

Detección Automática de Trastornos de la Voz desde una Perspectiva Práctica

Jazmín Vidal¹ 0009-0008-1306-7761, Dayana Ribas² 0000-0003-3813-4998, Cyntia Bonomi¹ 0009-0009-0089-5776, Eduardo LLeida² 0000-0001-9137-4013, Luciana Ferrer¹ 0000-0002-0426-8683, and Alfonso Ortega² 0000-0002-3886-7748

¹ Instituto de Investigación en Ciencias de la Computación (ICC), Universidad de Buenos Aires-CONICET, Argentina

² ViVoLab, Aragón Institute for Engineering Research (I3A), University of Zaragoza, Spain

Resumen. Los trastornos de la voz, como la disfonía, son frecuentes y en muchos casos, no reciben tratamiento hasta que se vuelven graves. Facilitar su detección podría contribuir a un diagnóstico temprano y a un tratamiento oportuno. En este estudio, abordamos los aspectos prácticos de la detección automática de trastornos de la voz (AVDD). Los datos anotados para trastornos de la voz suelen ser escasos debido a los desafíos involucrados en la recolección y anotación de dichos datos. Sin embargo, existen algunos conjuntos de datos relativamente grandes disponibles para un número reducido de dominios. En este contexto, proponemos el uso de una combinación de datos fuera del dominio y dentro del dominio para entrenar un sistema AVDD basado en redes neuronales profundas, y ofrecemos orientación sobre la cantidad mínima de datos dentro del dominio necesarios para lograr un rendimiento aceptable. Además, proponemos el uso de una métrica basada en costos, el costo esperado normalizado (EC), para evaluar el rendimiento de los sistemas AVDD de manera que refleje de cerca las necesidades de la aplicación. Como beneficio adicional, las decisiones óptimas para el EC pueden tomarse de forma fundamentada mediante la teoría de decisión de Bayes. Finalmente, argumentamos que, para aplicaciones médicas como AVDD, las decisiones categóricas deben ir acompañadas de puntuaciones interpretables que reflejen la confianza del sistema. Aquí mostramos que añadir una etapa de calibración entrenada con una pequeña cantidad de datos dentro del dominio puede mejorar estos modelos y apoyar a los profesionales en su toma de decisiones.

Palabras clave: Detección Automática de Trastornos de la Voz, Calibración, Modelos Auto-supervisados, proper scoring rules, aplicaciones en salud.

Automatic Voice Disorder Detection from a Practical Perspective

Abstract. Voice disorders, such as dysphonia, are common and often go untreated until they become severe. Assisting the detection of voice disorders could facilitate early diagnosis and subsequent treatment. In this study, we address the practical aspects of automatic voice disorders detection (AVDD). Data annotated for voice disorders is usually scarce due to challenges involved in data collection and annotation of such data. However, some relatively large datasets are available for a reduced number of domains. In this context, we propose using a combination of out-of-domain and in-domain data for training a deep neural network-based AVDD system and offer guidance on the minimum amount of in-domain data required to achieve acceptable performance. Further, we propose the use of a cost-based metric, the normalized expected cost (EC), to evaluate performance of AVDD systems in a way that closely reflects the needs of the application. As an added benefit, optimal decisions for the EC can be made in a principled way given by Bayes decision theory. Finally, we argue that for medical applications like AVDD, the categorical decisions need to be accompanied by interpretable scores that reflect the confidence of the system. Here, we show that adding a calibration stage-trained with a small amount of in-domain data can improve these models and support professionals in their decision-making.

Keywords: Automatic Voice Disorder Detection, Calibration, Self-supervised models, proper scoring rules, health applications

DOI: <https://doi.org/10.1016/j.jvoice.2024.03.001>

Publicado en: *Journal of Voice*

Factor de impacto: 2.5 (Q2).