

Proyecto curso IA-Bio 2024

Reconocimiento de enfermedades respiratorias por la tos

La tos es un sonido repentino, forzado y seco para liberar aire y despejar la irritación en la garganta o las vías respiratorias. La tos puede tener diversas causas que incluso pueden no ser una enfermedad subyacente. En este proyecto se desea realizar un análisis del audio de diferentes toses causadas por enfermedades respiratorias, específicamente, asma, gripe, COVID-19 y EPOC. El objetivo del proyecto consiste en calcular el espectrograma de las muestras de tos y utilizarlos para su discriminación. El proyecto tiene dos objetivos: (i) por un lado utilizar un *autoencoder* para procesar los espectrogramas con la intención de compactar la información y mejorar la representación al determinar posibles relaciones entre los elementos del espectrograma; (ii) por otro lado, utilizar una técnica de aumento de datos directamente sobre los espectrogramas para incrementar el conjunto de datos, sobretodo en las clases minoritarias.

Plan tentativo de actividades a desarrollar:

1. Familiarización con los corpus propuestos para abordar esta tarea.
2. Realizar una caracterización del audio basado en espectrogramas.
3. Utilizar un esquema de clasificación utilizado para imágenes cuando como entrada el espectrograma del audio.
4. Aplicar un autoencoder y comprobar su beneficio al cambiar la representación.
5. Aplicar la técnica de aumento de datos.
6. Evaluación del modelo propuesto y realizar un estudio de ablación para determinar el beneficio de los dos puntos anteriores.
7. Comparación con métodos del estado del arte para abordar dicha tarea.

Referencias:

Bertini, F., Allevi, D., Lutero, G., Calzà, L., & Montesi, D. (2022). An automatic Alzheimer's disease classifier based on spontaneous spoken English. *Computer Speech & Language*, 72, 101298.

Freitag, M., Amiriparian, S., Pugachevskiy, S., & Cummins, N. (2018). audeep: Unsupervised learning of representations from audio with deep recurrent neural networks. *Journal of Machine Learning Research*, 18(173), 1-5.

Ma, X., Yang, H., Chen, Q., Huang, D., & Wang, Y. (2016, October). Depaudionet: An efficient deep model for audio based depression classification. In *Proceedings of the 6th international workshop on audio/visual emotion challenge* (pp. 35-42).

Park, D. S., Chan, W., Zhang, Y., Chiu, C. C., Zoph, B., Cubuk, E. D., & Le, Q. V. (2019). Specaugment: A simple data augmentation method for automatic speech recognition. *arXiv preprint arXiv:1904.08779*.

Enlace para acceder a los datos:

https://drive.google.com/drive/folders/1yclnQ2lv3JzI7P8oqAHDlhn3wl6SVW-r?usp=drive_link

Se necesita solicitar acceso con su cuenta de Google.

El siguiente artículo describe los datos y detalla cómo están agrupados, así como experimentos iniciales que se hicieron con estos datos. Los positivos son pacientes con enfermedades respiratorias crónicas (Asma y EPOC) y los negativos son pacientes sanos o pacientes con enfermedad respiratoria aguda (COVID o gripa). El código de estos experimentos puede consultarse en libreta de Colab.

https://colab.research.google.com/drive/1GdH8y33PzTUNOc_uLfpQAU8k5ZyVRizJ?usp=sharing

Estos experimentos iniciales usan una representación pre-entrenada (wav2vec), y sólo usa un subconjunto de los datos. Los conjuntos de audio usados en el Colab que se probaron en el artículo son diferentes. Los csv que están en el Colab están en las secciones de los datos.

Link to access the CRD DB

/content/drive/MyDrive/InaoeProyecto/BDSegBalanceada/foldsCompleto

Link to access the Respiratory Conditions DB

/content/drive/MyDrive/InaoeProyecto/BDSegBalanceada/newfoldsCompleto