

“Satellite” Görüntü Veri Seti Üzerinde Çalışmalar

HAZIRLAYAN: RECEP BATTAL

PROJENİN AMACI

Bu proje, uydu görüntülerini sınıflandırmayı amaçlamaktadır. Çeşitli makine öğrenimi algoritmaları kullanarak, uydu görüntülerini belirli kategorilere (bulutlu(cloudy), çöl(desert), yeşil alan(green_area), su(water)) ayırmak hedeflenmiştir.

Uydu görüntülerinin otomatik olarak sınıflandırılması, uzaktan algılama (RS) görüntülerinin yorumlanmasında büyük ilerlemeler sağlamaktadır.

MATERYAL - METHOD

Bu projede kullanılan yöntemler:

- **Kütüphaneler:** pandas, numpy, matplotlib, seaborn, scikit-learn, xgboost, keras, time
- **Algoritmalar:** K-Nearest Neighbors, Random Forest Classifier, Decision Tree Classifier, Support Vector Machine, XGBoost Classifier, Gradient Boosting Classifier
- **Görüntü İşleme:** Keras kütüphanesi ile görüntülerin yüklenmesi ve numpy dizilerine dönüştürülmesi

VERİ SETİ - VERİ KULLANIMI

Context: Bu veri seti, 4 farklı sınıftan oluşan uydu görüntülerini içermektedir. Sensörler ve Google harita görüntüleri karıştırılmıştır.

Content: Son yıllarda uzaktan algılama (RS, 'remote sensing') görüntülerinin yorumlanmasında büyük ilerlemeler kaydedilmiştir. RS görüntülerine erişim daha kolay hale geldikçe, bu görüntülerin otomatik yorumlanmasına olan talep artmaktadır.

Bu bağlamda, benchmark veri setleri, akıllı yorumlama algoritmalarının geliştirilmesi ve test edilmesi için temel ön koşullar olarak hizmet etmektedir.

KULLANILAN MODELLER/MİMARİLER

Kullanılan Modeller:

- K-Nearest Neighbors (KNN)
- Random Forest Classifier
- Gradient Boosting Classifier
- Decision Tree Classifier
- Support Vector Machine (SVM)
- XGBoost Classifier

Bu modeller, eğitim ve test verileri üzerinde eğitilmiş elde edilen değerler ve eğitim süreleri açısından değerlendirilmiştir.

DENEY SONUÇLARI

Modellerin performansı çeşitli metrikler kullanılarak değerlendirilmiştir:

- Doğruluk (Accuracy)
- Hassasiyet (Precision)
- Duyarlılık (Recall)
- F1 Skoru
- Karışıklık Matrisi (Confusion Matrix)
- Model Eğitim Süresi

Bu metrikler, her bir modelin doğruluğunu ve başarımını görselleştirmek için kullanılmıştır.


MODELLER HAKKINDA

KNN Accuracy	: 0.84	Training Time : 0.02 seconds
Random Forest Accuracy	: 0.94	Training Time : 7.41 seconds
Decision Tree Accuracy	: 0.87	Training Time : 5.07 seconds
SVM Accuracy	: 0.90	Training Time : 4.66 seconds
XGBoost Accuracy	: 0.95	Training Time : 47.39 seconds
Gradient Boosting Accuracy	: 0.93	Training Time : 615.08 seconds

XGBoost Modeli:

- **Avantajlar:** En yüksek doğruluğa sahip olması (%95) ve çeşitli veri türlerinde iyi performans göstermesi.
- **Dezavantajlar:** Eğitim süresinin uzun olması (47.39 saniye), büyük veri setlerinde bu sürenin daha da artabileceği anlamına gelir.


Random Forest Modeli:

- **Avantajlar:** Yüksek doğruluğa sahip (%94) ve eğitim süresi XGBoost'a göre daha kısa (7.41 saniye).
 - **Dezavantajlar:** Eğitim süresi yine de bazı diğer modellerden daha uzun olabilir.
- 

Gradient Boosting Modeli:

- **Avantajlar:** Yüksek doğruluğa sahip (%93).
- **Dezavantajlar:** En uzun eğitim süresine sahip model (615.08 saniye), bu da modelin büyük veri setlerinde pratik olmamasına yol açabilir.

Support Vector Machine (SVM):

- **Avantajlar:** Yüksek doğruluğa sahip (%90) ve eğitim süresi orta derecede (4.66 saniye).
 - **Dezavantajlar:** Eğitim süresi hala bazı diğer modellerden daha uzun olabilir.
- 

Decision Tree Modeli:

- **Avantajlar:** Orta derecede doğruluğa sahip (%87) ve eğitim süresi orta derecede (5.07 saniye).
- **Dezavantajlar:** Doğruluğu bazı diğer modellerden algoritma yapısından dolayı daha düşük olabilir.

K-Nearest Neighbors (KNN):

- **Avantajlar:** En kısa eğitim süresine sahip model (0.02 saniye).
- **Dezavantajlar:** En düşük doğruluğa sahip model (%84).

TARTIŞMA

Deneylerden elde edilen bulgular:

- Modellerin performansı kıyaslanmış ve en iyi sonuç veren model belirlenmiştir.
- Güçlü Yönler: XGBoost modelinin yüksek doğruluk oranı ve genel performansı.
- Eksiklikler: Bazı kategorilerdeki düşük hassasiyet ve duyarlılık oranları.

Bu bulgular, uydu görüntülerinin sınıflandırılması konusunda gelecekte yapılacak çalışmalar için temel oluşturabilir.

Kullanılan verilerin yeniden boyutlandırıldığında kullanılan piksel sayısı görüntünün modeller içerisinde işlenmesine büyük oranda etki ettiği elde edilmiştir.

Veri Ön İşleme ve Boyutlandırma:

- Görüntü boyutları daha yüksek çözünürlükte tutulabilir. 28x28 yerine 64x64 veya 128x128 gibi boyutlar kullanılabilir.
- Veri arttırma (augmentation) teknikleri kullanılarak eğitim verisi zenginleştirilebilir.

Model İyileştirmeleri:

- Derin öğrenme modelleri (CNN) kullanılarak ve modellerin hiperparametre optimizasyonu yapılarak daha iyi sonuçlar elde edilebilir. Özellikle ResNet, VGG gibi önceden eğitilmiş modeller transfer öğrenme ile kullanılabilir. Grid Search veya Random Search gibi yöntemler kullanılabilir.

Veri Kümesi Büyütme:

- Daha fazla veri toplanarak eğitim veri kümesi genişletilebilir. Bu, modellerin genelleme yeteneğini artıracaktır.
- Farklı veri kaynakları kullanılarak veri çeşitliliği artırılabilir.

Ensemble Yöntemleri:

- Birden fazla modelin bir araya getirilerek ensemble yöntemleri kullanılabilir. Bu, genel model performansını artırabilir ve farklı modellerin güçlü yönlerinden yararlanılabilir.

REFERANSLAR

- Projede kullanılan kütüphaneler ve kaynaklar.
- Kullanılan veri setinin kaynağı.