

WiFi CSI Verileri ile İç Mekan Konum Tahmini: Teknik Rapor

İçindekiler

1. Giriş
2. Veri Seti Analizi
3. Veri Ön İşleme
4. Model Geliştirme
5. Sonuçlar ve Değerlendirme
6. Gelecek Çalışmalar

1. Giriş

Bu proje, WiFi Kanal Durum Bilgisi (CSI - Channel State Information) verilerini kullanarak iç mekanlarda hassas konum tespiti yapmayı amaçlamaktadır. CSI verileri, WiFi sinyallerinin iç mekanlarda nasıl yayıldığını gösteren zengin bir bilgi kaynağıdır.

1.1 Projenin Amacı

- İç mekan konumlandırma için makine öğrenmesi modelleri geliştirmek
- CSI verilerinden anlamlı özellikler çıkarmak
- Farklı model yaklaşımlarını karşılaştırmak
- 2-3 metre hassasiyetle konum tahmini yapmak

1.2 Teknik Altyapı

- Python 3.12 programlama dili
- Scikit-learn, TensorFlow gibi makine öğrenmesi kütüphaneleri
- NumPy, Pandas veri işleme kütüphaneleri
- Matplotlib, Seaborn görselleştirme araçları

2. Veri Seti Analizi

2.1 Veri Seti Yapısı

- 3 anten/alıcı
- Her anten için 30 alt taşıyıcı
- Her ölçüm için 1500 örnek
- Kompleks değerli veriler (gerçek + sanal kısım)

2.2 Sinyal Özellikleri

Genlik (Amplitude) Özellikleri: - Değer Aralığı: 0 - 55.72 - Ortalama: 19.35 - Standart Sapma: 10.18 - Alt taşıyıcılar arasında belirgin örüntüler

Faz (Phase) Özellikleri: - Değer Aralığı: -3.02 - 3.14 radyan ($-\pi$ - π) - Ortalama: 0.054 - Standart Sapma: 0.89 - Faz sarmalama örüntüleri

2.3 Veri Kalitesi

- Eksik veri yok
- Genlik değerleri iyi dağılmış
- Faz değerleri beklenen sarmalama davranışını gösteriyor

- Hem genlik hem faz örüntülerinde net yapı

3. Veri Ön İşleme

3.1 Faz Düzeltme

```
def phase_correction(phase_data):
    """Faz verilerini alt taşıyıcılar arasında düzeltir."""
    return np.unwrap(phase_data, axis=1)
```

3.2 Özellik Çıkarımı

Her anten için çıkarılan özellikler: - Genlik istatistikleri (ortalama, standart sapma, min, max, medyan) - Faz istatistikleri (ortalama, standart sapma, medyan) - Çeyreklik değerleri (Q1, Q3) - Zamansal özellikler (LSTM modeli için)

3.3 Veri Normalizasyonu

- StandardScaler kullanılarak özellikler normalize edildi
- Koordinat verileri ölçeklendirildi
- Aykırı değerler temizlendi

4. Model Geliştirme

4.1 Random Forest Regressor

```
rf_model = RandomForestRegressor(
    n_estimators=1000,
    max_depth=20,
    min_samples_split=5,
    min_samples_leaf=4,
    max_features='sqrt',
    bootstrap=True,
    oob_score=True,
    random_state=42,
    n_jobs=-1
)
```

Özellikler: - En tutarlı performans - Merkezi alanlarda daha iyi doğruluk - Muhafazakar tahminler

4.2 Gradient Boosting

```
gb_model = MultiOutputRegressor(
    GradientBoostingRegressor(
        n_estimators=300,
        learning_rate=0.01,
        max_depth=6,
        min_samples_split=5,
        min_samples_leaf=4,
        subsample=0.8,
        validation_fraction=0.2,
        n_iter_no_change=10,
        random_state=42
    )
)
```

)
)

Özellikler: - Random Forest'a benzer performans - Daha yüksek varyans - Yoğun eğitim verisi olan alanlarda daha iyi

4.3 Sinir Ağı

Mimarisi: - Giriş katmanı - 3 artık (residual) blok - Batch normalizasyon - Dropout (0.3) - Çıkış katmanı (2 nöron, X,Y koordinatları)

Özellikler: - Doğrusal olmayan örüntüleri yakalama - Değişken performans - Aşırı öğrenme eğilimi

4.4 LSTM Modeli

Zamansal özellikleri kullanarak: - Hareket örüntülerini yakalama - Sıralı veri analizi - Yüksek tahmin varyansı

5. Sonuçlar ve Değerlendirme

5.1 Model Performansları

Ortalama metrikler: - RMSE: 2-3 metre - MAE: 1.5-2.5 metre - R^2 Skoru: -0.02 ile -0.12 arası

5.2 Model Karşılaştırması

1. Random Forest:
 - En tutarlı sonuçlar
 - Merkezi alanlarda 2m altı hata
 - Düşük varyans
2. Gradient Boosting:
 - Random Forest'a yakın performans
 - Bazı bölgelerde daha iyi sonuçlar
 - Orta düzey varyans
3. Sinir Ağı:
 - Karmaşık örüntülerde iyi
 - Yüksek varyans
 - Eğitim verisi dağılımına hassas
4. LSTM:
 - Hareket tahmininde başarılı
 - En yüksek varyans
 - Zamansal ilişkileri yakalama

5.3 Hata Analizi

- Duvar/köşe yakınlarında daha yüksek hata
- Merkezi alanlarda daha düşük hata
- Bazı bölgelerde sistematik sapmalar

6. Gelecek Çalışmalar

6.1 Veri Toplama

- Yüksek hatalı bölgelerde ek veri

- Çevresel özelliklerin eklenmesi
- Daha iyi gürültü filtreleme

6.2 Model İyileştirmeleri

- Topluluk yöntemleri
- Belirsizlik tahmini
- Hiperparametre optimizasyonu

6.3 Özellik Mühendisliği

- Gelişmiş faz düzeltme
- Ek zamansal özellikler
- Sinyal gücü göstergeleri

Ekler

Ek-1: Örnek Görselleştirmeler

[Görselleştirmeler model_results/ dizininde bulunmaktadır]

Ek-2: Performans Grafikleri

[Detaylı performans grafikleri docs/figures/ dizininde bulunmaktadır]

Ek-3: Kod Dokümantasyonu

Tüm kod tabanı Türkçe ve İngilizce dokümantasyon içermektedir.