

Mel-Spektrogram ve CNN Kullanarak Müzik Türü Sınıflandırma

Ders: Derin Öğrenme

Proje Türü: Dönem Projesi

Öğrenci: Tunahan Sarı

Öğrenci No: 23120205084

1. Giriş

Bu projede müzik türü sınıflandırma problemi derin öğrenme yöntemleri kullanılarak ele alınmıştır. Amaç, bir ses dosyasını giriş olarak alıp ait olduğu müzik türünü otomatik olarak tahmin eden bir sistem geliştirmektir.

2. Veri Seti

Bu çalışmada GTZAN formatına uygun bir müzik veri seti kullanılmıştır. Veri seti 10 farklı müzik türünden oluşmaktadır: blues, classical, country, disco, hiphop, jazz, metal, pop, reggae ve rock. Veri seti boyutu nedeniyle GitHub reposuna eklenmemiştir.

3. Özellik Çıkarımı

Ham ses sinyalleri doğrudan modele verilmemiştir. Bunun yerine sesler mel-spektrogram temsiline dönüştürülmüştür. Bu temsil zaman ve frekans bilgisini içerir ve CNN modelleri için uygun bir iki boyutlu girdi sağlar.

4. Model Mimarisi

Sınıflandırma işlemi için Yapay Sinir Ağı (CNN) mimarisi kullanılmıştır. Model; konvolüsyon katmanları, aktivasyon fonksiyonları, havuzlama katmanları ve tam bağlı katmanlardan oluşmaktadır.

5. Yöntem Seçimi ve Karşılaştırmalı Analiz

Bu problem literatürde farklı yaklaşımlarla ele alınmaktadır. Klasik yöntemlerde MFCC / spectral features çıkarılıp SVM, Random Forest veya k-NN gibi modellerle sınıflandırma yapılır. Bu yaklaşımlar hızlıdır ancak özellik mühendisliğine bağımlıdır ve karmaşık örüntülerde sınırlı kalabilir.

Derin öğrenme tarafında iki yaygın yaklaşım vardır: (i) Ham dalga formunu 1D-CNN/TCN ile işlemek, (ii) Sesi zaman-frekans düzleminde temsil edip (mel-spectrogram) 2D-CNN ile sınıflandırmak. Ham dalga formu yöntemleri daha az ön-işleme gerektirir fakat daha fazla veri ve daha uzun eğitim süresi isteyebilir.

Bu projede mel-spectrogram + 2D-CNN seçilmiştir; çünkü mel-spectrogram, sesin frekans içeriğini görselleştirerek CNN'lerin güçlü olduğu "görüntü

benzeri” bir girdiye dönüştürür. Böylece model, manuel özellik tasarımına daha az ihtiyaç duyar ve GTZAN-style veri düzeninde pratik bir çözüm sunar. Ayrıca Gradio arayüzü ile gerçek kullanım senaryosuna uygun şekilde kolayca deploy edilebilir.

6. Eğitim Süreci

Model eğitimi sırasında Çapraz Entropi (Cross-Entropy) kayıp fonksiyonu ve Adam optimizasyon algoritması kullanılmıştır. Eğitim sürecinde en iyi performansı veren model ağırlıkları kaydedilmiştir.

7. Değerlendirme

Model performansı doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1-skoru metrikleri kullanılarak değerlendirilmiştir. Elde edilen sonuçlar modelin müzik türlerini ayırt etmede yeterli bir performans sergilediğini göstermektedir.

8. Dağıtım (Deployment)

Eğitilen model Gradio kullanılarak web tabanlı bir arayüz üzerinden kullanıma sunulmuştur. Kullanıcılar ses dosyası yükleyerek müzik türü tahmini alabilmektedir.

9. Sonuç:

Bu projede mel-spektrogram tabanlı CNN modeli ile müzik türü sınıflandırma problemi başarıyla çözülmüştür. Gelecekte daha büyük veri setleri ve farklı mimariler kullanılarak performans artırılabilir.

Kaynakça:

DATASET: [Link](#)