



Propuesta de Tesis

Ing. Juan Manuel Miguez

Buenos Aires

Junio 2022

Contenido

Título	3
Resumen	3
Antecedentes y Planteamiento del problema	3
Objetivos	4
Justificación del tema	5
Índice	7
1.0 Introducción	7
2.0 Fundamentos Teóricos	7
3.0 Caso de prueba sobre identificación de hablantes	7
4.0 Conclusiones	7
5.0 Futuras líneas de investigación	7
6.0 Bibliografía	7
7.0 Anexos	7
Bibliografía	8
Citas Bibliográficas	8
Referencias Bibliográficas	8
Documentos electrónicos	8

1. Título

Reconocimiento automático de hablantes empleando técnicas de aprendizaje profundo, en peritajes informáticos dentro del ámbito de la Ciudad de Buenos Aires.

2. Resumen

¿Puede un perito en informática, elaborar un dictamen que demuestre con convicción que dos registros de voces han sido realizados por la misma persona?

El reconocimiento automático de hablantes está dentro de las incumbencias del perito, no obstante, el método científico y el modelo de inteligencia artificial utilizados para llegar a la conclusión del informe, es objeto de controversias científicas.

Tras haber analizado los modelos, las controversias, y ejemplos de aplicación, nos encontramos en condiciones de aplicar un desarrollo que permite desambiguar conclusiones y emitir dictámenes periciales sobre la compatibilidad de registros de voz, en contextos acotados por la efectividad del proceso.

3. Antecedentes y Planteamiento del problema

Según publicaron D. L. Faigman, D. H. Kaye, M. J. Saks and J. Sanders en el año 2002, históricamente los científicos de todo el mundo desalentaron el reconocimiento automático de hablantes por controversias técnicas.

Los primeros registros sobre identificación de hablantes a fines de los años 40', empleaban técnicas de reconocimiento por escucha, haciendo uso de la discriminación empírica de voces por parte de oficiales forenses. Hacia fines de los años 70' aparecieron las primeras aplicaciones electrónicas biométricas pero con baja efectividad, no aplicables en el ámbito judicial. Actualmente, científicos del Conicet modelaron un proceso de aprendizaje automático que podría ser utilizado en casos judiciales reservados.

La solicitud de generación de prueba informática para la resolución de casos judiciales en el fuero penal, haciendo uso de técnicas relacionadas con Inteligencia Artificial, incluye la identificación forense de personas por métodos biométricos.

Debido a la capacitación de los letrados en ejercicio en el ámbito de la Ciudad de Buenos Aires en herramientas relacionadas con Tecnología de la Información, la

solicitud de este tipo de prueba se incrementó y comenzó a propagarse a otros fueros.

La designación de un experto capaz de elaborar un dictamen de este tenor, es crítico, ya que el procedimiento se realiza por sorteo sobre una base de peritos con restricciones a este tipo de tecnología; demorando en promedio entre 2 semanas a 2 meses la designación del especialista y si no resulta idóneo, se repite el proceso¹¹.

Esta actividad dejó de ser un tema exclusivo de la fuerza pública (policía, prefectura, gendarmería), siendo que ahora está dentro de las incumbencias del perito oficial de la especialidad en Informática (cualquier jurisdicción y fuero)

El Dr. Pedro Univaso³, en el año 2016, publicó que los avances en materia de inteligencia artificial relacionados con esta actividad, están en etapa de formalización de procesos de mejores prácticas con el fin de estandarizar las metodologías empleadas en cada laboratorio, que le permitan al profesional el modelado de su proceso de acuerdo a la necesidad.

A todas luces, un experto en la materia necesita desarrollar un proceso que determine automáticamente si dos registros de voz son compatibles dentro de un intervalo de confianza, para elaborar un informe pericial sustentable, sin necesidad de recurrir a un proveedor cuyo producto cerrado podría obligar a responder por tecnicismos que den lugar a controversias.

4. Objetivos

Objetivo general:

- Determinar qué modelo de aprendizaje automático puede ser utilizado por un experto en informática, para resolver el problema de identificación automática de hablantes. El mismo deberá permitir la elaboración de un dictamen, con un alto nivel de confianza, en el marco de un proceso judicial, y que además evite su impugnación por motivos técnicos.

Objetivos específicos:

- Desarrollar una maqueta que permita resolver el problema de identificación de hablantes, utilizando “deep learning”, identificando todas las etapas del proceso; desde la preparación de los datos de aprendizaje, la toma de la muestra de la voz, hasta la conclusión del proceso.
- Analizar en qué otros campos de aplicación podría ser utilizado este proceso.

5. Justificación del tema

El Diario Judicial¹ publicó el 22-Mayo-2022, que existen 85 mil abogados matriculados en la Ciudad Autónoma de Buenos Aires. De esa cantidad, según el Colegio Público de Abogados de la Capital Federal² (Cpacf), existen 71 mil abogados en ejercicio de actividades.

El listado de peritos informáticos que publicó el Poder Judicial de la Nación (www.pjn.gov.ar) al 29-Mayo-2022, cuenta con 38 ingenieros en informática y 76 analistas en informática.

Según sondeos realizados por el Consejo de Profesionales en Ciencias Informáticas de la Ciudad de Buenos Aires (<https://www.cpci.org.ar/>), la tercera parte de los abogados en ejercicio, mantiene en promedio 3 causas en curso con al menos 1 de ellas requiriendo algún peritaje informático.

En el citado caso ut-supra (CIV 034762/2017), hubo una demora de al menos 2 años en la designación de un perito idóneo, y una vez aceptado el cargo, el experto no logró acceder al expediente para conocer el tenor de la prueba presentada, propagando otra demora de 2 años.

Si el método de resolución de este tipo de problemática se lograra automatizar, entonces el profesional en informática podría agilizar el tiempo de respuesta de sus asignaciones en tiempo y forma. Si así fuera, el proceso de identificación automática de hablantes no dependería de la subjetividad humana, sino de un proceso estándar con un alto nivel de certeza, que permitiría elaborar informes muy precisos.

Si resultara aplicable el análisis por aprendizaje automático utilizando “deep learning”, se podría otorgar la libertad de elección de un profesional idóneo, para la realización de la tarea; independizando a los expertos de formar un vínculo con el único actor en el mercado que provee servicios de manera privada, y aceleraría el proceso judicial debido a tiempos de espera por falta de disponibilidad de expertos; otorgando además, beneficios a la comunidad de informáticos que incursionen en dicha actividad.

Como consecuencia de este análisis, una aplicación a futuro es la securización de aplicaciones mediante la integración del reconocimiento biométrico por voz, empleando -por ejemplo- protocolo OAuth2 con keycloak⁵ complementando otra modalidad de autenticación como por ejemplo “reconocimiento biométrico facial”.

Actualmente, el reconocimiento de voz es utilizado en las siguientes actividades:
De la voz al comando,

- Dictado automático: Es uno de los usos más comunes en la actualidad. Por ejemplo, el dictado de recetas médicas y diagnósticos o el dictado de textos legales, se usan corpus especiales para incrementar la precisión del sistema.
- Control por comandos: Consiste en dar órdenes a un procesador. Por ejemplo, "abrir navegador", "encender luces" o "cerrar ventana". Estos sistemas reconocen un vocabulario reducido, incrementando su rendimiento. Ver módulo de reconocimiento de voz de 7 comandos en venta online [7]
- Telefonía: Algunos sistemas PBX permiten a los usuarios ejecutar comandos de voz, en lugar de pulsar tonos. Por ejemplo, el usuario podría decir un número para navegar por un menú.
- Sistemas portátiles: Son sistemas de tamaño reducido, como relojes o teléfonos móviles, que tienen restricciones muy concretas de tamaño. Aquí la voz es una solución natural para introducir datos en sus aplicaciones.
- Sistemas diseñados para discapacitados: diseñados para personas con discapacidades, con restricciones para teclear con fluidez, así como para personas con problemas auditivos, que pueden usarlos para obtener texto escrito a partir de habla. Por ejemplo, para que los hipoacúsicos puedan recibir llamadas telefónicas.

Del comando a la voz,

- Generador de voz: Generadores de contenido de video o audio, narraciones virtuales, en lugar de contratar a un locutor profesional, se utilizan herramientas de Inteligencia Artificial. Se trata de software que otorga un aspecto realista, natural y de calidad a las voces⁶
- Clonar voces: software que mediante Inteligencia Artificial permite clonar la voz de una persona, de manera que luego de entrenar el modelo permite ingresar un texto y escuchar la voz clonada⁷. Existen controversias éticas acerca de la creación de una voz clonada.
- Ataques por clonación de voz (Deep Fake): para engañar a los sistemas biométricos de voz haciendo parecer que se escuchan reales, mediante la clonación de voz (Inteligencia Artificial para la simulación de voz sintética)⁸ se trata de una nueva generación de estafas de phishing, o de método de clonación de voz con fines malignos.
- Educación: Para la enseñanza educativa y los audios en los museos
- Audiolibros: Las voces de las celebridades se pueden usar para narrar libros, y las figuras históricas pueden contar sus propias historias con sus voces.
- Tecnología de asistencia: Para ayudar a las personas con discapacidades o con problemas de salud, que afectan su habla.
- Autenticación biométrica por voz: utilizada por bancos, para autenticar el acceso de personas a sus plataformas en línea.

6. Índice

1.0 Introducción

- 1.1 Objetivos
- 1.2 Estructura de la Tesis

2.0 Fundamentos Teóricos

- 2.1 Marco teórico sobre procesamiento de voz.
- 2.2 Modelos de aprendizaje automático usando técnicas de deep-learning.
- 2.3 Interpretación de resultados.
- 2.4 Identificación de Voz
- 2.5 Identificación de Hablantes

3.0 Caso de prueba sobre identificación de hablantes

- 3.1 Ajustes de configuración de la maqueta.
- 3.2 Carga del Set de datos.
- 3.3 Preparación de datos.
- 3.4 Preparación de ruido.
- 3.5 Recodificación de muestras con ruido.
- 3.6 Generación de Datasets.
- 3.7 Definición del Modelo.
- 3.8 Entrenamiento.
- 3.9 Validación del Modelo.
- 3.10 Elección de la muestra para ejecutar una demo.
- 3.11 Ejecución del proyecto.
- 3.12 Análisis de resultados.

4.0 Conclusiones

5.0 Futuras líneas de investigación

6.0 Bibliografía

7.0 Anexos

- 7.1 Código principal de ejemplo.
- 7.2 Código de soporte.

7. Bibliografía

Citas Bibliográficas

[11] Transcripción de la solicitud de generación de prueba y la designación de un experto: *"CIV 034762/2017, ARIAS, DIEGO ANTONIO C/ ARIAS DE PIÑEIRO, MARTA MAFALDA Y OTRO S/RENDICION DE CUENTAS, a fs.1105, C-PERITO INGENIERO EN INFORMÁTICA : Se designe por este juzgado y secretaría Perito Ingeniero en informática, quien previa aceptación del cargo por ante el actuario, deberá citar a Pablo Andrés Piñeiro, a fin de "auditar" el dvd que se agrega con la " llamada que le hiciera a Leonor Coca" y determinar si la voz le pertenece al co demandado Piñeiro. Se realizarán las pericias en presencia de la parte actora, con citación de la misma conforme el art. 471 del Código Procesal."* - <https://www.pjn.gov.ar>

Referencias Bibliográficas

[10] Andrés Felipe Mendoza Marín, Dagoberto Porras Plata (2016) - "Desarrollo de un sistema de verificación del hablante basado en modelos de mezclas gaussianas".

[4] D. L. Faigman, D. H. Kaye, M. J. Saks and J. Sanders (2002). "Modern scientific evidence: Vol. 2", In The law and science of expert testimony, pp.430-446

[3] Dr. Pedro Univaso (2016). "Towards a Unified Methodology for Forensic Speaker Identification", pp.2-6

[9] José Esteban Calle Chuchuca (2022) - "Diseño y desarrollo de un sistema de prototipo para reconocimiento automático del hablante empleando técnicas de aprendizaje profundo".

[12] Martín Calle (2019). "Reconocimiento de voz basado en características DNN Bottleneck"

Documentos electrónicos

[2] Cpacf. "Elecciones Colegio de abogados CABA":
<https://www.cronista.com/economia-politica/votan-los-abogados-portenos-cuando-es-la-eleccion-quienes-pueden-participar-y-que-listas-se-presentan/>

[1] Diario judicial (2022). "La matriculación de abogados en Capital Federal creció un 30 %":

<https://www.diariojudicial.com/nota/13402#:~:text=De%20esta%20manera%2C%20existen%20actualmente,31.945%20hombres%20y%2026.962%20mujeres.>

[8] Instituto Internacional de Seguridad Cibernética <http://www.iicibersecurity.com>
(2021) - "Ataques por clonación de voz" -
<https://www.youtube.com/watch?v=VnasUX7mxns>

[6] Mario Carvajal (2021) "Convertir texto a voz para con IA para tours virtuales"
<https://www.youtube.com/watch?v=6ce3PhAcwkE>

[7] Síntesis Vocal (2022) "Cómo clonar voces en Español con IA - y usarla en FakeYou"
- <https://www.youtube.com/watch?v=pd7pMrMNAdg>

[5] SoftJourn (2020) "Is Voice Authentication Secure Enough to be Your New Password?":
<https://softjourn.com/insights/security-considerations-in-voice-authentication>