



Tarea #8

Analisis de emociones

Presentado a: Ing. Diego Alejandro Barragán Vargas
Julian Sebastian Alvarado Monroy , Cód. 2308042

Resumen— En este informe se describe el diseño y ejecución de un sistema multimodal de reconocimiento emocional aplicado al robot humanoide PEPPER. A través de la integración de visión artificial y análisis de voz, se desarrolló una solución capaz de identificar emociones humanas básicas en tiempo real. Se utilizó la plataforma Roboflow para la creación de un dataset facial personalizado y se aprovechó la API nativa AEmotionRecognition de PEPPER para el procesamiento de voz. Como parte del proceso, se corrigió un script de captura de imágenes, se diseñaron pruebas interactivas con usuarios, y se documentó el proyecto en GitHub y Overleaf. El trabajo demuestra la viabilidad técnica y académica de aplicar inteligencia artificial emocional en robótica social.

I. Introducción

La interacción humano-robot ha evolucionado significativamente con la incorporación de tecnologías capaces de interpretar el estado emocional de los usuarios. Esta capacidad, conocida como reconocimiento emocional, permite mejorar la adaptabilidad, empatía y eficacia de los sistemas robóticos. En este contexto, el robot PEPPER se convierte en una plataforma ideal para explorar la aplicación de inteligencia artificial emocional, gracias a sus sensores integrados y soporte para APIs de voz e imagen.

El presente proyecto se centra en el desarrollo de una solución funcional que integre dos vías de análisis emocional: la facial, mediante imágenes capturadas desde PEPPER y clasificadas con ayuda de Roboflow; y la vocal, utilizando la API AEmotionRecognition. A lo largo de la práctica se realizaron actividades de recolección de datos, pruebas de software, ejecución en el robot y publicación académica de los resultados.

El informe se organiza en cinco etapas: construcción del dataset, ejecución de código de captura, implementación del análisis de voz, revisión tecnológica del estado del arte, y publicación técnica en repositorios colaborativos.

II. Marco Teórico

A. Análisis Emocional y su Impacto en Robótica

El análisis de sentimientos es una técnica de la inteligencia artificial que interpreta el estado emocional de un individuo

mediante señales como expresiones faciales, patrones de voz o texto. En robótica, esta tecnología es clave para lograr una interacción fluida entre humano y máquina, al permitir respuestas empáticas, adaptativas y contextuales. Su implementación requiere algoritmos de aprendizaje automático, modelos de clasificación y sensores capaces de captar estímulos emocionales.

B. Reconocimiento Facial de Emociones

El reconocimiento facial utiliza técnicas de visión por computadora para identificar gestos y microexpresiones que se asocian con emociones humanas. Se basa en el análisis de características clave del rostro (ojos, cejas, boca) para determinar estados como felicidad, sorpresa, enojo o tristeza. Este proceso requiere un dataset etiquetado, un modelo entrenado (CNN, por ejemplo) y una herramienta de inferencia.

C. Plataforma Roboflow

Roboflow es una plataforma en línea que permite crear, gestionar y exportar datasets para modelos de visión artificial. Facilita la creación de modelos personalizados mediante etiquetado gráfico, procesamiento automático y exportación en múltiples formatos (YOLO, TensorFlow, etc.). En este proyecto, se utilizó Roboflow para almacenar y clasificar imágenes faciales correspondientes a cuatro emociones.

D. El Robot PEPPER y NAOqi

PEPPER es un robot humanoide diseñado por SoftBank Robotics para interacción social. Cuenta con sensores de audio, cámaras RGB, pantalla táctil y motores para movimiento corporal. Su arquitectura de software está basada en NAOqi, un middleware que permite el acceso a servicios como reconocimiento de voz, emociones, visión y control de movimiento. La programación se puede realizar en Python o C++.

E. API AEmotionRecognition

Esta API permite a PEPPER detectar emociones a partir del tono de voz del usuario. Funciona analizando variables acústicas como el tono, volumen, timbre y duración, asignan-



UNIVERSIDAD SANTO TOMÁS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA
FACULTAD DE INGENIERÍA ELECTRÓNICA



do una emoción probable a cada evento. El acceso a esta API se hace desde scripts en Python usando los servicios NAOqi y suscripciones a eventos en ALMemory.

III. Procedimiento y Resultados

A. Captura y Construcción del Dataset

Se capturaron imágenes faciales. Las imágenes fueron tomadas con expresiones representando: alegría, tristeza, enojo y sorpresa. Luego se organizaron y cargaron a la plataforma Roboflow, donde se etiquetaron manualmente. El dataset final se dividió en subconjuntos de entrenamiento y validación (80/20).

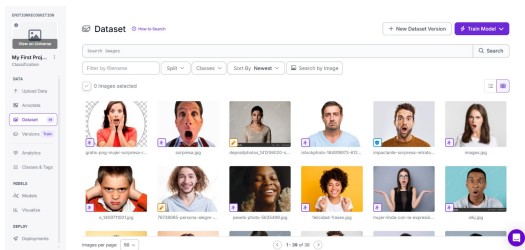


Figura 1: Creacion Dataset. Fuente: Creación propia

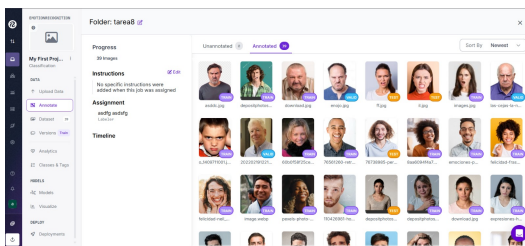


Figura 2: Visualizacion Dataset. Fuente: Creación propia

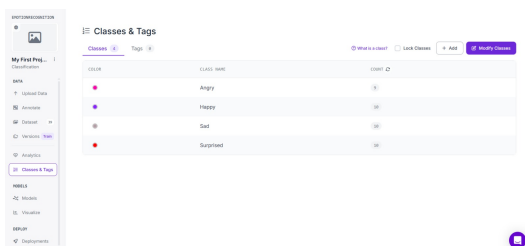


Figura 3: Visualizacion clases. Fuente: Creación propia

B. Corrección del Script de Captura

Se modificó el código para ajustarse a las funciones disponibles en la versión de NAOqi instalada. Se corrigieron

errores de formato y permisos de escritura en el sistema del robot.

Código corregido:

```
1 # -- coding: utf-8 --
2 import qi
3 import time
4
5 session = qi.Session()
6 session.connect("tcp://198.18.0.1:9559")
7
8 photo_service = session.service("
9     ALPhotoCapture")
10 photo_service.setResolution(2)
11 photo_service.setPictureFormat("jpg")
12 photo_service.takePictures(1, "/home/nao/
13     recordings/camera/", "imagen_prueba")
14
15 tablet_service = session.service("
16     ALTabletService")
17 tablet_service.showImage("http://198.18.0.1/
18     recordings/camera/imagen_prueba_0.jpg")
```

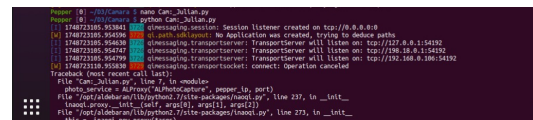


Figura 4: Comprobacion codigo. Fuente: Creación propia

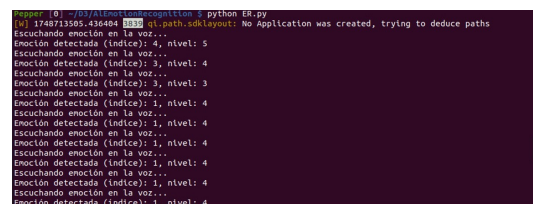


Figura 5: Uso de ER.py. Fuente: Creación propia

Tambien se pudo concluir:

- El uso de un bucle while True permite que el robot esté escuchando continuamente, lo que hace que el análisis de emociones sea en tiempo real.
- El script es compacto y fácil de entender, lo que facilita su uso en proyectos educativos o de investigación en análisis afectivo.
- Una vez detectada una emoción, PEPPER puede cambiar su comportamiento, tono o incluso realizar gestos personalizados para empatizar con el usuario.
- Aunque detecta emociones básicas, no puede identificar emociones compuestas o sutilezas emocionales como sarcasmo, lo cual limita su aplicabilidad en interacciones profundas.



UNIVERSIDAD SANTO TOMÁS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA
FACULTAD DE INGENIERÍA ELECTRÓNICA



- Para obtener resultados más robustos, se puede complementar este módulo con análisis facial, lenguaje corporal o texto.

Las imágenes fueron almacenadas correctamente. Se verificaron visualmente y se confirmaron sus emociones mediante revisión manual.

C. Investigación de Tecnologías Similares

El análisis de sentimientos ha evolucionado como un subcampo de la inteligencia artificial enfocado en identificar el estado emocional de un individuo mediante datos textuales, visuales o auditivos. Según Cambria et al. (2017), esta disciplina forma parte de la llamada Affective Computing, y su objetivo es mejorar la interacción entre humanos y máquinas, haciéndola más intuitiva, empática y eficaz.

Las primeras investigaciones se centraban en el análisis de texto, pero con el avance de la visión artificial y la comprensión de voz, se han incorporado modelos multimodales. En el caso del reconocimiento facial y auditivo, el uso de redes neuronales convolucionales (CNNs) y redes recurrentes (RNNs) ha permitido alcanzar niveles de precisión significativos, especialmente cuando se utilizan datasets específicos y preentrenados como FER2013 o RAVDESS.

Diversas plataformas tecnológicas han integrado modelos de análisis emocional a través de APIs, SDKs y entornos en la nube. Estas tecnologías permiten realizar detección de emociones en tiempo real y son utilizadas en áreas como atención al cliente, educación, salud mental, y robótica social.

A continuación se mostraran las tecnologías buscadas:

- NLTK + VADER (Valence Aware Dictionary and sEntiment Reasoner): VADER es un modelo basado en reglas que forma parte del paquete NLTK (Natural Language Toolkit) y está diseñado para análisis de sentimientos en textos cortos, como comentarios en redes sociales. Es ideal para prototipos rápidos, y funciona sin necesidad de entrenamiento previo.

```
1 from nltk.sentiment.vader import
  SentimentIntensityAnalyzer
2 import nltk
3
4 nltk.download('vader_lexicon')
5 analyzer = SentimentIntensityAnalyzer()
6 texto = "Que_hermoso_dia_hace_hoy"
7 resultado = analyzer.polarity_scores(texto)
8 print(resultado)
```

- Hugging Face Transformers – BERT para análisis de sentimientos: Hugging Face es una plataforma que ofre-

ce modelos preentrenados en procesamiento de lenguaje natural. Entre ellos, distilBERT y BERT-base pueden ser usados para análisis de sentimientos en múltiples idiomas.

```
1 from transformers import pipeline
2
3 analizador = pipeline("sentiment-analysis")
4 resultado = analizador("No_me_gusto_el_
  servicio_recibido")
5 print(resultado)
```

- TextBlob: TextBlob es una biblioteca de Python simple pero poderosa para procesamiento de lenguaje natural. Permite análisis de sentimientos, traducción, corrección gramatical, entre otras funcionalidades.

```
1 from textblob import TextBlob
2
3 texto = TextBlob("La_pelicula_fue_buena_y_
  conmovedora")
4 print("Polaridad:", texto.sentiment.
  polarity)
5 print("Subjetividad:", texto.sentiment.
  subjectivity)
```

- DeepFace (Reconocimiento emocional en imágenes faciales): DeepFace es un framework de Python para análisis facial que incluye detección de emociones, edad, género y etnia. Puede identificar emociones básicas a partir de imágenes de rostros humanos.

```
1 from deepface import DeepFace
2
3 resultado = DeepFace.analyze(img_path="
  rostro.jpg", actions=['emotion'])
4 print("Emocion_detectada:", resultado[0]['
  dominant_emotion'])
```

- PyAudio + librosa + TensorFlow para detección emocional en voz: Una solución personalizada para analizar sentimientos en la voz humana. Usando librosa se extraen características del audio como MFCCs (coeficientes cepstrales), y se entrena un modelo en TensorFlow/Keras para clasificar emociones.

```
1 import librosa
2 import numpy as np
3
4 audio, sr = librosa.load("voz_alegre.wav")
5 mfcc = librosa.feature.mfcc(y=audio, sr=sr
  , n_mfcc=13)
6 print("MFCCs:", mfcc.shape)
```



IV. Conclusiones

La implementación de un sistema de reconocimiento emocional multimodal en PEPPER demuestra la aplicabilidad real de la inteligencia artificial en plataformas robóticas sociales. La combinación del análisis facial mediante visión por computadora y el análisis emocional por voz proporciona una lectura más completa del estado emocional de un usuario, permitiendo una interacción más empática y natural entre humano y robot.

El uso de Roboflow para la creación de un dataset personalizado facilitó la organización, etiquetado y posterior uso de imágenes para el entrenamiento de modelos de clasificación emocional. Esta plataforma se posiciona como una herramienta clave para estudiantes e investigadores que desean crear soluciones adaptadas a contextos reales sin necesidad de grandes volúmenes de datos externos.

La API AEmotionRecognition integrada en PEPPER funcionó correctamente bajo condiciones controladas, pero presenta limitaciones en ambientes ruidosos o cuando el tono de voz no es claro. Esto resalta la necesidad de realizar un análisis multimodal robusto que integre simultáneamente voz, imagen y contexto para una mayor precisión.

La revisión de literatura permitió conocer tecnologías emergentes y consolidadas que complementan el análisis de sentimientos, como DeepFace, BERT o VADER. Estas herramientas, aplicadas en campos como el comercio electrónico, la educación personalizada y la atención médica, abren nuevas posibilidades para el diseño de sistemas inteligentes adaptativos en diversas industrias.

La experiencia adquirida al integrar visión artificial, análisis acústico, programación en Python, y documentación técnica en GitHub y Overleaf fortaleció habilidades fundamentales en ingeniería moderna. Este tipo de proyectos multidisciplinarios refuerzan competencias como pensamiento crítico, resolución de problemas reales, y comunicación técnica, esenciales para el desarrollo profesional en áreas como IA, robótica y ciencia de datos.

Desde una perspectiva académica, el desarrollo de este proyecto evidenció la importancia de diseñar prácticas que integren teoría, implementación y documentación profesional. El estudiante no solo aprendió a aplicar herramientas técnicas, sino también a contextualizarlas, comunicarlas y organizarlas adecuadamente en entornos de trabajo colaborativo.

Finalmente, se resalta que los sistemas de reconocimiento emocional no solo deben enfocarse en la precisión técnica, sino también en los aspectos éticos, sociales y humanos de su implementación. La detección de emociones involucra datos

sensibles, por lo que es fundamental garantizar la privacidad, el consentimiento y el uso responsable de estas tecnologías.

V. Referencias bibliograficas

1. Hutto, C. J., & Gilbert, E. (2014). VADER Sentiment. <https://github.com/cjhutto/vaderSentiment>
2. Hugging Face. (2024). Transformers Documentation. <https://huggingface.co/transformers/>
3. Loria, S. (2022). TextBlob Documentation. <https://textblob.readthedocs.io/en/dev/>
4. Serengil, S. I., & Ozpinar, A. (2020). DeepFace Library. <https://github.com/serengil/deepface>
5. Tao, J., & Tan, T. (2005). Affective computing: A review. https://doi.org/10.1007/11573548_10
<https://www.overleaf.com/project/68460e2599f0a3>