MACHINE LEARNING Y DEEP LEARNING EN LA INTERPRETACIÓN DEL LENGUAJE DE SEÑAS

Raúl Oscar Klenzi, Maria Isabel Masanet, Facundo Recabarren, Silvia Saez, Gustavo Conturso

Instituto de Informática / Departamento de Informática /Facultad de Ciencias Exactas Físicas y Naturales / Universidad Nacional de San Juan

Av. Ignacio de la Roza 590 (O), Complejo Universitario "Islas Malvinas", Rivadavia, San Juan, Teléfonos: 4260353, 4260355 Fax 0264-4234980, Sitio Web: http://www.exactas.unsj.edu.ar
e-mail: {rauloscarklenzi, mimasanet}@gmail.com

RESUMEN

La pandemia COVID19 puso en evidencia la gran dificultad de la comunidad hipoacúsica para comunicarse con el resto de la sociedad. El uso de barbijos, máscaras y barreras transparentes como alternativas atenuadoras de contagios generó una barrera casi infranqueable para esa comunidad cuando utilizaban la lectura de labios para establecer la comunicación. Por ello, la única alternativa que tenían era al uso de la Lengua de Señas (LS), herramienta que el resto de la sociedad, en su mayoría, desconoce; por lo que debían y deben recurrir a una tercera persona que no siendo hipoacúsica y comprendiendo la lengua de señas hiciera de interprete.

Los objetivos que persigue el presente trabajo se centran en la utilización de algoritmos derivados del Machine Learning (ML) y Deep Learning (DL) aplicado al reconocimiento de expresiones en LS a partir de una secuencia de imágenes (video), y lograr traducir a texto o audio estas expresiones, como así también el camino inverso, emulando con ello al interprete humano. Estos objetivos exigen la utilización de hardware veloz tipo GPU, gran capacidad de memoria, y algoritmos eficientes para el procesamiento de la información.

Palabras clave: Machine Learning, Deep Learning, LSA.

CONTEXTO

En el marco del proyecto “Evaluación de visualizaciones eficientes en Ciencia de Datos” concluido en diciembre de 2022 se comenzó la ejecución, bajo el asesoramiento de integrantes del proyecto, de un trabajo final de grado en la carrera Licenciatura en Ciencias de la Computación. El trabajo, “Sistema Web Intérprete de Lengua de Señas Argentina” permitió ir conociendo la punta de un iceberg evidenciado por las particularidades de la Lengua de Señas Argentina (LSA), el contexto de la comunidad sorda donde lo aplica, y la necesidad cierta y mundialmente explicitada de generar un proceso automático que permita la intercomunicación de la comunidad hipoacúsica con el resto de la sociedad. Esto habilitó que tras la incorporación de intérpretes de LSA al grupo de trabajo, se presentara el proyecto para el bienio 2023-2024 “MACHINE LEARNING Y DEEP LEARNING APLICADO A LENGUA DE SEÑAS ARGENTINA”, el cual próximamente será evaluado en el ámbito de la UNSJ por investigadores externos a nuestra Universidad. El background obtenido por los investigadores en el área del aprendizaje profundo aplicado al procesamiento de imágenes permite realizar la propuesta e intentar dar un primer paso en el objetivo de facilitar la comunicación entre ambas

comunidades, la hipoacúsica y la oyente, cuya necesidad y dura realidad se evidenció en forma acrecentada en la pandemia COVID19.

1. INTRODUCCIÓN

En las últimas dos décadas, el campo del Machine Learning (aprendizaje automático) ha jugado un papel relevante en el desarrollo de aplicaciones de software [1]. Se trata de un área de las ciencias de la computación y de la ciencia de datos (Data Science) que utiliza algoritmos y métodos estadísticos para aprender de los datos, extraer inferencias y reconocer patrones sin programación explícita. El Deep Learning (aprendizaje profundo) es una subárea del Machine Learning (Figura 1) que utiliza una red neuronal de varias capas para imitar el complejo procesamiento de información del cerebro humano [2].

La explosión de datos digitales en una amplia variedad de dominios, como la ciencia, la ingeniería, el internet de las cosas, la biomedicina, la atención médica y muchos sectores comerciales, ha declarado la era de los grandes volúmenes de datos (big data). Estos datos no pueden analizarse con las estadísticas clásicas, sino que requieren de las más modernas técnicas de ML y DL. Dado que el aprendizaje automático aprende de los datos en lugar de programar reglas de decisión codificadas, se está intentando utilizar el aprendizaje automático para crear computadoras que puedan resolver problemas como expertos humanos en el campo [1].

Las personas con discapacidad verbal y auditiva enfrentan dificultades en la comunicación, debido a esto, se sienten aisladas y dependientes [1]. Estas personas utilizan la lengua de señas para comunicarse, pero la interacción social se dificulta debido a que la mayoría de la población no puede interpretar la lengua de señas.

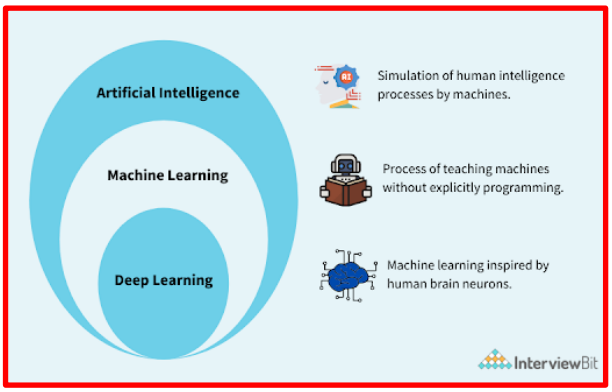

Figura 1 - Machine Learning y Deep Learning en la Inteligencia artificial.

Esta dificultad en la comunicación con el mundo que habla verbalmente, hace difícil obtener educación, trabajo y muchos otros requerimientos básicos. Para comunicarse con la sociedad, necesitan un intérprete de señas manual o automático. El primero puede convertir cualquier secuencia de gestos complicada a voz con todas las modulaciones emocionales requeridas para ello, pero reduce la privacidad de las personas involucradas en la conversación.

Un intérprete no manual o automático garantiza el factor de privacidad, pero aún se está investigando para desarrollar un sistema completo de reconocimiento de lengua de señas que pueda funcionar con la misma precisión que un intérprete humano. La complejidad del problema se evidencia en los desafíos involucrados en la interpretación automática de la lengua de señas, como ser:

• La lengua de señas es muy diversa; difiere de una geografía a otra, e incluso dentro de una región.

• La lengua de señas involucra gestos con una sola mano, gestos con dos manos, expresión facial y postura corporal.

• La conversación en lengua de señas contiene gestos estáticos y dinámicos realizados en una secuencia y, por lo tanto, encontrar el comienzo y el final de cada gesto en la

secuencia implica tareas de preprocesamiento en los datos (videos).

• Oclusión: Mientras se realiza el gesto, una mano puede ocluir a la otra, lo que dificulta la interpretación [1].

En este caso la propuesta contempla el desarrollo de un prototipo de aplicación web que interprete de forma automática expresiones de la Lengua de Señas Argentina (LSA) y emita, en formato escrito o de audio, la interpretación realizada. Dicha aplicación será testeada por los intérpretes de LSA integrantes del proyecto como así también por otros miembros de la comunidad hipoacúsica de San Juan.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación consideradas son las atinentes al procesamiento de imágenes mediante la utilización de algoritmos de ML y DL, que permitan hacer una interpretación adecuada y automática de los movimientos expresados por el hipoacúsico obteniendo así el correspondiente fragmento de texto o audio.

La investigación se orientará a integrar el Machine Learning y Deep Learning con la Lengua de Señas Argentina, con el objetivo de generar conocimiento de interés para desarrollos tecnológicos que faciliten la comunicación e interacción entre personas que utilizan la LSA y personas oyentes.

Técnicas de Machine Learning y Deep Learning, Visión por computadora (Computer Vision) y la Lengua de Señas Argentina son los principales conceptos sobre los que se basará la investigación propuesta.

2.1. Machine Learning

Dentro del área de la Inteligencia Artificial (IA) se desprende una rama específica conocida como Machine Learning (ML) o Aprendizaje de Máquina. Es una rama de algoritmos computacionales diseñados para emular la inteligencia humana mediante el aprendizaje del entorno [3].

2.2. Deep Learning

El Deep Learning es un área del Machine Learning que a través de una red neuronal trata de imitar el complejo procesamiento de información del cerebro humano.

Al igual que las neuronas biológicas en el cerebro humano, se utiliza un perceptrón como componente básico de una red neuronal artificial y realiza el procesamiento de la información [2].

El aprendizaje profundo está siendo ampliamente aceptado para tareas de visión por computadora, ya que ha demostrado capacidades casi humanas o incluso mejores para realizar numerosas tareas. Las redes profundas son efectivas en el reconocimiento visual, reconocimiento de objetos, seguimiento de movimiento, reconocimiento de acciones, estimación de la pose humana y la segmentación semántica [1].

2.2.1. Red Neuronal

Una red neuronal consiste en una gran cantidad de neuronas (o perceptrones) interconectadas entre sí en capas y es responsable del procesamiento paralelo masivo de datos de entrada para producir salidas (Figura 2). Consta de una capa de entrada, una capa de salida y varias capas ocultas.

En un sistema de aprendizaje profundo, las capas ocultas definen la profundidad de la red, la que depende de la naturaleza del problema en cuestión y del volumen de datos considerado para el procesamiento.

Las redes neuronales son potentes modelos de aprendizaje que logran resultados de vanguardia en una amplia gama de tareas de aprendizaje automático (ML) supervisado y no supervisado [4].

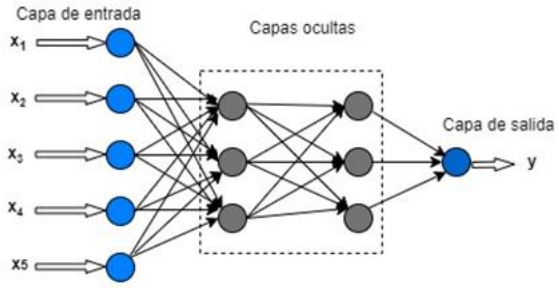

Figura 2 - - Arquitectura de una red neuronal con dos capas ocultas y una capa de salida [5].

2.3. Visión por computadora (Computer visión)

La visión por computadora se puede entender de distintas maneras, dependiendo del objetivo con el cual es usada. Puede ser utilizada para el procesamiento digital de imágenes, el análisis de imágenes, el reconocimiento de patrones o computación gráfica entre otras [6].

Implica interpretar y comprender el mundo a partir de sus imágenes o videos. Las tareas de la visión por computadora se pueden organizar jerárquicamente en niveles bajo, medio y alto. Mientras que las tareas de bajo nivel se centran en el procesamiento de imágenes para mejorarlas y extraer información, las tareas de nivel medio implican estimar las propiedades de los objetos en las imágenes, incluidas las propiedades geométricas, el movimiento y las categorías de objetos. Las tareas de alto nivel se centran en interpretar y comprender eventos o actividades en las imágenes/videos [7].

2.4. Lengua de señas

Existen muchas lenguas de señas en el mundo. La existencia de una o más lenguas en un mismo país o en más de un país depende de la historia de las comunidades Sordas de cada lugar [8]. La Lengua de Señas Argentina es la lengua de la comunidad sorda argentina. Es una lengua natural, con una estructura gramatical diferente a la del español. Esto significa que la LSA y el español son dos lenguas distintas.

Cada seña que compone a una lengua de señas está constituida por la combinación de 5 elementos: la configuración, la orientación, la ubicación, el movimiento y componentes no manuales (movimientos de hombros, cabeza, expresiones faciales).

La estructura básica entre el español y la Lengua de Señas Argentina difieren. En vez de la estructura del español ‘sujeto, verbo, objeto’ (Ejemplo: el perro corre la pelota), en la LSA el verbo va al final (Ejemplo: el perro a la pelota corre). Este orden es canónico, es decir, es el orden hacia el que la lengua tiende aunque en el discurso se viola por efectos semánticos y pragmáticos [9].

3. RESULTADOS OBTENIDOS Y ESPERADOS

Al momento, y fundamentalmente desde el desarrollo de un trabajo final de grado, se ha trabajado con el reconocimiento de letras expresadas en LSA como imágenes estáticas. Se han reconocido palabras en las que ya aparecen secuencias de imágenes (short video) y breves frases de entrenamiento realizadas por diferentes intérpretes; cuyos videos conforman el set de datos “LSA64” construido en [10]; y que, desde la representación realizada por diferentes personas hipoacúsicas se independizan del “quien” para centrarse en el “qué o cuál” es la secuencia analizada. Sin perder capacidad de reconocimiento.

Lo anterior llevó a tratar de reducir la tasa de frame que tiene cada video con el objetivo de disminuir el espacio de almacenamiento de las diferentes expresiones lingüísticas y acelerar el procesamiento sin perder capacidad de discernimiento y reconocimiento.

Para el preprocesamiento y desarrollo de los modelos de reconocimiento se usa el lenguaje de programación Python, el cual proporciona diferentes librerías como Matplotlib, Pandas, NumPy, Sci-Kit Learn, ScraPy, Keras, Tensor

Flow, entre otros, que ayudan a que los procesos de ciencia de datos y Machine Learning sean más manejables y eficientes. Por último, hay librerías que permiten el procesamiento de fotogramas, sin dudas necesarios para las tareas de preprocesamiento requeridas, como son MediaPipe, PoseNet, HandPose, OpenPose, entre otras. Estas librerías marcan los puntos del cuerpo, cara, brazos, manos, piernas en una secuencia de imágenes (video), necesarios para reconocer movimientos y con ello asociar a una expresión de la LSA.

4. FORMACIÓN DE RECURSOS HUMANOS

El proyecto de investigación está integrado por docentes investigadores responsables de cátedras de las carreras del DI-FCEFyN y que son inherentes a las diferentes columnas que dan soporte al área de conocimiento de la Ciencia de Datos, el Machine Learning y Deep Learning. Estos Docentes investigadores fomentan constantemente la participación de alumnos que integran el espacio e incitan a aquellos que cursan esas cátedras a abrazar la temática e integrarse al proyecto. Durante el pasado 2022, se defendieron dos tesis de grado en Licenciatura en Ciencias de la Computación.

Actualmente se encuentra en instancia de evaluación por la Comisión de Licenciatura un trabajo final, desde donde el grupo publicó en la JAIIO 2022 particularmente en el Congreso Argentino de AgroInformática - CAI 2022. Así mismo, se lleva adelante y seguramente se defenderá en el primer semestre de 2023 el trabajo final de Licenciatura en Ciencias de la Computación.

A nivel de posgrado, se ha presentado el informe final de una tesis de maestría, en etapa de evaluación por el tribunal asignado para su ulterior defensa.

5. BIBLIOGRAFÍA

[1] P. Singh, Ed., *Fundamentals and methods of machine and deep learning : algorithms, tools and applications*, 1st ed. USA, 2022.

[2] S. Dash, S. K. Pani, J. Rodrigues, and B. Majhi, Eds., *Deep learning, machine learning and IoT in biomedical and health informatics : techniques and applications*, Fourth. Massachusetts, 2022.

[3] I. El Naqa and M. J. Murphy, "What Is Machine Learning?," *Mach. Learn. Radiat. Oncol.*, pp. 3–11, 2015, doi: 10.1007/978-3-319-18305-3_1.

[4] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A Critical Review of Recurrent Neural Networks for Sequence Learning," 2015.

[5] C. C. Aggarwal, *Neural networks and deep learning : a textbook*, 1st ed. Cham, Switzerland: Springer Nature Switzerland AG, 2018.

[6] D. Mery, "Visión por Computador," Santiago de Chile, 2004.

[7] Q. Ji, "Computer vision applications," *Probabilistic Graph. Model. Comput. Vis.*, pp. 191–297, Jan. 2020, doi: 10.1016/B978-0-12-803467-5.00010-1.

[8] "Señario de términos y expresiones en Lengua de Señas Argentina – CAS." https://cas.org.ar/senario-de-terminos-y-expresiones-en-lengua-de-senas-argentina/ (accessed Dec. 03, 2022).

[9] M. I. Massone, "Lenguas de señas: 'cada comunidad desarrolló la propia por necesidad' | CONICET," 2012. Accessed: Nov. 11, 2022. [Online]. Available: https://www.conicet.gov.ar/lenguas-de-senas-cada-comunidad-desarrollo-la-propia-por-necesidad/.

[10] F. Ronchetti, "Reconocimiento de gestos dinámicos y su aplicación al lenguaje de señas," Mar. 2017, doi: 10.35537/10915/59330.