

CSE 454 Veri Madenciliği

Ödev 2 Rapor

Ahmed Semih Özmekik
171044039

17 Aralık 2020

Özet

Seçtiğimiz konferans makalesi: Y. Wang et al., Iterative anomaly detection, 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, 2017, pp. 586-589.

Bu raporda konferansın detayları, makalenin yöntemi, sonuçları, sonuçların literatürle tartışılması ve literatüre katkısı anlatılacaktır.

1 Giriş

Gerçek dünya veri kümelerinde çoğu zaman bazı vakaların (veri noktalarının), verilerin çoğundan farklı davrandığı görülür. Bu tür veri noktaları, makine öğreniminde anomali-ler ve istatistiklerde aykırı değerler olarak adlandırılır. Aykırı değerler hatalardan kaynaklanıyor olabilir, ancak istisnai durumlarda kaydedilmiş olabilir veya başka bir popülasyona ait olabilir. Ya verilerden çıkarılan sonuçlar üzerinde zararlı bir etkiye sahip olabilecek ya da değerli bilgileri içerebilecek anormal durumları tespit edebilmek çok önemlidir.¹

Anomali tespiti (AD)² son yıllarda çok ilgi gören bir alandır. Dolandırıcılık tespiti ve ağ saldırı tespiti dahil olmak üzere çok çeşitli uygulamalara sahiptir.³ Sensörler gibi çeşitli gelişen teknolojilerin hızlı büyümesi, bağlı cihazlar, akıllı ev aletleri, akıllı şehirler, 5G iletişim ortamı, akıllı telefonlar, mobil bulut, sağlık uygulaması, multimedya, sanal gerçeklik ve otonom otomobiller, bir ağda akacak devasa gerçek zamanlı veri birikimine katkıda bulunmuştur. Beklenen bu büyüme büyük veride güvenlik için tehlike uyandırmıştır ve dolayısıyla anomali tespiti önemli bir alan haline gelmiştir.⁴

AD, birçok bilinmeyen (görsel olarak gözlemlenemeyen veya önceki bilgilerle tespit edilemeyen) hedefi ortaya çıkarmayı sağlayan hiperspektral görüntüleme teknolojisindeki ilerlemeler sayesinde son yıllarda büyük ilgi gördü.

Bu raporda, AD üzerine yayınlanan bir konferans makalesini inceleyerek, makaledeki yöntemi, sonuçları ve katkıları tartışacağız.

¹Riveiro, M, Pallotta, G, Vespe, M. Maritime anomaly detection: A review. WIREs Data Mining Knowl. Discov. 2018; 8:e1266.

²Bu raporun geri kalanında bir anomali tespiti belirtilen algoritmalar için AD kısaltması kullanılacaktır.

³Caleb C. Noble and Diane J. Cook. 2003. Graph-based anomaly detection. In Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '03). Association for Computing Machinery, New York, NY, USA, 631–636.

⁴Ariyaluran Habeeb, R. A., Nasaruddin, F., Gani, A., Targio Hashem, I. A., Ahmed, E., Imran, M. (2018). Real-time big data processing for anomaly detection: A Survey. International Journal of Information Management.

2 Konferans

IGARSS (IEEE International Symposium on Geoscience and Remote Sensing), Yerbilimi ve Uzaktan Algılama Derneği (Geoscience and Remote Sensing Society)'nin en iyi konferansıdır. Amerika, Avrupa veya Afrika, Asya veya Avustralya'da üç yıllık bir döngüye göre seçilen yerlerde yıllık olarak düzenlenmektedir.

IGARSS bugün uzaktan algılamada önde gelen bir etkinlik olarak kabul edilmektedir ve en son gelişmeler hakkında güncel bilgiler elde etmek, fikir alışverişinde bulunmak, araştırma alanında gelecekteki eğilimleri belirlemek ve uluslararası uzaktan algılama topluluğu ile iletişim kurmak için ideal bir forum sağlamaktadır.

Seçtiğimiz makale, IGARSS 2017 konferansında yayınlanan, *Iterative anomaly detection*⁵ adındaki makaledir. Sonraki bölümlerde bu makaledeki yöntemi, deneysel sonuçları ve makalenin literatürüne katkısını inceleyeceğiz.

3 Yöntem

Çalışma, anomali tespitinin tamamen kör bir ortamda (herhangi bir ön bilgiye sahip olmadan), gerçekleştirilmesi gerektiğine ve bundan hareketle iki noktanın göz önüne alınması gerektiğine vurgu yapmaktadır: Birincisi, herhangi bir ön bilgi kullanmadan anomalilerin nasıl bulunacağı sorusudur. İkincisi ise anomaliler tespit edildiğinde bir anormalliğin diğerinden nasıl ayırt edileceği sorusudur.

3.1 Diğer Yöntemler

Çalışma önce, literatürde ilk soruna yönelik geliştirilen bazı diğer çalışmaları göstererek, bunlardaki anomali detektörlerinin hepsinin genel bir problemten muzdarip olduğunu altını çizmektedir: Bütün bu anomali dedektörlerinde anomalilerin segmentasyonunun yapıldığı noktada uygun ve belirli bir eşik değerine ihtiyaç bulunmaktadır. Bu bir problem teşkil etmektedir.

İkinci probleme (anomalilerin ayrıştırılması ve kategorizasyonu) yönelik olarak geliştirilen çalışmalardan bahsederek, sınıflandırma probleminin üzerinde bu çalışmanın kısa/üstünkörü durduğunu altını çizmektedir.

Çalışmada referans olarak alınan iki AD çalışmasını inceleyelim.

3.1.1 K-AD⁶

Bu yöntem, verinin kovaryans matrisi \mathbf{K}' 'yı kullandığı için, **K-AD** olarak adlandırılmaktadır ve $\delta^{K-AD}(r)$ şeklinde tanımlanmıştır:

$$\delta^{K-AD}(r) = (r - \mu)^T K^{-1}(r - \mu) \quad (1)$$

⁵Y. Wang et al., Iterative anomaly detection, 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Fort Worth, TX, 2017, pp. 586-589.

⁶I.S. Reed and X. Yu, Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution, IEEE Trans. on Acoustic, Speech and Signal Process. , vol. 38, pp. 1760-1770, 1990.

1. r veri vektörü,
2. μ verilerin ortalama vektörü,
3. K verilerin kovaryans matrisidir.

3.1.2 R-AD⁷

K-AD'den farkı **K** veri kovaryans matrisinin, veri korelasyon matrisi olarak alınmasıdır ve $\delta^{R-AD}(r)$ şeklinde tanımlanmıştır:

$$\delta^{R-AD}(r) = (r)^T R^{-1}(r) \quad (2)$$

3.2 Yinelemeli Anomali Tespiti (IAD)

Çalışma, anomali tespitini yinelemeli olarak gerçeklemektedir. Bir anomalinin varlığı, çevresindeki komşularıyla, yani bir komşuluk durumuyla yakından ilişkili olduğundan, varlığının tespiti için uzamsal veriler çok önemlidir. Önerilen IAD, anomali tespitinde uzamsal bilgi toplamak için bir uzamsal filtreden yararlanır. Fikirde önce, bir Gauss penceresi sayesinde anomali tespit haritası yakalanır, ardından ise bu haritaya bir Gauss filtresi uygulanmaktadır.

IAD'ye göre anomaliler genellikle uzamsal olarak onları çevreleyen veri vektörleriyle bağlantılı olduğundan AD'ye uzamsal bilginin eklenmesi yararlı olacaktır. **K-AD** ve **R-AD**'nin ikisi de, yani bahsi geçen önceki yöntemler, uzamsal korelasyonu hesaba katmamaktadır. IAD, bu yerel uzamsal bilgileri dahil ederek AD'nin yeni bir versiyonunu geliştirmiştir.

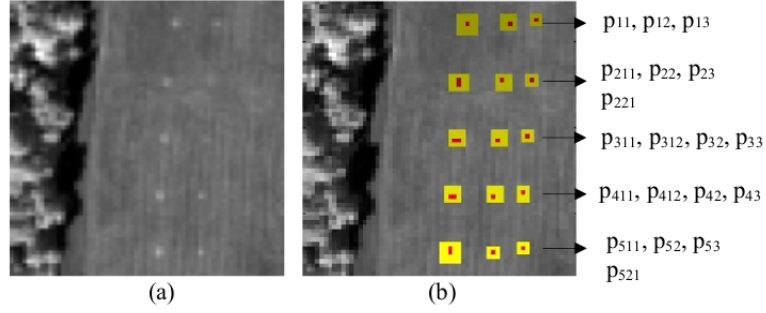
IAD'nin algoritması şu şekilde aktarılabilir (AD ya R-AD ya da K-AD olabilir):

1. (a) $\Omega^{(0)} = \{B_l\}_{l=1}^L$ orijinal hiperspektral görüntü olsun.
 (b) $\{r_i^{(0)} = (r_{i1}, \dots, r_{iL})^T\}_{i=1}^N, \Omega^{(0)}$ içinden bir veri vektörü olsun.
 (c) $\delta^{R-AD}(r), \Omega^{(0)}$ üzerinde bir AD (K-AD ya da R-AD) ve $k = 0$ olsun.
2. $k = k + 1$ olsun. Her k 'inci iterasyonda, Gaussian filtresiyle $|B|_{AD}^{(k)}$ 'yi blurlanır. $|B|_{AD}^{(k)}$, burada $\delta_{k-1}^{AD}, B_{AD}^k$ tarafından üretilen tespit haritası olmaktadır. Sonuç görüntü, $|B|_{GFAD}^{(k)}$ Gaussian filtresi ile belirtilmiştir.
3. $\Omega^k = \Omega^{k-1} \cup \{|B|_{GFAD}^{(k)}\}$. Anomali sınıflarını temsil eden $T^{(k)}$ sınıfları bulunur.
4. Eğer $T^{(k)}$ sınıfı durma kuralını sağlıyorsa, 5.adıma, sağlamıyorsa 2.adıma geçilir.
5. $T^{(k)}$ istenen sonuç olan, etiketlenmiş veri kümesidir.

⁷C.-I Chang, Hyperspectral Imaging: Techniques for Spectral Detection and Classification , Kluwer Academic Publishers, New York, 2003.

4 Sonuçlar

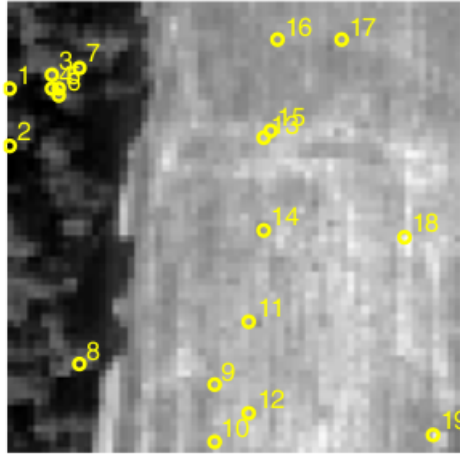
Aşağıdaki figürde **HYDICE** verikümesinin bir örneği (a) gösterilmektedir. Veri kümesinde 3 farklı boyutta, 3×3 , 2×2 , 1×1 15 panel bulunmakta olup, bunlar (b)'de gösterilmiştir.



Şekil 1

(b)'de bu 19 R'nin uzamsal konumları gösterilmiştir. Kırmızı piksellerin (R pikselleri) panel merkezi olduğu panel pikselleri pikseller ve sarı renkli pikseller (Y piksel) panel pikselleridir.

Yukarıdaki figür temel referans değerlerini belirtmektedir. Bir sonraki resme baktığımızda sistemin sonuçları görmekteyiz.



Şekil 2

5 Katkı

AD literatürde geniş çapta incelenmiştir, ancak hedef sınıflandırmayı gerçekleştirmek için AD'nin nasıl kullanılacağı araştırılmamıştır. Bu çalışma, ilgilenilen hedefleri denetimsiz (unsupervised) bir şekilde sınıflandırmak için (CEM ile) AD'nin, yinelemeli bir sürümünü (IAD) geliştirmektedir.

IAD'deki esas ve katkı, Gauss filtreli uzamsal bilgileri, yinelemeli olarak ve geri bildirim döngüleri aracılığıyla sisteme dahil etmektir. Ardından ATGP'yi tespit edilen anor-

mallikler arasında ayırt etmek ve sınıfları bulmak için kullanmaktadır, böylece her bir temsilci/sınıf farklı bir anormallik sınıfı belirtir.

IAD'nin literatür yaptığı katkıyı özetleyecek olursak, bahsi geçen diğer AD methodlarını tartışırken söylediğimiz gibi, her biri uzamsal bilgileri, yani bir veri noktasına ait komşuluk bilgisini hesaba katmamaktadır. Fakat IAD bu uzamsal bilgiyi sistemine dahil ederek katkı sunmuştur.