

4.3 Reconocimiento de eventos acústicos

El contexto de esta tarea se encuentra en la unidad de curas intensivas de neonatos (NICU). En este contexto, hay muchos ruidos que se tienen que filtrar para estudiar los patrones relevantes. Se pretende grabar y etiquetar datos recogidos de micrófonos instalados en las incubadoras de las NICU. La base de datos incluirá información sobre las variables fisiológicas relevantes y los patrones de sueño.

4.4 Síntesis de voz

El aprendizaje profundo se ha integrado en síntesis de voz principalmente aplicado a la modelización paramétrica (Ling et al., 2015)

La tarea de síntesis de voz es básicamente una tarea de regresión. Con tal de producir voz natural y continua se pueden utilizar técnicas de generación paramétrica. En esta área, proponemos investigar representaciones de la voz que permitan usar redes neuronales. También pretendemos proponer y evaluar técnicas de aprendizaje profundo para reducir el ruido de la voz generada e incluir expresividad en la voz final.

4.5 Traducción automática

En este caso, el aprendizaje profundo se ha usado para mejorar los sistemas estadísticos ya existentes y también ha permitido desarrollar un nuevo paradigma de traducción usando un modelado de secuencia a secuencia. Como en las otras áreas, la lista de trabajos es muy extensa (Costa-jussà et al., 2017).

La traducción automática se puede aplicar a la voz o al texto. El objetivo al final de este proyecto es construir un sistema de traducción de voz a texto, ya sea concatenando técnicas de reconocimiento de voz y traducción de texto o planteando un sistema directo de voz a texto traducido. En el primer caso, se integrarán las mejoras del paquete de reconocimiento de voz y las mejoras que aporta un paradigma de traducción automática basado en redes neuronales. En el segundo caso, se diseñará una nueva arquitectura neuronal para afrontar el reto.

5 Impacto del proyecto

Las tecnologías de voz pueden facilitar el acceso a la información (comunicación hombre-máquina) y la comunicación humana. Los dispositivos electrónicos se están convirtiendo

en imprescindibles. El uso de la voz en estos dispositivos es cada vez más esencial y también puede abrir una nueva gama de posibilidades. Estas tecnologías también pueden aplicarse a múltiples campos específicos, como mejorar la comunicación y la comprensión de los seres humanos, ayudar a las personas discapacitadas y ancianas, mejorar los servicios ofrecidos en los medios de comunicación, etc. El empleo de dispositivos de voz con voces inadecuadas (género, edad, acento, dialecto, tono) o sistemas de reconocimiento de voz que no funcionan en condiciones ruidosas pueden desalentar a los usuarios. El desarrollo que estamos proponiendo de la tecnología de voz será la clave para aplicaciones robustas de alta calidad. Asimismo, la traducción es un aspecto importante para reducir las barreras internacionales y lograr el pleno entendimiento entre las personas, preservando al mismo tiempo las sociedades multilingües. Esperamos realizar traducciones de voz en tiempo real y de alta calidad con concatenación e integración de reconocimiento profundo de voz y tecnologías de traducción automática. Esto representaría un progreso claro en los negocios y las relaciones políticas, así como en las áreas de ocio y educación.

Nuestra propuesta de investigación sobre detección de eventos acústicos también incluye su aplicación específica en unidades de cuidados intensivos neonatales (NICU). En este caso, se diferenciarán los factores de ruido microambiental y los signos fisiológicos y así los clínicos podrán proponer mejores protocolos NICU.

6 Página web

En la página web del proyecto

<http://www.tsc.upc.edu/deepvoice/>

se puede consultar el equipo de investigación. En la misma página también se harán públicos los principales resultados alcanzados con el progreso de DeepVoice.

Bibliografía

- Bengio, Y. 2009. Learning deep architectures for ai. *Found. Trends Mach. Learn.*, 2(1):1–127, Enero.
- Costa-jussà, M. R., A. Allauzen, L. Barrault, K. Cho, y H. Schwenk. 2017. Introduction to the Special Issue on Deep Learning Approaches for Machine Translation. *Accepted for publication in Computer Speech*

- and Language, Special Issue in Deep learning for Machine Translation.*
- Hannun, A. Y., C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates, y A. Y. Ng. 2014. Deep speech: Scaling up end-to-end speech recognition. *CoRR*, abs/1412.5567.
- Hinton, G. E., S. Osindero, y Y. Teh. 2006. A fast learning algorithm for deep belief nets. *Neural Comput.*, 18(7):1527–1554, Julio.
- Ling, Z., S. Kang, H. Zen, A. W. Senior, M. Schuster, X. Qian, H. M. Meng, y L. Deng. 2015. Deep learning for acoustic modeling in parametric speech generation: A systematic review of existing techniques and future trends. *IEEE Signal Process. Mag.*, 32(3):35–52.
- Richardson, F., D. A. Reynolds, y N. Dehak. 2015. A unified deep neural network for speaker and language recognition. *CoRR*, abs/1504.00923.
- Xia, B. y C. Bao. 2014. Wiener filtering based speech enhancement with weighted denoising auto-encoder and noise classification. *Speech Communication*, 60:13–29.