Hatasız Dağılım Öğrenme ile Sınırlı Geri Besleme Durumunda Çevrimiçi Anomali Sezimi
Online Anomaly Detection In Case Of Limited Feedback With Accurate Distribution Learning

Iman Marivani[1], Dariush Kari[1], Ali Emirhan Kurt[2], Eren Manış[2]
[1]Elektrik ve Elektronik Mühendisliği Bölümü, Bilkent Üniversitesi, Ankara, Türkiye
{marivani,kari}@ee.bilkent.edu.tr
[2]Bilgisayar Teknolojisi ve Bilişim Sistemleri Bölümü, Bilkent Üniversitesi, Ankara, Türkiye
{ali.kurt,eren.manis}@ug.bilkent.edu.tr

***Özetçe* —Ardışık anomali tespiti için yüksek performanslı bir algoritma önerilmektedir. Önerilen algoritma sıralı olarak akan veri üzerinde çalışmakta, üstel aileyi kullanarak var olan dağılımı hatasız olarak tahmin etmekte ve mevcut gözleme atanan olasılığın bir eşiğin altına düştüğünde anomali bildirmektedir. Tahmini nominal dağılım, mevcut gözleme bir olasılık değeri atamak için kullanılır ve eşiği ayarlamak için kullanıcıdan sınırlı geribildirim alınır. Algoritmanın yüksek performansı, anormal verilerin güncelleme sürecini bozmasının önlenmesi ile, nominal dağılımın doğru bir şekilde tahmin edilmesinden kaynaklanmaktadır. Yöntem, geniş bir veri dağılımı üzerinde başarılı bir şekilde çalışabilmesi açısından geneldir. Algoritmanın performansı en gelişmiş yöntemlere göre zamanla değişen dağılımlar üzerinde gösterilmektedir.**

***Anahtar Kelimeler*—anomali tespiti, çevirimiçi öğrenme, üstel aile, olasılık ataması, sınırlı geribildirim**

***Abstract*—We propose a high-performance algorithm for sequential anomaly detection. The proposed algorithm sequentially runs over data streams, accurately estimates the nominal distribution using exponential family and then declares an anomaly when the assigned likelihood of the current observation is less than a threshold. We use the estimated nominal distribution to assign a likelihood to the current observation and employ limited feedback from the end user to adjust the threshold. The high performance of our algorithm is due to accurate estimation of the nominal distribution, where we achieve this by preventing anomalous data to corrupt the update pro- cess. Our method is generic in the sense that it can operate successfully over a wide range of data distributions. We demonstrate the performance of our algorithm with respect to the state-of-the-art over time varying distributions.**

***Keywords*—anomaly detection, online learning, exponential family, likelihood assignment, limited feedback**

I. GİRİŞ

Anomali tespiti, beklenen bir davranışı takip etmeyen kalıp verileri saptama sorununu ifade eder [1]. Dolandırıcılık algılama, siber güvenlik için saldırı tespiti, hata tespiti ve askeri uygulamalarda düşman aktivitelerini kontrol etme gibi anormallik tespiti için çok çeşitli uygulamalar vardır [1]–[4]. Örneğin, bir bilgisayar ağındaki mevcut anormal veri trafiği, bir bilgisayar saldırısından, hassas verilerin yetkisiz bir yere gönderilmesi ile olabilir. [5]. Bu makalede üstel aileyi kullanarak nominal veri dağılımının doğru tahminine dayanan çevrimiçi anomali tespiti için yüksek performanslı bir algoritma önermekteyiz.

978-1-5090-6494-6/17/$31.00

Anomali sezimi, gürültü giderme [6] ve gürültü azaltma modifikasyonu [7], hepsi de verilerdeki istenmeyen ses ile uğraşan farklı görevlerdir.Verilerin kesin bir analizini yapmak için öncelikle gürültüyü veya olağandışı davranışları (kalıpları) kaldırmak gerekir. Örneğin, verilerin dağılımını tahmin etmek için, veri setindeki aykırı değerlerin (anormallikler) etkisini saptanmalı ve kaldırılmalıdır. Bu amaçla, nominal dağılımın ardışık tahmini üzerine kurulmuş ve sınırlı geribildirime dayalı dinamik eşik ayarında çevrimiçi anomali sezimi için yeni bir algoritma sunmaktayız [8].

Bu anlamda, her t adımında, bir vektör dizisinin, $\mathbf{x}_t$'nin bir örneğini gözlemliyoruz ve $\mathbf{x}_t$'nin geçmiş gözlemlere göre $(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{t-1})$ anormal olup olmadığını belirlememiz gerekiyor. İkili sistemde karar, $\hat{y}_t$ ile verilir; bu, -1 (anormal-olmayan ya da nominal veriler) ya da $+1$ (anormal) olabilir. Tahminimizi beyan ettikten sonra rasgele gözlemin gerçek etiketinin geribildirimini alıyor ve gelecek karar verme sisteminin modelini güncellemek için kullanıyoruz [9], [10].

Önerilen yöntemimizde, $\mathbf{x}_t$'yi izledikten sonra her zaman t, $\mathbf{x}_t$'ye bir olasılık atamak için geçmiş gözlemleri kullanırız. Atanmış ihtimali p_t cinsinden belirtelim. Bu olasılık kararı, üstel aile modelini kullandığımız tahmini olasılık dağılımımıza dayanmaktadır. Daha sonra, eğer $p_t < \delta_t$ ise bir anomali ($\hat{y}_t = +1$) beyan ederiz, burada δ_t olasılık üzerinde pozitif bir eşiktir. Bu yaklaşımın arkasındaki sezgisel düşünce, yeni bir gözlem $\mathbf{x}_t$'nin geçmişteki bilgilerimize dayanarak olası olmadığı durumda anormal olduğunu göstermektir. Ayrıca, çevrimiçi anomali sezimi için algoritmamızda yüksek performans elde etmek için nominal dağılımın doğru bir tahminini yapmaya çalışıyoruz. Bu amaçla, dağıtım tahmininin güncellenmesinde anormallikleri önleme olasılığına ilişkin ikinci bir eşik tanımlıyoruz. Kabaca konuşmak gerekirse, veriler arasında anormallikleri tespit edebilirz, her birinin belirli bir güveni vardır ve dağılımı güncellemek için bir anormallik olma ihtimali yüksek olanları kullanmazlar. Bununla birlikte, tahmini dağılımı güncellemek için daha az olası anormal örnekleri kullanmaya devam etmekteyiz. Tahminimiz $\hat{y}_t$, alınan geri bildirimden farklıysa, yani, $y_t \neq \hat{y}_t$ ise, o zaman bir kayıp yaşar ve eşikleri doğru bir şekilde güncelleriz.

Bu makale şu şekilde organize edilmiştir: II. Bölümde problemi tanımlıyoruz. III, önerilen modeli açıklar ve nominal verilerin dağılımını tam olarak tahmin etmek için bir yöntem sunar. IV'te, algoritmamızın performansını değerlendiriyoruz ve algoritmamızın önceki yöntemlere göre performansı önemli ölçüde geliştirdiğini gösteriyoruz. V. bölümdeki notlarla bir-

likte makaleyi tamamlıyoruz.

II. Problem formülasyonu

Makale boyunca tüm vektörler sütun vektörleri olup koyu renkli küçük harflerle gösteriyoruz. $\mathbf{x}$ ve $\mathbf{y}$'nin iç çarpımını $\langle \mathbf{x}, \mathbf{y} \rangle$ ile gösteriyoruz. Ayrıca, $\mathbf{x}_t \in \mathbb{R}^d$ mevcut gözlem ve $\mathbf{x}^{t-1} = (\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_{t-1})$, t zamanı öncesi tüm gözlemleri gösterir. Düzenli (anormal olmayan) gözlemler, X_t süreci için bilinmeyen, karşılık gelen dağılım $f_X(\mathbf{x})$ ile bağımsız ve aynı şekilde dağıtılan rastgele değişken gerçeklemeleridir.

Bu makalede, sıralı çevrimiçi anomali tespitini çalışıyoruz. Her $t = 1, 2, ...$ zamanında yeni bir gözlem $\mathbf{x}_t$ alıyoruz ve hedefimiz, önceki gözlemler $\mathbf{x}^{t-1}$'e dayanarak, $\mathbf{x}_t$'nin anormallik olup olmadığını belirlemek için anomalileri sıralı olarak tespit etmektedir. Anomali, geçmişe dayanan bilgiye dayanmayan bir gözlemdir [1]. Dolayısıyla, anomaliler, türetilen süreç X_t'den daha farklı bir kaynaktan gelmektedir. $\mathbf{x}_t$'yi değil $\mathbf{x}^{t-1}$'i gözlemledikten sonra, bu gözlemleri, üstel yayılım ailesini kullanarak $f_X(\mathbf{x})$'i tahmin etmek için kullandık [11]. Bu amaçla, $f_X(\mathbf{x})$ parametresini $\boldsymbol{\theta}_t$ parametresiyle üstel ailenin bir üyesi olarak modelliyoruz. Bölüm III'te $\mathbf{x}^{t-1}$'i kullanarak $\boldsymbol{\theta}_t$ parametresinin nasıl tahmin edileceğini gösteriyoruz. $f_X(\mathbf{x})$'i aşağıdaki gibi hesaplıyoruz

$$\hat{f}_X(\mathbf{x}_t) = e^{\langle \hat{\boldsymbol{\theta}}, \phi(\mathbf{x}_t) \rangle - T(\boldsymbol{\theta})},$$

burada, $T(\boldsymbol{\theta})$, log bölümleme fonksiyonu olarak adlandırılır ve aşağıdaki gibi tanımlanır

$$T(\boldsymbol{\theta}) = \log \int_X e^{\langle \boldsymbol{\theta}, \phi(\mathbf{x}) \rangle} d\nu(X).$$

Tanım, $\hat{f}_X(\mathbf{x})$'in 1 ile bütünleşmesi ve $\phi(\mathbf{x})$ in yeterli bir istatistik olmasıdır. $\mathbf{x}_t$'yi gözlemledikten sonra tahmin edilen dağılım $\hat{f}_X(\mathbf{x})$'ı kullanarak p_t'ı, $\mathbf{x}_t$'ın olasılığını atamak için aşağıdaki şekildeki gibi kullanıyoruz

$$p_t = \hat{f}_X(\mathbf{x}_t). \tag{1}$$

$\mathbf{x}_t$'ın anormal bir gözlem olup olmadığını belirleme olasılığına ilişkin bir eşik değeri δ_t tanımlıyoruz. Sezgisel olarak, yeni bir gözlem olan $\mathbf{x}_t$ tahmini dağılıma dayanma olasılığı düşükse bir anomalidir. Dolayısıyla, $\mathbf{x}_t$'ye atanmış olasılık δ_t'den küçükse, yani $\mathbf{x}^{t-1}$'e göre bunu beklemediğimiz anlamına gelir ve bir anomali bildiririz. $y_t = 1$ ve $y_t = -1$ etiketleri sırasıyla bir anomali ve normal bir gözlemi göstermektedir. $\mathbf{x}_t$ öngörülen etiketini $\hat{y}_t$ ile gösteririz. Anomali tespiti görevinde, N zaman adımları üzerinden algoritma tarafından yapılan hataların sayısıyla ilgileniyoruz; diğer bir deyişle, $L_{\delta_t}(\mathbf{x}^N) = \sum_{t=1}^{N} 1_{\{\hat{y}_t \neq y_t\}}$. Ayrıca, $L_{\delta_t}(\mathbf{x}^N) - L_{\delta}(\mathbf{x}^N)$, geri görüşte seçilebilecek herhangi bir sabit eşik δ'ya göre pişmanlıktır [12]. Dahası, bize doğru etiketleri sağlayacak bir uzman sistem olduğunu varsayıyoruz. Dolayısıyla, en iyi eşiğin bulunması için, eşiği güncellemek için geri bildirimi kullanıyoruz.

III. Önerilen model

Bu yazıda, ardışık çevrimiçi anomali tespiti için yüksek performanslı bir algoritma öneriyoruz ve değerlendiriyoruz. Algoritmamızı üç büyük adımda tanımlıyoruz. İlk olarak, verilerin $\mathbf{x}^{t-1}$'lerini kullanarak olasılık yayılımını tahmin ediyoruz. II. bölümde tanıtıldığı gibi üstel aileyi kullanarak bilinmeyen yayılım $f_X(\mathbf{x})$'e modelliyoruz. Sonrasında, tahmin edilen yayılımımızda $\hat{f}_X(\mathbf{x})$ yanlışlığını ölçmek için negatif bir log kaybı $l_t(\boldsymbol{\theta})$ tanımlıyoruz. Kayıp fonksiyonunu şu şekilde tanımlıyoruz

$$\begin{aligned} l_t(\boldsymbol{\theta}) &= -\log p_t \\ &= -\log \hat{f}_X(\mathbf{x}_t) \\ &= -\langle \boldsymbol{\theta}, \phi(\mathbf{x}_t) \rangle + T(\boldsymbol{\theta}), \end{aligned} \tag{2}$$

Üstel aileyi kullanmamızın iki ana nedeni vardır. Birincisi, üstel aileleri kullanarak, yeterince zengin bir dağıtım modeli sınıfı oluşturabiliriz. İkincisi, üstel ailenin [13] özelliklerinden, log bölme fonksiyonunun T üzerindeki türevlerinin yeterli istatistiğin kümülatifleri olduğunu biliyoruz. Böylece, $\phi(\mathbf{x})$'in kovaryans matrisi olan Hessen $\nabla^2 T(\boldsymbol{\theta})$, pozitif yarı-belirsizdir ve dolayısıyla $T(\boldsymbol{\theta})$'nin, θ'nın konveks bir fonksiyonu olduğu sonucuna varırız. Sonuç olarak, $\langle \boldsymbol{\theta}, \phi(\mathbf{x}) \rangle$, $\boldsymbol{\theta}$'nın doğrusal bir fonksiyonudur ve kayıp fonksiyonuda $\boldsymbol{\theta}$'nın konveks bir fonksiyonudur. Her t zamanı için, bu kayıp fonksiyonunu, stokastik gradyan düşüşü metodunu [14] kullanarak tahmin parametresini $\hat{\boldsymbol{\theta}}$ güncellemek için kullanıyoruz.

$$\hat{\boldsymbol{\theta}}_{t+1} = \hat{\boldsymbol{\theta}}_t - \eta_t \nabla l_t(\hat{\boldsymbol{\theta}}_t), \tag{3}$$

burada η_t, olumlu bir güncelleme adımı boyutudur.

Daha sonra, geçerli gözlem $\mathbf{x}_t$'ye bir olasılık atamak için tahmini yayılım $\hat{f}_X(\mathbf{x})$'i ve (1) denklemini kullanırız. Gözlem $\mathbf{x}_t$ olasılığını atadıktan sonra, benzerlik eşik değerin altına düşerse, anomali bildirme olasılığı üzerine bir eşik δ_t gerekir. Anomali tespiti için bu yaklaşım popüler ve etkilidir [1], [10], ancak bu eşiğin belirlenmesi zor bir sorundur [15]. Eşik eğerini güncellemek için, uzman sistemler varsa, insan müdahalesi gibi uzman sistemlerdeki geri bildirimleri kullanabilirsiniz. Bununla birlikte, uygulamada bir uzman sistemden veya bir kişiden gelen geribildirim almak pahalı olduğundan ve önemli çaba gerektirdiğinden, algoritmamızda rastgele istek geri bildirimi kullanırız. Tahsis edilen benzerlik δ_t eşiğine yakın olduğunda, $\hat{y}_t$'ye ilişkin kendinden emin bir tahminimiz bulunmadığını ve bu durumda daha yüksek ihtimalle geri bildirim isteyeceğimiz unutulmamalıdır. Her t zamanında, aşağıdaki olasılık ile Bernoulli rastgele değişkeni çizelim

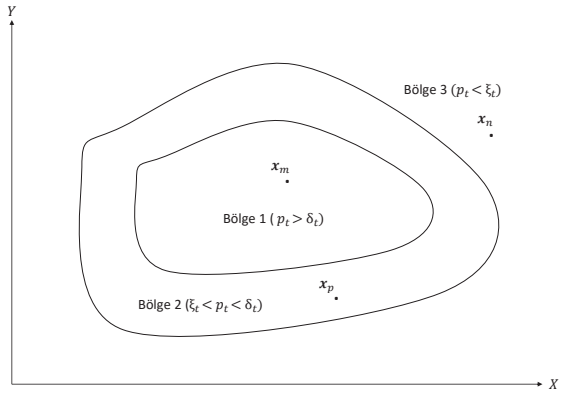
$$Pr[U_t = 1] = 1/(1 + \exp(-|p_t - \delta_t|)), \tag{4}$$

öyle ki, eğer $U_t = 1$ ise geribildirim istiyoruz [9], [16]. Uzman tarafından sağlanan, öngörülen etiket ve gerçek etiket aynı olmadığında eşiği aşağıdaki denkleme göre güncelliyoruz

$$\delta_{t+1} = \underset{\delta \in [\delta_{min}, \delta_{max}]}{\operatorname{argmin}} (\delta - \delta_t - \varphi y_t 1_{\{\hat{y}_t \neq y_t\}})^2 \tag{5}$$

burada φ, eşiği güncellemek için kullanılan pozitif bir adım büyüklüğüdür.

Şuana kadar, (1)-(5)'e dayalı anormallikleri tespit etmek için bir model kullandık. Son adımda, performansı artırmak için bir yöntem önermekteyiz. Modelimiz yayılım tahminine dayandığından tahmini hassasiyetimizi arttırarak anomali tespiti için daha iyi bir algoritma tasarlıyoruz. Anormallikleri $f_X(\mathbf{x})$ tahminini ve X_t sürecindeki verileri düzenli olarak güncellemek için kullanırsak, tahminimizin kesinliğini azalttığımızı vurguluyoruz. Bununla birlikte, hangi gözlemlerin tahminimizi güncellememizi önlemek için anomaliler olduğunu bilmiyoruz. Bu soruna hitap etmek ve doğru bir $f_X(\mathbf{x})$ tahmini elde etmek için δ_t'den küçük ikinci bir eşik ξ_t'yi tanımlıyoruz. İlk eşik δ_t, bir gözlemin anomali olup olmadığına karar vermektedir. İkinci eşik ξ_t ise tespit edilen anomaliye ne kadar emin olduğumuzu değerlendirmektedir.



Şekil 1: İki boyutlu bir veri dizisi için çift eşikli fikrin illüstrasyonudur. Her zaman t'de, bölge 1, olasılıkla δ_t'den büyük bir gözlem seti gösterir ve bu bölgedeki verileri normal olarak etiketliyoruz. 2. ve 3. bölgelerdeki herhangi bir gözlemi bir anomali olarak etiketliyoruz, ancak bölge 3 için tahminimizden daha eminiz. Dolayısıyla, nominal dağılım tahminimizi güncellemek için bölge 3'deki gözlemleri kullanmazken, bölge 2'deki verileri kullanıyoruz.

Algorithm 1 Çift eşikli anomali tespiti

Parametreler: gerçek sayılar $\delta_{max} > \delta_{min}$, $\varphi > 0$, $0 < k < 1$
Sıfırlama: $\delta_1 \in [\delta_{min}, \delta_{max}]$, $\lambda_1^{[2]} > \lambda_1^{[1]}$, $\delta_1 > \xi_1$, $\hat{\boldsymbol{\theta}}_1$
for t = 1,2,...,n **do**
  Yeni gözlem bulunması $\mathbf{x}_t$
  Olasılığın atanması $\hat{p}_t$
  if $p_t < \delta_t$ **then**
    $\hat{y}_t = 1$
    if $p_t < \xi_t$ **then**
      $\eta_t = 0$
    else
      $\eta_t = \lambda_t^{[1]}$
    end if
  else
    $\hat{y}_t = -1$
    $\eta_t = \lambda_t^{[2]}$
  end if
  Kaybı yakalamak $l_t(\hat{\boldsymbol{\theta}}_t) = -\log(\hat{p}_t) = -\langle \hat{\boldsymbol{\theta}}_t \, , \, \phi(\mathbf{x}_t) \rangle + T(\hat{\boldsymbol{\theta}}_t)$
  güncelleme $\hat{\boldsymbol{\theta}}_{t+1} = \hat{\boldsymbol{\theta}}_t - \eta_t \nabla l_t(\hat{\boldsymbol{\theta}}_t)$
  Bernoulli rastgele değişkeni U_t çizimi
  $Pr[U_t = 1] = 1/(1 + \exp(-|p_t - \delta_t|))$
  if $U_t = 1$ **then**
    Geri bildirim isteği y_t ve
    $\delta_{t+1} = \underset{\delta \in [\delta_{min}, \delta_{max}]}{\operatorname{argmin}} \; (\delta - \delta_t - \varphi y_t 1_{\{\hat{y}_t \neq y_t\}})^2$
    $\xi_{t+1} = k \times \delta_{t+1}$
  else
    $\delta_{t+1} = \delta_t$
    $\xi_{t+1} = k \times \delta_{t+1}$
  end if
end for

Tahsis edilen olasılık ikinci eşik ξ_t'nin altına düşerse, bu gözlem $\mathbf{x}_t$'yi tahmin parametresini güncellemek için kullanmıyoruz, yani, $\eta_t = 0$'yı (Şekil 1'deki bölge 3) ayarladık. Bu nedenle p_t düşük olduğunda, yüksek ihtimalle $\mathbf{x}_t$ bir anomalidir diyebiliriz. Ayrıca, ξ_t ve δ_t arasındaki muhtemel gözlem anomalileri olarak algılanır, ancak kendi etiketinden emin değiliz. Bu nedenle, tahminimizi güncellerken, bu gözlemler için küçük bir adım boyutu kullanıyoruz. Yani $\eta_t = \lambda_t^{[1]}$, burada $\lambda_t^{[1]}$, bu bölge için muhtemelen zamanla değişen spesifik bir pozitif adım boyutudur (Şekil 1'deki bölge 2). Son olarak, $\mathbf{x}_t$ olasılığı δ_t'den büyükse, onu normal bir veri örneği olarak etiketliyoruz ve bu tür veri için (Şekil 1'deki bölge 1) $\lambda_t^{[1]}$'den daha büyük bir adım boyutu tanımlıyoruz, yani $\eta_t = \lambda_t^{[2]}$.

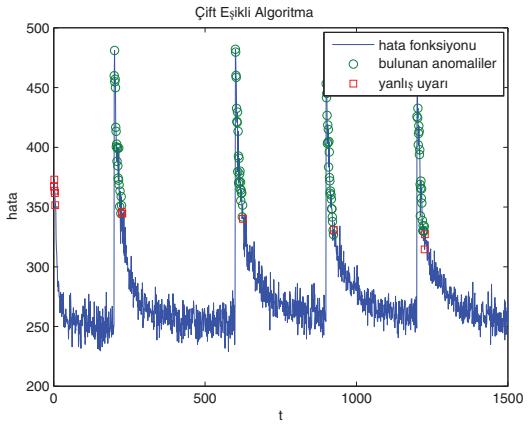
Örnek olarak, Şekil 1, iki boyutlu bir veri dizisi için çift eşikli yöntem fikrini tasvir ediyoruz. Her t zamanda, nominal veri ve iki eşik tahmini bir bölüme sahibiz. Bölge 1, muhtemel δ_t'den büyük bir gözlem seti gösterir ve bu bölgedeki verileri normal olarak olarak etiketliyoruz. 2. ve 3. bölgelerdeki herhangi bir gözlemi bir anomali olarak etiketliyoruz, ancak bölge 3 için tahminimizden daha eminiz. Dolayısıyla, nominal dağılım tahminimizi güncellemek için bölge 3'deki gözlemleri kullanmıyoruz. $f_X(\mathbf{x})$ tahminimizi güncellemek için bu kuralı kullanarak, sıralı çevrimiçi anormallik tespit ayarında iyi performans gösteren bir algoritma tasarladık. Algoritma 1, modelimizi ayrıntılı olarak açıklamaktadır. Arzulanan dağılımın kesin bir kestirimini elde etmenin diğer bir yolu, $l_t(\boldsymbol{\theta})$'yi temel alan bir adım boyutu tanımlamaktadır. Örneğin, anomalileri filtrelemek için $\eta_t = 1/l_t(\boldsymbol{\theta})$ değerini kullanabiliriz. Ancak, bu durumu tartışmıyoruz, çünkü çift eşikli algoritmada daha az güncelleme ve daha iyi performansa sahibiz. Bölüm IV'te, algoritmamızı performansını değerlendirmek için bir veri kümesine uyguluyoruz, daha sonra yöntemimizin ve önceki çalışmaların bir karşılaştırmasını sunuyoruz.

IV. DENEYLER

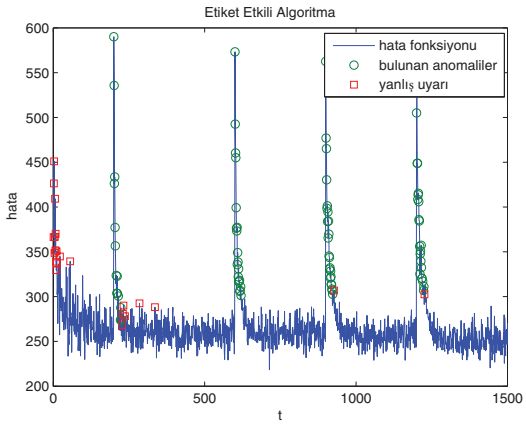
Algoritmamızda üstel ailenin kullanılması nedeniyle, çok çeşitli dağılımları modellemek için kullanılabilir [11], [13]. Özellikle, performans değerlendirmesi ve gösterimsel basitlilik uğruna, bağımsız olarak ve aynı şekilde dağıtılan rasgele değişken Bernoulli örnekleri $\mathbf{x}_t$'yi $\boldsymbol{\rho}_t = (\alpha_{1,t}, \alpha_{2,t}, \alpha_{3,t}, ..., \alpha_{500,t})$'e göre çiziyoruz; burada $\alpha_{i,t} \in [0 \; 1]$, vetör $\mathbf{x}_t$ ve $1 \leq t \leq 1500$'deki i elementi için Bernoulli parametresidir. $\boldsymbol{\rho}_t$'nin 5 zaman arasında sabit olduğunu ve değişim noktalarının $t = 200, 600, 900$ ve 1200 zamanlarında olduğunu varsayıyoruz. Bu değişim noktası veri setimizde anomaliler üretmek içindir. Hedef, geçmiş verilere göre daha az olası anomali tespit etmektir. Bunu takiben, her değişim noktasından sonra 25 gözlemin gerçek anomaliler olduğunu düşünüyoruz. Bu vektörler yeni bir dağılımdan yani bir önceki aralıktan farklı bir $\boldsymbol{\rho}_t$ ile üretilir. Dolayısıyla, mevcut dağılım tahmininde beklenmedik ve amaç onları tespit etmektir.

Bağımsız ve aynı şekilde dağıtılmış rastgele değişkenler için Bernoulli örnekleri, üstel aile bileşenlerini $\phi(\mathbf{x}) = \mathbf{x}$ ve $T(\boldsymbol{\theta}) = \log(1 + \exp(\boldsymbol{\theta}))$ olarak elde etmek basittir. Ayrıca parametreleri $\varphi = 0.1$, $\delta_{max} = 10$ ve $\delta_{min} = 0$ olarak ayarladık. Önerilen yöntem ile önceki çalışmaların performans karşılaştırması için algoritmayı ve etiket etkili algoritmayı [9], [12] bu veri üzerinde yürüttük. Algoritmalar arasında iyi bir karşılaştırma sağlamak için, etiket etkili algoritmada ve algoritmamızın 1. bölgesindeki aynı adım boyutunu kullanıyoruz (Şekil 1). Deneylerimizde farklı parametreler kullanarak, bu iki algoritmanın ROC eğrilerini Şekil 3'de tasvir ediyoruz. Şekil 3'te gösterildiği gibi. 3 algoritmamız çevrimiçi anomali tespiti için üstün bir performans sağlar.

Belirli bir parametre kümesi için, rasgele üretilen 15 farklı veri dizisi üzerinde her iki algoritmayı da yürütüyoruz ve sonuç Şekil 2'de tasvir edilmiştir.

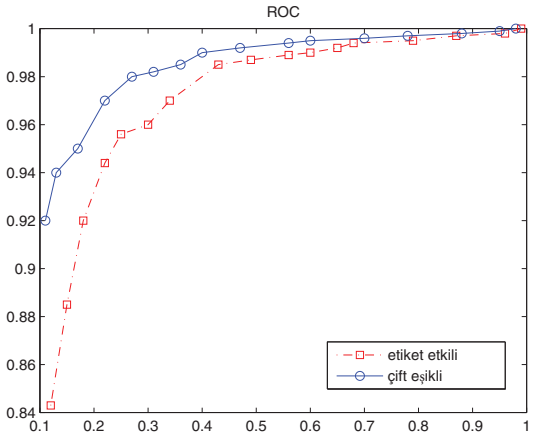


(a)



(b)

Şekil 2: Kayıp parsel üzerinde bulunan (a) çift eşikli algoritma ve (b) etiket etkili algoritma ile anomaliler (daireler ile gösterilir) ve yanlış alarmlar (kareler ile gösterilir) tespit edilir.



Şekil 3: ROC eğrilerini kullanarak çift eşikli ve etiket etkili algoritmaların performans karşılaştırması.

V. SONUÇ

Sıralı anomali tespiti için, verileri akışlı olarak aldığımız, üstel dağıtım gruplarını kullanarak nominal dağılımı doğru bir şekilde tahmin ettiğimiz ve geçerli gözlemin atanmış olasılığa bağlı olarak daha az olası olduğu zaman bir anomali beyan etmek için bir yöntem önermekteyiz. Tahmini dağılımı, mevcut gözlem olasılığını belirlemek için kullanırız. Bu amaçla, tahminimizde olasılık hakkında iki ayrı eşik belirleyerek anormal veriden kaçınırız. Sonuç olarak, algoritmamız nominal dağılımı doğru bir şekilde tahmin eder ve yüksek bir performansa ulaşır. Metodumuz, altta yatan nominal dağılımların geniş bir yelpazesinde iyi çalışabilmesi açısından jeneriktir.

KAYNAKÇA

[1] Varun Chandola, Arindam Banerjee, and Vipin Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 15:1–15:58, July 2009.

[2] J. Hong, C. C. Liu, and M. Govindarasu, "Integrated anomaly detection for cyber security of the substations," *IEEE Transactions on Smart Grid*, vol. 5, no. 4, pp. 1643–1653, July 2014.

[3] R. Buschkes, D. Kesdogan, and P. Reichl, "How to increase security in mobile networks by anomaly detection," in *Computer Security Applications Conference, 1998. Proceedings. 14th Annual*, Dec 1998, pp. 3–12.

[4] Nong Ye, Yebin Zhang, and C. M. Borror, "Robustness of the markov-chain model for cyber-attack detection," *IEEE Transactions on Reliability*, vol. 53, no. 1, pp. 116–123, March 2004.

[5] V. Kumar, "Parallel and distributed computing for cybersecurity," *IEEE Distributed Systems Online*, vol. 6, no. 10, 2005.

[6] H. S. Teng, K. Chen, and S. C. Lu, "Adaptive real-time anomaly detection using inductively generated sequential patterns," in *Research in Security and Privacy, 1990. Proceedings., 1990 IEEE Computer Society Symposium on*, May 1990, pp. 278–284.

[7] P. J. Rousseeuw and A. M. Leroy, *Robust Regression and Outlier Detection*, John Wiley & Sons, Inc., New York, NY, USA, 1987.

[8] Katy S. Azoury and M. K. Warmuth, "Relative loss bounds for on-line density estimation with the exponential family of distributions," in *MACHINE LEARNING*. 2000, p. 2001, Morgan Kaufmann.

[9] M. Raginsky, R. M. Willett, C. Horn, J. Silva, and R. F. Marcia, "Sequential anomaly detection in the presence of noise and limited feedback," *IEEE Transactions on Information Theory*, vol. 58, no. 8, pp. 5544–5562, Aug 2012.

[10] H. Ozkan, O. S. Pelvan, and S. S. Kozat, "Data imputation through the identification of local anomalies," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 26, no. 10, pp. 2381–2395, Oct 2015.

[11] Chyong-Hwa Sheu Andrew R. Barron, "Approximation of density functions by sequences of exponential families," *The Annals of Statistics*, vol. 19, no. 3, pp. 1347–1369, 1991.

[12] Nicolo Cesa-Bianchi and Gabor Lugosi, *Prediction, Learning, and Games*, Cambridge University Press, New York, NY, USA, 2006.

[13] Martin J. Wainwright and Michael I. Jordan, "Graphical models, exponential families, and variational inference," *Foundations and Trends® in Machine Learning*, vol. 1, no. 1–2, pp. 1–305, 2008.

[14] Shai Shalev-Shwartz, "Online learning and online convex optimization," *Foundations and Trends® in Machine Learning*, vol. 4, no. 2, pp. 107–194, 2012.

[15] V. Saligrama, J. Konrad, and P. m. Jodoin, "Video anomaly identification," *IEEE Signal Processing Magazine*, vol. 27, no. 5, pp. 18–33, Sept 2010.

[16] N. Cesa-Bianchi, G. Lugosi, and G. Stoltz, "Minimizing regret with label efficient prediction," *IEEE Transactions on Information Theory*, vol. 51, no. 6, pp. 2152–2162, June 2005.