

Arıza etiketi yokken: endüstride denetimsiz makine öğrenmesi

Etiketsiz sensör geçmişinden neyin öğrenilebileceğini, etiket olmadan dedektörlerin nasıl değerlendirileceğini ve çalışma modlarının neden önce ele alınması gerektiğini öğrenin.

Kısaca

- Tesislerde sensör geçmişi bol, etiketli arıza kaydı ise azdır; bu yüzden çoğu yöntem normal öğrenip sapmaları işaretler.
- Denetimsiz bir model “olağandışı” diyebilir, ama “arıza” diyemez; her uyarının nedeni operatör incelemesiyle doğrulanmalıdır.
- Etiketsiz bir modeli, geçmiş olayların yeniden oynatılması, haftalık uyarı sayısı ve zamana göre bölünmüş verilerle değerlendirin.

01 / Etiket neden az olur

Bir tesisin geçmiş veri arşivinde yıllarca sensör kaydı birikmiş olabilir, ama bu kayıtların içinde etiketlenmiş arıza genellikle azdır. Arızalar nadirdir; bakım kayıtları eksik, tutarsız ya da serbest metin olabilir; her olayın kök nedeni hiç yazılmamış olabilir. Denetimsiz öğrenme, eğitimde hedef etiket kullanmayan yöntemleri anlatır. Bu yöntemler normal davranışı öğrenir ve normalden sapan anları işaretler. Etiketli arıza azken normal veri genellikle bolca bulunur; yöntem seçimi de bu dengeye göre yapılır.

Burada dil dikkatli seçilmelidir. Model “olağandışı” diyebilir, “arıza” diyemez. Bir noktanın olağandışı olması arızalı olduğu anlamına gelmez; başka bir çalışma koşulu, bir ölçüm hatası ya da yeni bir ürün reçetesi de aynı sapmayı üretebilir. Bu ayrım raporda ve operatör eğitiminde açık tutulmalıdır.

02 / Yöntem aileleri

Birincisi istatistiksel sınırlardır: her sensörün normal aralığını, ortalama ve standart sapmasını izler. Basit ve açıklanabilir, ama değişkenler arasındaki ilişkiyi görmez.

İkincisi mesafe ve yoğunluk tabanlı yöntemlerdir. Mahalanobis uzaklığı, değişkenler arasındaki ilişkiyi hesaba katarak bir noktanın normal bulutun ne kadar dışında olduğunu ölçer. Üçüncüsü izolasyon ağaçlarıdır; az ve farklı noktaların az bölmeye ayrıldığı fikrine dayanır. Dördüncüsü yeniden yapılandırma yöntemleridir: PCA tabanlı bir model normal veriyi düşük boyutlu bir yapıya sıkıştırıp geri açar, geri açılmayan kısım sapmayı gösterir. Buna ek olarak, çalışma modlarını bulmak için k-means gibi kümeleme yöntemleri kullanılır. Hangi aileyi seçeceğinize verinin boyutuna, değişkenler arası bağa ve açıklanabilirlik ihtiyacına bağlıdır. Her ailenin varsayımı farklıdır; aynı veriyi iki yöntemle çalıştırıp sonuçları karşılaştırmak, tek bir yöntemle bağlı kalmaktan daha bilgilendiricidir.

03 / Çalışma modları önce gelir

Bir makinenin tek bir “normal” hâli olmayabilir. Bir pompa iki farklı devirde, bir hat üç farklı ürün reçetesinde çalışıyorsa her mod kendi normal dağılımına sahiptir. Modları ayırmadan tek bir model kurarsanız, meşru bir mod geçişi bile sürekli alarm üretir.

İki çözüm yaygındır. Birincisi, önce kümeleme ile modları bulup her mod için ayrı bir model kurmaktır. İkincisi, çalışma modunu gösteren değişkeni (devir, ürün kodu, yük) modele katmaktır. Mod bilgisi yoksa, devir ya da yük değişimlerini gösteren bir sinyal üzerinden geçiş anlarını bulun ve değerlendirmeden çıkarın. Modları bulduktan sonra her mod için ayrı eşik belirlemek sonucu daha açıklanabilir yapar.

04 / Etiket olmadan değerlendirme

Doğruluk yüzdesi etiket gerektirir; etiket yoksa birkaç dolaylı ölçüt kullanılır. Birincisi geçmiş olayların yeniden oynatılmasıdır: model o dönemde ne zaman uyardı ve bakım ekibi ne zaman müdahale etti? İkincisi operatörlerin işaretli dönemleri incelemesidir. Üçüncüsü haftalık uyarı sayısıdır.

Hipotetik bir hesap: 12 haftalık bir dönemde 48 uyarı çıktıysa, haftalık ortalama $48 / 12 = 4$ uyarıdır. Bilinen 5 olaydan 3’ü önceden işaretlendiyse yakalama oranı $3 / 5 = \%60$ olur. Ancak 5 olay istatistik için çok azdır. Bu üç olayın önceden işaret süreleri 36, 10 ve 4 saat ise ortalama 10 saattir. Son olarak değerlendirmeyi zamana göre bölün: verileri rastgele karıştırmak, gelecekteki bilgiyi eğitime sızdırır. Geçmiş olay sayısı ne kadar azsa her bir olayın sonuca etkisi o kadar büyüktür; bu yüzden tek bir olaydan genel bir başarı sonucu çıkarmayın.

05 / Sınırlar ve yanlış okumalar

Denetimsiz bir model arızanın türünü ya da kök nedenini bilmez. Yalnızca verinin eğitim dağılımına benzemediğini söyler. Eğitim verisinde zaten bir arıza dönemi varsa model o dönemi normal sayabilir. Mevsim değişimi, yeni hammadde veya sensör değişikliği gibi kaymalar da yanlış alarm üretir. Bu nedenle sonuç bir teşhis değil, bir inceleme önceliği olarak sunulmalıdır.

Sık atlanan bir risk de ölçüt seçimidir. Haftalık uyarı sayısı düşük görünse bile bakım ekibi bu uyarıları gerçek bir olaya bağlayamıyorsa model işe yaramıyordur. Ortalama bir ölçütün kötü bir dönemi gizleyebileceğini de aklınızda tutun. Bir uyarı, arızanın başlangıcından sonra da gelebilir; bu yüzden uyarı süresini bakım kayıtlarındaki olay zamanıyla karşılaştırarak ölçün.

06 / Uyarıyı karara ve etikete dönüştürmek

Ham uyarı bir karar değildir. Bir triyaj kuralı tanımlayın: örneğin üç ardışık pencerede eşik aşılsa operatöre bildirim gitsin, tek bir sıçrama ise yalnızca kayda geçsin. Her uyarı için operatörün verdiği karar (gerçek, yanlış ya da belirsiz) saklanmalıdır.

Bu geri bildirim zamanla bir etiket setine dönüşür. Yeterli sayıda doğrulanmış olay ve tutarlı bir etiketleme süreci oluştuğunda denetimli bir model düşünülebilir. O zamana kadar denetimsiz yöntem, etiket üretmenin bir aracıdır. Bu geçişin hızı, sahadaki kararların ne kadar tutarlı kaydedildiğine bağlıdır. Operatör geri bildirimini ayrı bir tabloda tutun; her kayıta tarih, varlık, uyarı zamanı ve karar alanları bulunsun.

Uygulayın

Sentinel’de örnek sensör tablosundaki iki işaretli dönemi karşılaştırın; birinin bir çalışma modu geçişine, diğerinin gerçek bir sapmaya benzeyip benzemediğini bir cümleyle yazın.

Kaynaklar ve çevrimiçi rehber

Güncel çevrimiçi rehberi açın

Unsupervised learning (Wikipedia)

Anomaly detection (Wikipedia)

scikit-learn: novelty and outlier detection

k-means clustering (Wikipedia)

Önerilen pilot adımları saha doğruluk iddiası içermez. Sentetik örnekler gerçek tesis ölçümü değildir.