

SİNYAL VE GÖRÜNTÜ İŞLEME PROJE RAPORU

GERÇEK Mİ, SAHTE Mİ? - CHATTERBOX İLE ÜRETİLMİŞ SENTETİK SESLERİN TESPİTİ

Proje Ekibi:

- **Eda TEKEŞ** (eda.t.23@ogr.iu.edu.tr)
- **Selen GÜNEL** (seleng@ogr.iu.edu.tr)
- **Zehra ÖZTÜRK** (zehraozturk2023@ogr.iu.edu.tr)

1. GİRİŞ VE AMAÇ

Metinden sese (TTS) sistemlerindeki hızlı ilerleme, insan sesinden neredeyse ayırt edilemeyen yapay seslerin üretilmesini mümkün kılmıştır. Bu gelişme; kimlik hırsızlığı ve sosyal mühendislik saldırıları dahil olmak üzere önemli güvenlik risklerini de beraberinde getirmektedir.

Bu projenin amacı; gerçek insan sesleri ile **Chatterbox TTS** modeli kullanılarak üretilen deepfake sesleri ayırt edebilecek sağlam bir sistem geliştirmektir. Sistem, yüksek doğrulukta sınıflandırma sağlamak için dijital sinyal işleme tekniklerini ve klasik makine öğrenmesi algoritmalarını kullanmaktadır.

2. VERİ SETİ OLUŞTURMA

2.1 Gerçek Ses Kayıtları

Gerçek ses örnekleri, etik kurallar çerçevesinde gönüllülerden toplanmıştır. Kayıtlar aşağıdaki teknik özelliklere sahiptir:

- **Format:** .wav (16-bit PCM)
- **Kanal:** Mono
- **Örnekleme Hızı:** 16 kHz
- **Süre:** Klip başına 5–15 saniye

2.2 Deepfake Ses Üretimi

Sentetik sesler **Chatterbox TTS** modeli kullanılarak üretilmiştir. Gerçek kayıtlar, modele "ses istemi" (audio prompt) olarak verilmiş ve benzer sentetik karşılıkları oluşturulmuştur. Bu sayede, her gerçek sesin bir deepfake karşılığının olduğu dengeli bir veri seti elde edilmiştir.

3. VERİ ÖN İŞLEME

Tüm ses dosyaları 16 kHz standartına getirilmiş ve mono kanala dönüştürülmüştür. Kayıtların başındaki ve sonundaki sessiz bölümler, gürültüyü azaltmak ve özellik çıkarımı verimliliğini artırmak amacıyla `librosa.effects.trim` kullanılarak temizlenmiştir.

4. ÖZELLİK ÇIKARIMI

Proje hem spektral hem de zaman-frekans alanı özelliklerine odaklanmaktadır:

4.1 Spektral ve Zaman-Frekans Özellikleri

- **MFCC (13 katsayı):** Sesin güç spektrumunu yakalar.
- **Delta ve Delta-Delta MFCC:** MFCC katsayılarının değişim yörüngelerini temsil eder.
- **Sıfır Geçiş Oranı (ZCR):** Sinyaldeki işaret değişim hızını tespit eder.
- **Spektral Centroid / Flatness / Rolloff:** Spektrumun şeklini ve "parlaklığını" tanımlar.
- **RMS Enerjisi:** Ses seviyesini ve enerji düzeylerini temsil eder.

4.2 İstatistiksel Özetleme

Her bir özellik matrisi için, sınıflandırıcıya girdi olacak sabit uzunlukta bir özellik vektörü oluşturmak amacıyla şu istatistikler hesaplanmıştır:

- Ortalama (Mean), Standart Sapma (Std), Maksimum ve Minimum.

5. MODEL GELİŞTİRME

- **Veri Bölme:** Veri seti, sınıf dengesini korumak için **tabakalı bölme** (stratified splitting) yöntemiyle %80 eğitim ve %20 test olarak ayrılmıştır.
- **Ölçekleme:** Özellikleri normalize etmek ve model performansını optimize etmek için StandardScaler uygulanmıştır.
- **Sınıflandırıcı:** Yüksek boyutlu uzaylardaki etkinliği nedeniyle **RBF çekirdekli Destek Vektör Makineleri (SVM)** seçilmiştir.
 - *Parametreler:* C=15.0, Gamma='scale'.

6. PERFORMANS VE DEĞERLENDİRME

Model, gerçek sesleri yanlış alarm üretmeden tanımada yüksek hassasiyet göstermiştir.

- **Genel Doğruluk:** %93.75
- **Karmaşıklık Matrisi Analizi:** 8 gerçek ve 7 sahte ses doğru sınıflandırılmış, sadece bir sahte ses yanlışlıkla gerçek olarak atanmıştır.

7. ETİK DEĞERLENDİRME

Bu proje etik ilkelere bağlı kalınarak yürütülmüştür:

- Tüm katılımcılardan bilgilendirilmiş onam alınmıştır.
- Kişisel bilgileri korumak amacıyla ses verileri anonimleştirilmiştir.
- Sentetik sesler sadece araştırma amaçlı üretilmiş ve kullanılmıştır.

8. LİTERATÜR KARŞILAŞTIRMASI

MFCC özelliklerini SVM sınıflandırıcısı ile birleştirme yaklaşımı, ses adli tıbbında köklü bir yöntemdir. Bu proje; klasik sinyal işlemenin, veri yoğunluklu derin öğrenme modellerine kıyasla, özel deepfake tespit görevleri için kararlı ve yorumlanabilir bir çözüm sunduğunu doğrulamaktadır.

9. SONUÇLAR

Geliştirilen sistem **%93.75 genel doğruluk** oranına ulaşmıştır. Sentetik sesler için kesinlik (precision) 1.00 ve F1-skoru 0.93 olarak hesaplanmıştır; bu da modelin Chatterbox mimarisi tarafından üretilen sesleri tespit etmede son derece güvenilir olduğunu göstermektedir.

10. GELECEK ÇALIŞMALAR

Gelecekteki olası geliştirmeler şunları içerebilir:

- Veri setinin farklı diller ve konuşmacılarla genişletilmesi.
- Perde (pitch) ve formant frekansları gibi prozodik özelliklerin entegre edilmesi.
- Gerçek zamanlı bir tespit uygulaması geliştirilmesi.

11. KAYNAKÇA

- Kinnunen, T., & Li, H. (2010). *An overview of text-independent speaker recognition: From features to supervectors*. Speech Communication.
- Theodoridis, S., & Koutroumbas, K. (2009). *Pattern Recognition*. Academic Press.
- McFee, B., et al. (2015). *librosa: Audio and Music Signal Analysis in Python*. SciPy Conference.
- Pedregosa, F., et al. (2011). *Scikit-learn: Machine Learning in Python*. Journal of Machine Learning Research.
- Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*. Springer.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep Learning*. MIT Press.
- Wu, Z., et al. (2015). *Spoofing and countermeasures for speaker verification: A survey*. Speech Communication.
- Chatterbox TTS GitHub Repository: <https://github.com/resemble-ai/chatterbox>