

NHẬN DIỆN CẢM XÚC QUA GIỌNG NÓI SỬ DỤNG KOLMOGOROV-ARNOLD NETWORKS

Dương Thanh Nguyên

Trường Đại học Công nghệ thông tin, Thành phố Hồ Chí Minh, Việt Nam

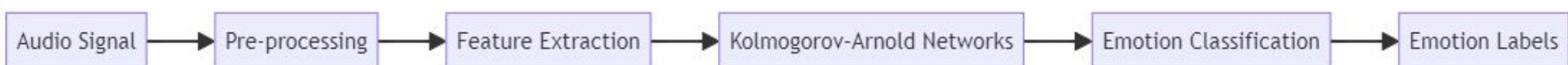
What ?

- Đề xuất sử dụng Kolmogorov-Arnold Networks (KANs) để cải thiện độ chính xác và hiệu suất, so sánh với các mô hình truyền thống.
- Đánh giá hiệu suất KANs trong môi trường thực tế, khẳng định tính ưu việt của phương pháp này.
- Xây dựng mô hình tối ưu cho nhận diện cảm xúc qua giọng nói dựa trên KANs.

Why ?

- Cần cải thiện tương tác giữa máy tính và con người để tạo trải nghiệm người dùng tốt hơn.
- Nhu cầu về hỗ trợ tâm lý và chăm sóc khách hàng đang tăng cao.
- Công nghệ nhận diện cảm xúc chưa đạt hiệu quả cao với phương pháp học máy truyền thống.

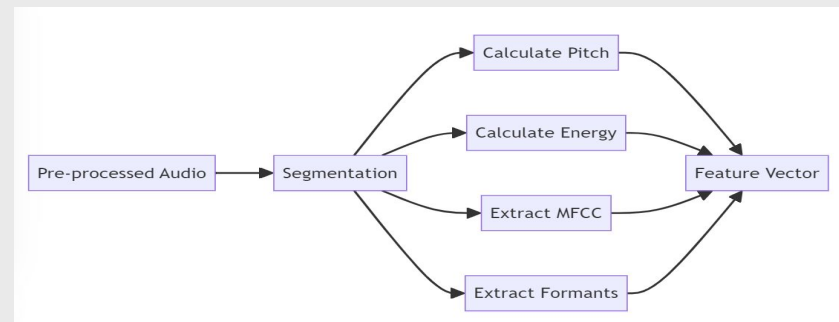
Overview



Pre-processed



Feature Extraction



Description

1. Mục tiêu

- Cải thiện độ chính xác trong nhận diện cảm xúc qua giọng nói sử dụng KANs
- Khám phá hiệu quả của KANs trong ứng dụng thực tế
- Xây dựng mô hình tối ưu cho nhận diện cảm xúc qua giọng nói từ KANs

2. Nội dung

- Nghiên cứu và phân tích đặc trưng giọng nói để phân biệt cảm xúc.
- Xây dựng và so sánh hiệu suất của mô hình KANs với các mô hình truyền thống (LSTM, CNN, RNN).
- Đánh giá khả năng và hiệu quả của KANs trong các tình huống thực tế và môi trường âm thanh đa dạng.
- Tối ưu hóa kiến trúc và các tham số của mô hình KANs để đạt được độ chính xác tốt nhất trong nhận diện cảm xúc qua giọng nói.

3. Phương Pháp

- Sử dụng các tập dữ liệu giọng nói đã được gán nhãn cảm xúc như RAVDESS, IEMOCAP.
- Tiền xử lý dữ liệu bao gồm loại bỏ nhiễu và trích xuất các đặc trưng như MFCCs, formants, pitch.
- Xây dựng mô hình sử dụng KANs và huấn luyện trên dữ liệu đã tiền xử lý.
- Đánh giá hiệu suất bằng các chỉ số như accuracy, precision, recall, F1-score và so sánh với các mô hình truyền thống (LSTM, CNN, RNN).
- Xác định các tình huống ứng dụng thực tế và thực hiện thử nghiệm trên các dữ liệu giọng nói từ các môi trường âm thanh khác nhau.
- Tối ưu hóa kiến trúc và tham số của mô hình sử dụng KANs bằng cách thử nghiệm và điều chỉnh các yếu tố như số lớp, số nơ-ron, hàm kích hoạt.

4. Kết quả mong đợi

- KANs mang lại độ chính xác cao hơn và hiệu suất tốt hơn so với các phương pháp truyền thống (LSTM, CNN, RNN) trong nhận diện cảm xúc qua giọng nói.
- KANs có khả năng ứng dụng hiệu quả trong các tình huống thực tế, duy trì hiệu suất tốt trong các môi trường âm thanh phức tạp.
- Xây dựng được mô hình KANs tối ưu với hiệu suất tính toán cao và tốc độ huấn luyện nhanh, dễ dàng triển khai và ứng dụng.