselde, (9,9), (0,9), (0,6), (0,0) koordinatlarına sahip nöronların fark edilir biçimde dışarıda kaldığını görebiliyoruz. Bu nöronlara atanan veri noktaları, veri içindeki potansiyel anomali noktaları olarak tespit edilecektir.

Hayatın her alanında ender rastlanan, beklenmeyen sıra dışı olaylarla karşılaşabiliriz. Nassim Taleb bu tür olayları, yani anomalileri, siyah kuğu olarak niteliyor (Taleb, 2007/2008). Taleb'e göre şimdiye kadar etrafımızda gördüğümüz tüm kuğuların beyaz olması nedeniyle, siyah bir kuğunun var olabileceğini hayal etmememiz oldukça zordur (Hogarth vd., 2012). Siyah kuğu ile karşılaşmanın, diğer bir ifadeyle beklenmeyen bir duruma maruz kalmanın yıkıcı sonuçları olabilir. Bu durumu kavramak için Bertrand Russell'ın Tümevarımcı Hindi Hikayesi'ni hatırlamak faydalı olacaktır (Russell, 1912/2017). Bu hikâyede, yemi her gün aksatılmadan verilen bir hindinin insanın dost olduğuna dair inancı her geçen gün daha da pekişmektedir. Şükran Günü geldiğinde, elinde yem yerine bıçak ile gelen insanı gördüğü zaman, hindi geçmişe bakarak gelecek tahmininde bulunmanın fena halde yanıltıcı olduğunu anlar. Fakat artık hindi için çok geçtir. Ana soru: Hindi açısından bakıldığında kelimenin tam anlamıyla düzen yıkıcı bu olay gerçekten tahmin edilemez bir olay mıydı? Yapay öğrenme mühendislerinin, Russell'ın tümevarımcı hindisinin durumuna düşmemesi için ne yapması gerekiyor?

Veri zenginleştirme ya da veriyi çeşitlendirme, bizi hindinin düştüğü tuzaktan kurtarabilir. Modelimize, yem ile birlikte hindiyi besleyen insanın davranışlarındaki değişimi de ölçmenin ve eklemenin bir yolunu bulmamız gerekmektedir. Ayrıca, çevredeki diğer veri kaynakları ile kendi verimiz arasında ilişki kurmak da faydalı olacaktır. Daha önceki yıllarda, çevredeki farklı hindilerin başına neler gelmişti?

İnsan doğası gereği siyah kuğuları tanımak için programlanmamıştır ve hayatın doğrusal ilişkiler ile ilerlediğine inanmayı ister. Aksi takdirde hayat dayanılmaz olacaktır. Hayatın karmaşıklığına göz kapamayı isteriz. Ders kitaplarımız bile, sadece baş edebildiği doğrusal sistemleri anlatır. Halbuki ne tarih ne toplum ne de başka bir kompleks sistem aşamalar halinde düz ve pürüzsüz bir şekilde ilerler. Varlığını hissettirmeyen küçük değişimler birikerek, ani kırılmalara ve seviyeler arası sıçramalara neden olur. Bu nedenle doğrusal olmayan, kompleks sistemler bakış açısı geliştirmemiz gerekmektedir. Bugünün ve önümüzdeki on yılın bilimi veri bilimidir. Ama bu bilim dalı geleceğin bilimi olan karmaşık sistemler ile birleştirilmelidir.

6. Sonuç

Bu çalışmada, çok önemli olduğuna inandığımız bilgi sayım, veri bilimi, yapay zekâ ve karmaşık sistem yaklaşımlarını Türk bilim dünyasının dikkatine sunmayı hedefledik. Bu çalışmadaki amacımız, farklı yaklaşımlarının evreni anlamak için bize birbirini tamamlayan bakış açıları kazandıracağını vurgulamaktır. Bu yaklaşımlara yapılan eleştiriler iki başlıkta toplanmaktadır. Birincisi, yapay öğrenmede, çoğu zaman, problemin kaynağını derinlemesine kavramadan, o probleme çözüm getirmekle yetinilmesidir. Bu çalışmada bu eleştirinin o kadar da doğru olmadığını, RNN, CNN ve SOM algoritmalarının verideki ilişkileri yakalama konusunda ne kadar başarılı olduğunu anlatmaya çalıştık. İkinci eleştiri ise, Kompleks Sistemler disiplinine yapılan bir eleştiriydi. Kompleks Sistemler evrendeki sorunlarımızla ilgili bize derin bir kavrayış, güçlü bir sezgi, sağladığı halde problemlere çözüm getirmekte şimdilik biraz zorluk çektiğine dair bir eleştiriydi. Bu eleştiri cevap vermek için de esas itibari ile denetimsiz bir derin öğrenme modeli olan Kendi Kendine Örgütlenen Haritalar'ı ayrıntılı bir şekilde inceledik. Bu algoritma Kompleks Sistem araştırmacılarının aşına olduğu biçimde,

birbiri ile ilişkili nöronlar arasında yarışmacı bir öğrenme yaklaşımını benimser. Nöronlar kendi kendine organize olarak, konulduğu kabın şeklini alan bir su gibi, çok boyutlu verinin gizli topogafyasına uyum sağlar.

Karmaşıklığı giderek artan sorunlarla boğuşan dünyamızda birbirini tamamlama potansiyeli çok yüksek, bu disiplinleri birbiriyle buluşturmak yönünde adımlar atmalı, çalıştaylar düzenlemeli ve Santa Fe'ye benzer enstitüler kurmalıyız. Bu yönde atılacak her adımın çok değerli olduğuna inanıyoruz. Sizleri de Karmaşık Sistemler ve Veri Bilimi topluluğumuza bekleriz (KaVe, 2017).

Teşekkür

Bu çalışmada, Sarıyer Belediyesi Kültür Eğitim Destek Merkezi bünyesinde yürütmekte olduğum Liseler için Yapay Zekaya Giriş Eğitiminde kullandığım materyallerden faydalandım. Ayrıca Tübitak 1512 BIGG ile desteklenen Summarify Yazılım Tek. Ar. Ge. A. Ş. ve KKB Kredi Kayıt Bürosu bünyesinde Tübitak Teydeb programı ile desteklenen MPP-62 Proaktif Denetimsiz Sistem Verisinde Anomali Tespiti projesini kapsamında yürütülen çalışmalardan da faydalandım. Her iki kurumun ARGE departmanlarına, İstanbul Bilgi Üniversitesi'ne, Karmaşık Sistemler ve Veri Bilimi Topluluğu'na ve Sarıyer Akademi'ye teşekkürü borç biliriz.

Kaynakça

- Alpaydın, E. (2011). *Yapay öğrenme*. Boğaziçi Üniversitesi Yayınevi.
- Çetin, U. (2020). *Yapay öğrenme ile anomali tespiti: Siyah kuğuları yakalamak*. İdeaport.
- Çetin, U., & Bingöl, H. O. (2014). Attention competition with advertisement. *Physical Review E*, 90(3), Article 032801. <https://doi.org/10.1103/PhysRevE.90.032801>
- Çetin, U., & Bingöl, H. O. (2015, 15 Aralık). *Karmaşık sistemler ve etmen temelli benzetim* [Complex systems and agent-based simulation]. Akademik Bilişim Konferansları. <https://ab.org.tr/ab16/bildiri/370.pdf>
- Çetin, U., & Bingöl, H. O. (2016). The dose of the threat makes the resistance for cooperation. *Advances in Complex Systems*, 19(08), Article 1650015. <https://doi.org/10.1142/S0219525916500156>
- Çetin, U., & Gündoğmuş, Y. E. (2019). Feature selection with evolving, fast and slow using two parallel genetic algorithms. IEEE (Ed.), *2019 4th International Conference on Computer Science and Engineering (UBMK)* içinde (ss. 699-703). <https://doi.org/10.1109/ICSE44797.2019.00117>
- Çetin, U. & Taşgın, M. (2020). Anomaly detection with multivariate K-sigma score using Monte Carlo. IEEE (Ed.), *2020 5th International Conference on Computer Science and Engineering (UBMK)* içinde (ss. 94-98). <https://doi.org/10.1109/ICSE44797.2020.00021>
- Hogarth, R. M., & Soyer, E. (2012, Kasım). Belirsizliğin doğası: Siyah kuğular ve sisli mayın tarlaları. *Harvard Business Review - Turkey Edition*, 23-25. <https://hburturkiye.com/dergi/belirsizligin-dogasi-siyah-kugular-ve-sisli-mayin-tarlalari>
- Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Lehman, J., & Miikkulainen, R. (2013). Neuroevolution. *Scholarpedia*, 8(6), Article 30977. <https://doi.org/10.4249/scholarpedia.30977>
- Liu, F. T., Ting, K. M., & Zhou, Z. H. (2008). Isolation forest. IEEE (Ed.), *2008 Eighth IEEE International Conference on Data Mining* içinde (ss. 413-422). <https://doi.org/10.1109/ICDM.2008.47>
- Lloyd, S. (2006). *Programming the universe: A quantum computer scientist takes on the cosmos*. Vintage.
- Karmaşık Sistemler ve Veri Bilimi Topluluğu. (2017). *Home*. [LinkedIn sayfası]. <https://www.linkedin.com/company/kavetr> adresinden 02.11.2020 tarihinde erişildi.
- Kohonen, T. (1982). Automatic formation of topological maps of patterns in a self-organizing system. *Proceedings of the 2nd Scandinavian Conference on Image Analysis* içinde, (ss. 214-220).
- McCulloch, W. S., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *The Bulletin of Mathematical Biophysics*, 5(4), 115-133. <https://doi.org/10.1007/BF02400246>
- Minsky, M., & Papert, S. A. (1969). *Perceptrons, expanded edition: An introduction to computational geometry*. MIT Press.
- Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, 323, 533-536. <https://doi.org/10.1038/323533a0>
- Russell, B. (2017). *Felsefe sorunları* (V. Hacıkadiroğlu, Çev.). Kabcacı Yayınları. (Orijinal eserin yayım tarihi 1912)
- Russell, S., & Norvig, P. (2009). *Artificial intelligence: A modern approach*. Prentice Hall Press.

- Schultz, S. K. (2001). Principles of neural science. *American Journal of Psychiatry*, 158(4), 662. <https://doi.org/10.1176/appi.ajp.158.4.662>
- Spillenger, P. (Writer), & Lucas, N. (Producer). (2009). Plants (Episode 9) [Documentary]. In M. Gunton (Executive Producer), *Life*. BBC Earth. <https://www.bbc.co.uk/programmes/b00p90d6>
- Taleb, N. N. (2008). *Siyah kuğu: Olasılıksız görünenin etkisi* (N. Arıbaş Erbil, Çev.). Varlık Yayınları. (Orijinal eserin yayım tarihi 2007)