

## **Reporte: Speech Emotion Recognition Two Decades in a Nutshell, Benchmarks, and Ongoing Trends**

El reconocimiento de emociones en el habla (SER) ha avanzado desde que comenzó hace más de veinte años. La idea es que las máquinas puedan interpretar nuestras emociones a partir de cómo hablamos, tanto en el tono como en las palabras que usamos. Esto permitirá que la inteligencia artificial pueda entendernos mejor y responder de manera más humana, mejorando desde la atención al cliente hasta la salud mental.

### **Modelos de Representación de Emociones**

Para que las máquinas logren entender las emociones humanas, se utilizan modelos de representación que pueden ser de clases discretas, como las emociones básicas que propuso Ekman, o de dimensiones continuas, que describen las emociones a través de ejes como la activación y la valencia. Etiquetar datos emocionales es un reto importante debido a su naturaleza subjetiva y se realiza mediante autoevaluación o anotación externa, usando métodos como la votación mayoritaria para asegurar que las etiquetas sean confiables.

### **Técnicas y Herramientas en SER**

Los sistemas SER tradicionales se basan en la extracción de características tanto del audio, como la información espectral y prosódica, como del texto, utilizando técnicas avanzadas como "word embeddings". El aprendizaje automático desempeña un papel crucial en este proceso, con modelos tradicionales como las máquinas de soporte vectorial (SVM) y los bosques aleatorios (RF). Además, existe un creciente uso de redes neuronales profundas, que tienen la capacidad de aprender directamente de los datos.

### **Tendencias y Desafíos**

El campo del SER está siendo impulsado por tendencias como el modelado holístico del hablante, el aprendizaje débilmente supervisado y por transferencia, y el uso de características aprendidas directamente de los datos. Sin embargo, SER enfrenta desafíos significativos, como la escasez de datos etiquetados, la necesidad de robustez ante variaciones culturales y lingüísticas, y la distinción entre la emoción expresada y la sentida.

### **Futuro del SER**

A pesar de estos desafíos, el futuro del SER tiene expectativas de modelos más holísticos y adaptativos que integren la información contextual. Se espera que esta tecnología se utilice ampliamente en aplicaciones de salud, seguridad e interfaces de usuario más inteligentes, transformando nuestra interacción con la tecnología.