Parametrik Olmayan Yoğunluk Tahmincileri ile Ardışık Anomali Tespiti Sequential Anomaly Detection Using Nonparametric Density Estimators

Mine Kerpiççi ve Süleyman S. Kozat
Elektrik ve Elektronik Mühendisliği Bölümü
İhsan Doğramacı Bilkent Üniversitesi
Ankara, Türkiye

Hüseyin Özkan
Mühendislik ve Doğa Bilimleri Fakültesi
Sabancı Üniversitesi
İstanbul, Türkiye

Özetçe —Bu bildiri, gözlemlenen verideki anomalileri, gözetimsiz bir çerçevede, iki aşamalı yöntemle bulmak için anomali tespit algoritması tanıtılmıştır. İlk aşamada, ardışık olarak gözlemlenen verinin yoğunluğu çekirdek temelli özgün bir yöntemle tahmin edilmektedir. Bu amaçla, gözlem alanı bölünmekte ve her bölgede parametrik olmayan Çekirdek Yoğunluk Tahmincisi (ÇYT) veri dağılımına dair hiçbir varsayımda bulunulmadan kullanılmaktadır. Sonra, yoğunluk tahmini eşik değeriyle karşılaştırılarak verinin anomali olup olmadığına karar verilmektedir. Ayrıca, çekirdek temelli yöntemlerdeki bant genişliği seçimi problemi de verimli bir şekilde çözülmektedir. Bu amaçla, her bir bölgeye çekirdek bant genişliği seti atanmakta ve her tahmincinin ait olduğu bölgeye göre en iyi bant genişliği seçeneğine zamanla ulaşması sağlanmaktadır. Sayısal örneklerde, tanıtılan algoritmanın literatürde sıklıkla kullanılan anomali tespit metodlarına göre yüksek performans artışı elde ettiği gösterilmektedir.

Anahtar Kelimeler—Gözetimsiz ardışık anomali tespiti, çekirdek yoğunluk tahmini, bant genişliği seçimi.

Abstract—In this paper, we introduce an online anomaly detection algorithm to detect the anomalies in the observed data with two step approach in an unsupervised framework. In the first step, we estimate the density of the sequentially observed data with a novel kernel based approach. To this end, we partition the observation space and use nonparametric Kernel Density Estimator (KDE) in each region on a partition such that we do not assume any underlying distribution for the data. Then, we compare the estimated density of the data with a threshold to decide whether it is anomalous. We also solve the bandwidth selection problem in kernel based approaches in an efficient way. For this, we assign a set of kernel bandwidth values to each region, and make each estimator to converge to the best bandwidth choice for the corresponding subspaces in time. In our experiments, we show that our algorithm significantly outperforms the anomaly detection algorithms, which are highly used in the literature.

Keywords—Unsupervised sequential anomaly detection, kernel density estimation, bandwidth selection.

I. GİRİŞ

Siber güvenlik ve gözetleme [1] gibi çeşitli alanlardaki uygulamalarından ötürü, anomali tespiti problemi literatürde

oldukça ilgi uyandırmıştır. Bu bildiri, ardışık olarak gözlemlenen verinin $x_t \in \mathbb{R}^m$ anomali olup olmadığına önceden gözlemlenen verileri $\{x_1, x_2, \dots, x_{t-1}\}$ kullanarak karar veren çevrimiçi anomali tespit algoritması tanıtılmıştır. Tanıtılan algoritma veri dağılımına dair hiçbir varsayımda bulunmadan gözlemlenen verinin yoğunluğunu tahmin etmektedir. Bu amaçla, parametrik olmayan çekirdek yoğunluk tahmin yöntemi [2], [3] rastgele Fourier özellikleri [4], [5] kullanılarak verimli bir uygulamayla kullanılmıştır. Gözlem alanı bölgelere ayrılmış ve her bölgede çekirdek yoğunluk tahmincisi bant genişlik seçeneklerini içeren bir set ile kullanılmıştır. Böylece, tanıtılan algoritma her bölge için en iyi bant genişliği seçeneklerini ayrı ayrı bulmaktadır.

Literatürde, anomali verinin tahmin edilemez ve nadir doğası nedeniyle tekdüze dağılıma sahip olduğu kabul edilmektedir [1]. Bu nedenle, Neyman-Pearson (NP) testi istenen yanlış alarm oranı için sezme gücünü maksimum yaparak, yoğunluk tahminini eşikle karşılaştırmayı önermektedir. Bu nedenle, literatürdeki en önemli anomali tespit yaklaşımlarından birisi iki aşamalı yöntem olan yoğunluk tahmini ve bu tahminin eşik değeriyle karşılaştırılmasıdır [1]. Tanıtılan yöntem de bu temele dayanmaktadır ve uygun eşik değeri seçimiyle NP testi elde edilebilmektedir.

Tanıtılan algoritma, yoğunluk tahmin yöntemini çevrimiçi durumda özgün bir yaklaşımla kullanmaktadır. Bu amaçla, hiçbir veri dağılımı varsayımında bulunmadan parametrik olmayan çekirdek yoğunluk tahmini yöntemi kullanılmıştır çünkü gerçek hayattaki uygulamaların çoğunda veri dağılımı önceden bilinmemektedir. Anomali tespiti problemi, Destekçi Vektör Makinesi (DVT) [6] metoduyla toplu veri dizisi kullanılarak çalışılmıştır. Bu nedenle, yüksek performans sağlasa da hesaplama karmaşıklıkları yüksek olduğu için gerçek hayattaki uygulamalarda kullanımı maliyetlidir. Çevrimiçi kullanılan yöntemlerden Temel Bileşen Analizine dayanan yöntem (çevrimiçi TBA) [7] ise verideki temel bileşeni kullanarak anomali olup olmadığına karar vermektedir. Bu yaklaşım düşük boyutlu veriler için yüksek performans sağlasa da, boyut sayısı arttıkça kaybolan bilgi artışından dolayı performansı düşmektedir.

Çekirdek yoğunluk tahmini yönteminde bant genişliği seçimi problemi literatürde oldukça işlenen konulardandır [8], [9]. Bu metodlar veri setinin tamamını kullanarak bant genişliğini tüm gözlem alanına göre belirledikleri için verideki bölgesel değişiklikleri takip edemezler. [10], [11] bant genişliği seçimi için çevrimiçi yöntemler sunmaktadır. Ancak bu yön-

Bu proje Türkiye Bilimsel ve Teknolojik Araştırma Kurumu (TÜBİTAK) tarafından 118E268 No.lu sözleşme ile desteklenmektedir.

nedenle, çekirdek fonksiyonu aşağıdaki şekilde yazılabilir.

$$w(\mathbf{x}_t, \mathbf{x}_k; g) = \int_{\mathbf{a}, b} f_{\mathbf{a}, b}(\mathbf{a}, b; g) \sqrt{2} \cos(\mathbf{a}' \mathbf{x}_t + b) \sqrt{2} \cos(\mathbf{a}' \mathbf{x}_k + b) \partial \mathbf{a} \partial b, \quad (7)$$

$$f_{\mathbf{a}, b}(\mathbf{a}, b; g) = (4g\pi)^{-\frac{m}{2}} (e)^{-\frac{\|\mathbf{a}\|^2}{4g}} \times \frac{1}{2\pi} 1_{\{0 \leq b \leq 2\pi\}}.$$

Böylece, (6)'deki değişim fonksiyonu şu şekilde tanımlanmaktadır.

$$\varphi(\mathbf{x}; g) \triangleq \sqrt{\frac{2}{n}} \left(\frac{g}{\pi}\right)^{m/2} \left[\cos\left(\sqrt{2g} \mathbf{a}'_1 \mathbf{x} + b_1\right), \right. \\ \left. \cos\left(\sqrt{2g} \mathbf{a}'_2 \mathbf{x} + b_2\right), \dots, \cos\left(\sqrt{2g} \mathbf{a}'_n \mathbf{x} + b_n\right) \right]'. \quad (8)$$

Yukarıdaki $\{(\mathbf{a}_i, b_i)\}_{i=1}^n$, $f_{\mathbf{a}, b}(\mathbf{a}, b; \frac{1}{2})$ dağılımının bağımsız aynı şekilde dağıtılmış örnekleridir. Dolayısıyla, $\mathbf{a}'$ 'nin elemanları birbirinden bağımsız, ortalaması 0 ve varyansı 1 olan normal dağılıma sahip rastgele değişkenlerdir. Tanıtılan algoritma, $\varphi(\mathbf{x}; g)$ üzerinde değişiklik yapılarak Gaussian çekirdek fonksiyonu dışındaki diğer fonksiyonlar için de kullanılabilir.

Tanıtılan algoritma, çekirdek yoğunluk tahminindeki bant genişliği seçimi problemini de çözmektedir. Bu amaçla, gözlem alanının her bölgesine bant genişliği seti $\mathbb{H} = \{g_1, g_2, \dots, g_{|\mathbb{H}|}\}$ atamakta ve her tahmincinin kendi bölgesi için en uygun bant genişliği değerini zamanla seçmesini sağlamaktadır. Bölgelerdeki bant genişliği değerlerinin $g \in \mathbb{H}$ olasılıkları $\rho_t^{\lambda_r}(g)$ olarak tanımlanmıştır ($\sum_{g \in \mathbb{H}} \rho_t^{\lambda_r}(g) = 1, \forall t$). İlk başta, $t = 0$ anında, bir bölgedeki bütün bant genişliği değerlerinin olasılıkları aynıdır $\rho_0^{\lambda_r}(g) = \frac{1}{|\mathbb{H}|}, \forall g \in \mathbb{H}$ ve zamanla Algoritma 1'deki gibi değiştirilirler.

λ_r bölgesindeki tahmincinin yoğunluk tahmini, setteki farklı bant genişliği değerleriyle hesaplanan tahminlerin kombinasyonudur.

$$\hat{f}^{\lambda_r}(\mathbf{x}_t) = \sum_{\forall g \in \mathbb{H}} \rho_t^{\lambda_r}(g) \hat{f}^{\lambda_r}(\mathbf{x}_t; g). \quad (9)$$

Yukarıdaki $\hat{f}^{\lambda_r}(\mathbf{x}_t; g)$, 4'deki fonksiyonun g bant genişliğiyle hesaplanmasıyla bulunan yoğunluk tahminidir.

Her tahminci sadece kendi bölgesine λ_r denk gelen veriyi $\mathbf{x}_t$ kullanır. Dolayısıyla, her t anında $\mathbf{x}_t$ verisinin denk geldiği sadece bir bölge vardır ve t anına kadar herhangi bir λ_r bölgesi için hesaplanan çekirdek fonksiyonu sayısı en fazla $t - 1$ olabilir. (6)'daki fonksiyon yaklaşımı, (9)'daki yerine yazılarak aşağıdaki ifade elde edilir.

$$\hat{f}^{\lambda_r}(\mathbf{x}_t) = \sum_{\forall g \in \mathbb{H}} \rho_t^{\lambda_r}(g) \varphi'(\mathbf{x}_t; g) \varphi_t^{\lambda_r}(g). \quad (10)$$

Yukarıdaki ifadedeki, $\lambda_r \in \{1, 2, \dots, R\}$ gözlem alanının bölgeyi temsil etmektedir ve

$$\varphi_t^{\lambda_r}(g) \triangleq \frac{1}{t-1} \sum_{r: \mathbf{x}_k \in \lambda_r, 1 \leq r < t-1} \varphi(\mathbf{x}_k; g).$$

$\varphi_t^{\lambda_r}(g)$, her bölge λ_r için ayrı ayrı tanımlanmakta ve değeri λ_r bölgesine denk gelen verilere göre güncellenmektedir.

Bant genişliği olasılık güncellemelerini oluşturmak için kayıp fonksiyonu $k(f_t(\mathbf{x}_t)) = -\log(f_t(\mathbf{x}_t))$ (negatif logaritmik

TABLO I: Doluluk veri seti için algoritmaların EAA değerleri

BÇYT	Çekirdek Yoğunluk Tahmincisi				
	$g = 0.32$	$g = 0.64$	$g = 1.28$	$g = 2.56$	$g = 5.12$
0.9558	0.6561	0.7817	0.8159	0.8386	0.9371

olabilirlik) olarak tanımlanmıştır [12]. Bu durumda, t anında g bant genişliğinin λ_r bölgesindeki olabilirliği aşağıdaki gibi gösterilmektedir.

$$Q_t^{\lambda_r}(g) = Q_{t-1}^{\lambda_r}(g) \exp(g \log \varphi'(\mathbf{x}_t; g) \varphi_t^{\lambda_r}(g)).$$

Böylece, normalleştirme işlemi aşağıdaki gibi uygulanarak, g bant genişliğinin λ_r bölgesi için olasılığı elde edilir.

$$\rho_t^{\lambda_k}(g) = \frac{Q_t^{\lambda_k}(g)}{\sum_{\forall g} Q_t^{\lambda_k}(g)}.$$

Son olarak, yoğunluk tahmini $\hat{f}(\mathbf{x}_t)$, eşik değeriyle ν karşılaştırılarak $\mathbf{x}_t$ verisinin anomali olup olmadığına (1)'deki ifadeye göre karar verilir. Bu durumda, yoğunluk tahmini eşik değerinden düşükse gözlemlenen veri $\mathbf{x}_t$ anomali, eşik değerinden yüksekse gözlemlenen veri $\mathbf{x}_t$ normaldir.

IV. SAYISAL ÖRNEKLER

Bu bölümde, tanıtılan anomali tespit yönteminin performansı doluluk veri seti [13] kullanılarak Destekçi Vektör Makinesi (DVM) [6], çevrimiçi Temel Bileşen Analizi temelli anomali tespit yöntemi (çevrimiçi TBA) [7] ve bant genişliğini gözlemlenen veriye en yakın verinin uzaklığıyla orantılı olarak değiştiren ÇYT (U-ÇYT) yöntemleriyle karşılaştırılmıştır. U-ÇYT yönteminin kullandığı bant genişliği seçim metodu literatürde oldukça kullanılmaktadır [14]. Ayrıca, tanıtılan metodun bant genişliği seçme performansını göstermek için bölgelere atanan bant genişliği değerlerinin olasılıklarının zamanla değişimi de verilmiştir. Son olarak, performans karşılaştırması için çekirdek yoğunluk tahmini yönteminin sabit bant genişliğiyle verdiği sonuçlar da eklenmiştir. Algoritmaların performansları alıcı iletim karakteristiği (AIK) altında kalan alan yani eğri altı alan (EAA) değerlerine göre karşılaştırılmıştır. Tanıtılan algoritma "Bölgesel Çekirdek Yoğunluk Tahmini (BÇYT)" olarak adlandırılmaktadır ve deneylerde bölge sayısı temsilen $R = 8$ olarak sabitlenmiştir. Ancak bu sayının artırılması performansı daha da arttıracaktır. Bant genişliği seti ise her bölge λ_r için $\mathbb{H} = \{0.01 * 2^{k-1}\}, k \in \{1, 2, \dots, 10\}$ olarak belirlenmiştir. Kullanılan doluluk veri setinde [13], dolu oda etiketi normal, boş oda etiketi ise anomali olarak düşünülmüştür.

Gözlem alanının bütün bölgelerinde bant genişliği olasılıklarının zamanla değişimi farklıdır. Örnek olarak, 7. bölgedeki bant genişliği olasılıklarının zamanla değişimi Şekil 2b'de verilmiştir. Bu sonuçlar doluluk veri seti kullanılarak $\mathbb{H} = \{0.01 * 2^{k-1}\}, \forall k \in \{1, 2, \dots, 10\}$ bant genişliği setiyle elde edilmiştir. Bu durumda başlangıç bant genişliği olasılıkları $\rho_0^{\lambda_r}(g) = \frac{1}{10}, \forall g \in \mathbb{H}$. Tanıtılan algoritma 7. bölge için zamanla artan olasılıkla $g = 5.12$ değerini seçmiştir ($g = \frac{1}{2^{6.2}}$). Bu da, $\mathbb{H}$ setindeki en küçük çekirdek bant genişliği seçeneğidir $c = 0.31$. ÇYT yöntemi az sayıda veriyle büyük bant genişliği değeri, çok veriyle ise küçük bant genişliği kullandığında yüksek performans sağlamaktadır. Dolayısıyla, gözlemlenen sayısal örnekler de bununla tutarlıdır.

Şekil 2a'da görüldüğü gibi tanıtılan algoritma BÇYT karşılaştırılan algoritmalar arasında en yüksek EAA değerine

Şekil 2: Doluluk veri seti için elde edilen (a) bant genişliği olasılıkları (b) AİK eğrileri.

ulaşarak en iyi performansı sağlamıştır. Bunun nedeni, tanıtılan algoritmanın bölgesel yaklaşımı sayesinde veri dizisindeki bölgesel değişimlere duyarlı olması ve böylece verinin olasılık dağılımını doğru olarak takip edebilmesidir.

Aynı zamanda, tanıtılan algoritma hesaplama maliyeti açısından da oldukça avantajlıdır. Örneğin, DVT için hesaplama karmaşıklığı eğitim aşamasında $O(T_e^3)$, test aşamasında $O(T_e T_{test})$ (T_e : eğitim için kullanılan veri sayısı, T_{test} : test için kullanılan veri sayısı). Öte yandan, tanıtılan algoritma hiçbir eğitim aşaması gerektirmemektedir ve hesaplama karmaşıklığı sadece $O(kT)$ (T : toplam veri sayısı, k : bant genişliği setindeki toplam eleman sayısı). Böylece, tanıtılan algoritma oldukça düşük hesaplama karmaşıklığıyla çevrimiçi çalışmaktadır.

Ayrıca, tanıtılan özgün algoritmayla elde edilen performans artışını göstermek için, tüm gözlem alanı için sabit bant genişliği kullanan ÇYT yöntemi de Tablo I'e eklenmiştir. Tanıtılan algoritma BÇYT'de kullanılan bölgesel yaklaşımla bant genişliği değeri değiştirilmesi veri setinin yoğunluğunun takip edilmesine olanak sağlamaktadır. Bu nedenle, Tablo I'de görüldüğü gibi BÇYT, sabit bant genişliği kullanan ÇYT yöntemine göre önemli bir performans artışı sağlamaktadır. Ayrıca, sabit bant genişliği kullanan ÇYT için en uygun bant genişliğini çevrimiçi ortamda önceden bilmek mümkün olmayacaktır. Öte yandan, tanıtılan algoritma BÇYT, bant genişliğini verilen sete göre zamanla değiştirerek sonuçlardaki yüksek performansı çevrimiçi ortamda elde etmektedir.

V. SONUÇLAR

Bu bildiriye, eğitimsiz çerçevede çalışan anomali tespit metodu tanıtılmıştır. Tanıtılan algoritma parametrik olmayan çekirdek yoğunluk tahmini temelli olup çevrimiçi çalışmaktadır. Öncelikle, gözlemlenen verinin yoğunluğu, veri dağılımına dair hiçbir varsayımda bulunulmadan, geçmiş gözlemler kullanılarak özgün bir yöntemle tahmin edilmektedir. Sonra, yoğunluk tahmini eşik değeriyle karşılaştırılarak gözlemlenen verinin anomali olup olmadığına karar verilmektedir. Aynı zamanda, tanıtılan özgün yöntemle, çekirdek yoğunluk tahmini yöntemindeki bant genişliği seçimi problemi de verimli bir şekilde çözülmüştür. Bu amaçla, gözlem alanı bölgelere ayrılarak her bölgeye bant genişliği seti atanmış ve her tahmincinin kendi

bölgesi için en uygun bant genişliği değerini zamanla değiştirdiği olasılıklara göre seçmesi sağlanmıştır. Sayısal örneklerde, literatürde yaygın olarak kullanılan anomali tespit metodlarına göre önemli performans artışı olduğu gösterilmiştir.

KAYNAKLAR

- [1] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 15:1–15:58, Jul. 2009.
- [2] B. W. Silverman, *Density estimation for statistics and data analysis*. CRC press, 1986, vol. 26.
- [3] T. Ahmed, "Online anomaly detection using kde," in *GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference*, Nov 2009.
- [4] A. Rahimi and B. Recht, "Random features for large-scale kernel machines," in *Advances in neural information processing systems*, 2008.
- [5] F. Porikli and H. Ozkan, "Data driven frequency mapping for computationally scalable object detection," in *2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, Aug 2011, pp. 30–35.
- [6] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.
- [7] Y. Lee, Y. Yeh, and Y. F. Wang, "Anomaly detection via online oversampling principal component analysis," *IEEE Transactions on Knowledge and Data Engineering*, vol. 25, no. 7, July 2013.
- [8] M. C. Jones, J. S. Marron, and S. J. Sheather, "A brief survey of bandwidth selection for density estimation," *Journal of the American Statistical Association*, vol. 91, no. 433, pp. 401–407, 1996.
- [9] S. J. Sheather and M. C. Jones, "A reliable data-based bandwidth selection method for kernel density estimation," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 53, no. 3, 1991.
- [10] M. Kristan, D. Skočaj, and A. Leonardis, "Online kernel density estimation for interactive learning," *Image and Vision Computing*, vol. 28, no. 7, pp. 1106–1116, 2010.
- [11] M. Kristan, A. Leonardis, and D. Skočaj, "Multivariate online kernel density estimation with gaussian kernels," *Pattern Recognition*, vol. 44, no. 10, pp. 2630 – 2642, 2011.
- [12] E. Hazan, A. Agarwal, and S. Kale, "Logarithmic regret algorithms for online convex optimization," *Machine Learning*, vol. 69, Dec 2007.
- [13] L. M. Candanedo and V. Feldheim, "Accurate occupancy detection of an office room from light, temperature, humidity and co2 measurements using statistical learning models," *Energy and Buildings*, vol. 112, 2016.
- [14] L. J. Latecki, A. Lazarevic, and D. Pokrajac, "Outlier detection with kernel density functions," in *Machine Learning and Data Mining in Pattern Recognition*, P. Perner, Ed., 2007.