



Universidad Francisco de Vitoria

GRADO EN INGENIERÍA MATEMÁTICA

Clasificador de expresiones auditivas en entornos sociales y de cuidado

EQUIPO:

Miguel Egido Morales, Jonás Manuel García Vallejo, Alejandra Llord Muñoz de la Espada, Alfredo Robledano Abasolo

Abril 2024

Índice

1. Resumen	1
2. Introducción	1
2.1. Motivación	1
2.2. Objetivos	2
2.2.1. Objetivo general	2
2.2.2. Objetivos específicos	2
2.3. Alcance e impacto	3
3. Estado de la cuestión	3
3.1. Investigaciones existentes	3
3.2. Tabla resumen	6
4. Resolución	6
4.1. Organización	6
4.2. Desarrollo	6
5. Conclusión	6

1. Resumen

Breve descripción de lo que se trata en el documento

2. Introducción

El reconocimiento y clasificación de sonidos en entornos sociales y de cuidado han emergido como áreas de investigación fundamentales con aplicaciones prácticas significativas. En este contexto, el presente proyecto se centra en la identificación precisa de sonidos específicos como tos, risas, lloros y suspiros. Estos sonidos, expresiones auditivas fundamentales en la comunicación humana, ofrecen un riquísimo campo de estudio con posibilidades de aplicación que van desde la salud hasta el bienestar emocional y la automatización del hogar.

2.1. Motivación

La motivación en este proyecto es impulsada por la necesidad de mejorar la calidad de vida y el cuidado de las personas a través de la tecnología. La identificación precisa de sonidos asociados con expresiones emocionales y estados de salud específicos proporciona una herramienta valiosa en entornos médicos, hogares de ancianos, y situaciones sociales diversas. Este proyecto aspira a crear un sistema robusto que no solo ofrezca un análisis acertado de los sonidos, sino que también pueda adaptarse a contextos cambiantes, mejorando así la eficiencia y la efectividad en la toma de decisiones.

2.2. Objetivos

2.2.1. Objetivo general

Identificar y extraer diferentes tipos de sonidos a partir de fragmentos de audio para su posterior análisis e implementación en diferentes modelos sociales. Para ello aplicaremos la Transformada de Fourier para limpiar el sonido de nuestros fragmentos de audio. Posteriormente implementaremos un modelo de clasificación, el cual entrenaremos y haremos predicciones con nuevos datos.

2.2.2. Objetivos específicos

1. **Explorar y adquirir datos de audio:** Recopilar una amplia variedad de muestras de audio que representen diferentes tipos de sonidos relevantes para el análisis y modelado social.
2. **Preprocesamiento de datos de audio:** Utilizar herramientas de procesamiento de señales para preprocesar los datos de audio, realizando la normalización del volumen y la eliminación de ruido de fondo no deseados. Este último lo realizaremos mediante la aplicación de la transformada de fourier para representar la señal en el dominio de la frecuencia, analizar este espectro de frecuencia para identificar las componentes de sonido deseadas y el ruido no deseado y por último realizar una eliminación selectiva de frecuencias no relevantes para mejorar la calidad de la señal y reducir el impacto del ruido.
3. **Implementar la Transformada de Fourier:** Utilizaremos la biblioteca *numpy* de Python, que permitirá transformar los datos de audio del dominio temporal al dominio de la frecuencia.
4. **Análisis de frecuencia de sonido:** Analizar los espectros de frecuencia resultantes para identificar las características espectrales distintivas de diferentes tipos de sonidos, como tonos, frecuencias dominantes y distribuciones de energía en diferentes bandas de frecuencia.
5. **Extracción de características relevantes:** Identificar y extraer características relevantes de los espectros de frecuencia que puedan ser útiles para la clasificación y análisis de diferentes tipos de sonidos, como la amplitud máxima, la frecuencia dominante o la dispersión espectral.
6. **Desarrollo de modelos de clasificación:** Utilizaremos técnicas de aprendizaje automático, como redes neuronales, para desarrollar un modelo que pueda clasificar automáticamente los diferentes tipos de sonidos basados en las características extraídas.
7. **Validación y evaluación del modelo:** Evaluaremos el rendimiento de los modelos de clasificación con nuevos datos para evaluar el rendimiento de nuestro modelo. Utilizando métricas como la precisión.
8. **Implementación y despliegue del modelo:** Integraremos el modelo de cla-

sificación final en una aplicación para su implementación práctica (de manera únicamente teórica), lo que llegará a poder realizar una detección y clasificación a tiempo real en entornos sociales específicos.

2.3. Alcance e impacto

3. Estado de la cuestión

3.1. Investigaciones existentes

Artículo 1

Título: “Desarrollo y análisis de clasificadores de señales de audio”

ID: [1]

URL: <https://riunet.upv.es/handle/10251/90005>

Resumen:

En este artículo podemos encontrar un análisis sobre la clasificación y reconocimiento de señales de audio, donde podemos destacar su relevancia en aplicaciones modernas que combinan procesamiento de audio y aprendizaje automático. En este artículo se mencionan temas bien establecidos como el reconocimiento de voz y la música, así como un campo emergente, la clasificación automática de ruidos ambientales. Para el aprendizaje automático emplea un modelo supervisado, donde los datos de audio están etiquetados en clases. Se describe un proceso de clasificación, que abarca desde la extracción de características del audio hasta la aplicación de algoritmos de aprendizaje automático para esta tarea. También, se proporciona un sistema de clasificación y etiquetado de audio en tiempo real como ejemplo concreto de aplicación práctica, incluyendo una estructura de clasificación jerárquica.

Por qué es útil:

Esta información puede ser muy útil para orientar nuestro proyecto en el desarrollo de un modelo que diferencie audios de tos, estornudos y risas. También es importante para nuestro proyecto la extracción de características del audio y el uso de algoritmos de aprendizaje automático para la clasificación. Además, se menciona un enfoque supervisado, que es importante cuando tienes datos etiquetados en clases. Por último, el artículo menciona el desarrollo de un software para la clasificación de audio en tiempo real, lo cual puede ser útil para implementar el modelo en un entorno práctico.

Artículo 2

Título: “Environmental sound recognition: a survey”

ID: [3]

URL: <https://www.cambridge.org/core/journals/apsipa-transactions-on-signal-and-information-processing/article/environmental-sound-recognition-a-survey/96211C365DC7B250CEEFFB15026A5CDF>

Resumen:

El artículo ofrece una revisión detallada y comprensible de los avances más recientes en el campo del reconocimiento de sonidos ambientales, organizada en cuatro secciones principales. Estas incluyen una explicación de los enfoques básicos de procesamiento de sonidos ambientales, técnicas para manejar sonidos ambientales estacionarios y no estacionarios, así como una comparación del rendimiento entre diferentes métodos. Por último, el artículo discute las conclusiones obtenidas y señala las tendencias futuras en la investigación y desarrollo de este campo específico.

Por qué es útil:

Este artículo es útil porque proporciona revisión detallada de los avances de campo del reconocimiento de sonidos ambientales. Al revisar las técnicas presentadas en el artículo, podrás identificar técnicas relevantes que podrían ser aplicables a tu proyecto de diferenciar entre audios de tos y risa. Por ejemplo, podrías encontrar métodos para manejar aspectos no estacionarios de los sonidos, lo cual puede ser relevante para identificar características distintivas de la tos y la risa en los audios.

Artículo 3

Título: “Clasificación automática de sonidos utilizando lenguaje máquina”

ID: [4]

URL: <https://idus.us.es/handle/11441/101412>

Resumen:

Este artículo realiza la implementación de dos modelos para clasificar sonidos ambientales mediante aprendizaje automático. Ambos resaltan la importancia del procesamiento del espectrograma de audio y el preprocesamiento de datos. La finalidad de este artículo es crear un sistema para clasificar sonidos ambientales con precisión aceptable, explorando métodos de extracción de características y utilizando inteligencia artificial.

Por qué es útil:

Este trabajo es útil ya que aborda un tema muy conectado al nuestro, nuestro objetivo es la clasificación de sonidos producidos por las personas. Aunque este artículo clasifique sonidos ambientales nos puede ayudar a relacionar el enfoque que podemos darle a nuestro trabajo. Además utiliza la transformada de fourier, luego también es útil para empezar a conocerla para poder trabajar con ella.

Artículo 4

Título: “Reconocimiento automático de instrumentos mediante aprendizaje máquina”

ID: [5]

URL: <https://idus.us.es/handle/11441/91372>

Resumen:

Este proyecto aborda el reconocimiento de instrumentos musicales mediante múltiples algoritmos de aprendizaje automático, incluyendo KNN, redes neuronales, PCA, LDA y Random Forest. Se evalúan distintos enfoques y se concluye que, en algunos casos, los algoritmos más simples, como KNN, ofrecen resultados óptimos. Se destaca la importancia de la reducción de dimensiones y se exploran las dificultades asociadas

con instrumentos específicos.

Por qué es útil:

La utilidad de la información del estudio de instrumentos se refleja en la experiencia adquirida en el procesamiento de datos de audio y en la selección de algoritmos eficaces para la clasificación, realizando un contraste de estos para elegir el mejor candidato. Además, conceptos como la reducción de dimensiones y la adaptabilidad a contextos cambiantes pueden aplicarse al reconocimiento de sonidos en nuestro proyecto.

Artículo 5

Título: “Classification of Nonverbal Human Produced Audio Events: A Pilot Study”

ID: [2]

URL: <https://espace2.etsmtl.ca/id/eprint/17579/>

Resumen:

En este estudio piloto, se plantea la necesidad de avanzar en el estado del arte sobre las capacidades de procesamiento en micro-controladores de sonidos no verbales como la tos y carraspeo. La correcta clasificación de este tipo de sonidos no verbales captados a través de un pequeño dispositivo intra-aural podría mejorar la atención médica. Las técnicas propuestas que se evalúan en este estudio piloto son Gaussian Mixture Model (GMM), Support Vector Machine y Multi-Layer Perceptron.

Por qué es útil:

El estudio trata técnicas que se adecúan enormemente al tipo de problema que se quiere resolver. Entender su funcionamiento y complejidad permite aumentar el abanico de posibilidades en la tarea de desarrollar un modelo clasificador de sonidos no verbales.

Artículo 6

Título: “A Bag-of-Audio-Words Approach for Snore Sounds’ Excitation Localisation”

ID: [6]

URL: <https://ieeexplore.ieee.org/abstract/document/7776181/>

Resumen:

En esta publicación se estudia como mejorar la calidad del sueño y tratar patologías graves en personas que padecen de ronquidos y de apnea del sueño, a partir del análisis de sonidos. Se trata de un campo de estudio eminentemente práctico que destaca por la importancia y los riesgos en la salud que generan este tipo de enfermedades. Se utilizan técnicas como Bag-of-audio-words para mejorar la eficiencia en clasificadores de sonidos.

Por qué es útil:

El enfoque que se da al análisis de sonidos es muy práctico y aunque tal vez muy específico, permite hacerse a la idea de como puede ser una aplicación práctica de un clasificador de sonidos, además de mejorar la precisión con la técnica Bag-of-audio-words.

Artículo 7

Artículo 8

3.2. Tabla resumen

4. Resolución

4.1. Organización

Organización temporal y de trabajo

4.2. Desarrollo

Desarrollo y resolución completa

5. Conclusión

Referencias

- [1] Fabián Aguirre Martín et al. «Desarrollo y análisis de clasificadores de señales de audio». En: (2017).
- [2] Rachel E Bouserhal et al. «Classification of nonverbal human produced audio events: a pilot study». En: (2018).
- [3] Sachin Chachada y C-C Jay Kuo. «Environmental sound recognition: A survey». En: *APSIPA Transactions on Signal and Information Processing* 3 (2014), e14.
- [4] Patricio Rodríguez Ramírez. «Clasificación automática de sonidos utilizando aprendizaje máquina». En: (2020).
- [5] Aurora Salgado Díaz del Río. «Reconocimiento automático de instrumentos mediante aprendizaje máquina». En: (2019).
- [6] Maximilian Schmitt et al. «A bag-of-audio-words approach for snore sounds' excitation localisation». En: *Speech Communication; 12. ITG Symposium*. VDE. 2016, págs. 1-5.

Artículo	Título	Análisis	Objetivo del estudio	Relación y aportación
3	Clasificación automática de sonidos utilizando lenguaje máquina	Crear un sistema de clasificación de sonidos ambientales. Una vez implantado el modelo estudiar los parámetros de decisión.	Explorar métodos de extracción de características de audio. Conseguir una precisión aceptable. Elegir el mejor modelo de inteligencia artificial.	Clasificador de sonidos y Transformada de Fourier
4	Reconocimiento automático de instrumentos mediante aprendizaje máquina	Desarrollar métodos para el reconocimiento de instrumentos musicales utilizando algoritmos de aprendizaje automático	Identificar qué algoritmo es más óptimo para la clasificación. Se exploran aspectos como la reducción de dimensiones y las dificultades asociadas a cada instrumento	Contraste de algoritmos de clasificación y analogía de los instrumentos musicales con sonidos producidos por las personas.
5	Classification of Nonverbal Human Produced Audio Events: A Pilot Study	Evaluación de la técnicas GMM, SVM y Perceptrón Multicapa en el análisis de sonidos no verbales.	En 10 categorías de sonidos no verbales, medir la eficiencia de 3 técnicas comunmente utilizadas en el procesamiento de sonidos.	Guía útil en la elección del algoritmo utilizado en el desarrollo de un clasificador de sonidos no verbales.
6	A Bag-of-Audio-Words Approach for Snore Sounds' Excitation Localisation	Análisis de sonidos relacionados con el sueño (ronquidos, apnea del sueño, ...).	Realizar un enfoque en la clasificación basado en la técnica Bag-of-Audio-Words.	Entender un caso práctico de uso y el enfoque mediante la técnica Bag-of-Audio-Words.