

# Lenguaje natural aplicado a los sistemas conversacionales para el aprendizaje de lenguas extranjeras

A.Vega (1), M.Bilbao(1), M.Falappa(2)

(1) Grupo de Investigación en Inteligencia Artificial Facultad de Ingeniería UNPSJB  
Ruta 1 Km 4 Comodoro Rivadavia - Chubut

[alejandrovega@unpata.edu.ar](mailto:alejandrovega@unpata.edu.ar), [martinbilbao@ing.unp.edu.ar](mailto:martinbilbao@ing.unp.edu.ar),

(2) Departamento de Ciencias e Ingeniería en Computación UNS

San Andrés 800 – Campus de Palihue - Bahía Blanca

[mfalappa@cs.uns.edu.ar](mailto:mfalappa@cs.uns.edu.ar)

## RESUMEN

La computación afectiva es un área de investigación emergente que reúne a investigadores y profesionales de varios campos, que van desde el estudio de la IA, el procesamiento del lenguaje natural hasta las ciencias cognitivas y sociales (Poria et al., 2017, 98-127). Las mismas se centran en el estudio y desarrollo de sistemas y dispositivos que pueden reconocer, interpretar, procesar y simular los afectos humanos.

El objetivo de la computación afectiva es lograr que una máquina o dispositivo sea capaz de simular la empatía i.e: “La capacidad de comprender el estado emocional de un individuo y, a menudo, expresar una respuesta que resuena con él” (Rodríguez et al., 2021), generando una inteligencia emocional que sea capaz de reconocer y responder de una manera más adecuada ante situaciones cotidianas. Para la detección de las emociones en el último tiempo se ha empezado a trabajar con sistemas que indican la utilización de más de una modalidad, también llamados multimodales, como por ejemplo la modalidad visual, de audio, de texto o

entradas fisiológicas. En este estudio se enfocará en el uso de información audiovisual, es decir bimodal.

Actualmente se utiliza el término deep learning (aprendizaje profundo), a partir de Big Data emocional. En (Shamim Hossain & Muhammad, 2019), se propone el procesamiento de una señal de voz para obtener un Espectrograma de Mel, y transformar las frecuencias de sonido según las escuchamos los humanos. Esto sirve, por ejemplo, para clasificaciones de canciones o clasificaciones del dictado (speech recognition). Luego, se extraen unos cuadros representativos de un segmento de video que alimentan una red neuronal convolucional (CNN, Convolutional Neural Network), y las salidas de esta red se fusionan utilizando dos máquinas de aprendizaje extremo (ELMs, extreme learning machines) consecutivas. La salida de esta fusión las califica una máquina de vectores de soporte (SVM, support vector machine) indicando la emoción encontrada, utilizando bases de datos emocionales audiovisuales. Por último, el estudio indica

si los sistemas que involucran CNN y ELMs son eficaces.

Por otro lado, existe el NLP (Procesamiento del Lenguaje Natural) como rama de la IA que ayuda a una computadora a entender, interpretar y manipular el lenguaje, y es una herramienta importante en el campo de la comunicación humano-máquina. Los avances de la última década han dado como resultados que estas se utilicen cada vez más en una gran variedad de campos como la atención médica, finanzas, marketing, etc. Se podría decir que el objetivo principal de las NLP es traducir el lenguaje humano (o natural) en comandos que puedan ser ejecutados por una computadora.

**Palabras claves: Computación Afectiva, Empatía, Sistemas Bimodales**

## CONTEXTO

El proyecto de investigación se desarrolla bajo convenio de actividad interinstitucional entre el Departamento de Ciencias e Ingeniería en Computación de la UNS y el Grupo de Investigación y el Grupo de Investigación en Inteligencia Artificial de la UNPSJB.

### 1. INTRODUCCION

Los avances tecnológicos actuales en el área de TICs, permiten disponer de cámaras y micrófonos en entornos familiares, sean dedicados o mediante dispositivos como smartphones o notebooks. Junto al creciente uso de redes

sociales, estos dispositivos permiten capturar grandes cantidades de información para la creación de una base de datos audiovisual para el análisis de sentimientos en tiempo real. Así, se puede dotar a los sistemas de la capacidad de interactuar con los usuarios de una manera más empática. Esto quiere decir, obtener una respuesta lo más cercana posible a la humana en las interacciones humano-máquina. El estudio de la computación afectiva y del análisis de los sentimientos son las claves para obtener Inteligencias artificiales (IAs), más complejas y cercanas al comportamiento humano, como así también para todos los campos de estudios que derivan de ellas.

El objetivo de la computación afectiva es lograr que una máquina o dispositivo sea capaz de simular la empatía i.e: “La capacidad de comprender el estado emocional de un individuo y, a menudo, expresar una respuesta que resuene con él” (Rodríguez et al., 2021), generando una inteligencia emocional que sea capaz de reconocer y responder de una manera más adecuada ante situaciones cotidianas. Para la detección de las emociones en el último tiempo se ha empezado a trabajar con sistemas que indican la utilización de más de una modalidad, también llamados multimodales, como por ejemplo la modalidad visual, de audio, de texto o entradas fisiológicas. En este estudio se enfocará en el uso de información audiovisual, es decir bimodal.

## **2. LINEAS DE INVESTIGACION, DESARROLLO E INNOVACION**

El principal objetivo de este trabajo es la obtención de información relacionada a la computación afectiva y el análisis de sentimientos como campos interdisciplinarios para el procesamiento y comprensión del lenguaje natural (NLP) aplicando enfoques híbridos. De esta manera, se podrá brindar una herramienta de aprendizaje de idiomas extranjeros con la cual sea posible tener una conversación fluida con una máquina y sea capaz de evaluar a la misma en tiempo real y de una manera empática.

### **Objetivos Específicos:**

- Evaluar y seleccionar diferentes métricas para bases de datos y repositorios con información sobre los sistemas conversacionales basados en lenguajes extranjeros.
- Desarrollar un enfoque híbrido (análisis de videos y audio) utilizado para la obtención de datos para los procesos de reconocimiento de emociones.
- Mejorar la respuesta en la interacción humano-computadora, a través del empleo de computación afectiva para el aprendizaje de lenguas extranjeras en los sistemas conversacionales actuales.
- Implementar algoritmos de computación afectiva que puedan ejecutarse en tiempo real.
- Diseñar una app de asistencia conversacional basada en el idioma

español y con aplicación de computación afectiva.

## **3. RESULTADOS Y OBJETIVOS ESPERADOS**

Realizar una buena obtención de datos tanto visuales como auditivos, para la conformación de una buena base de datos, para poder cumplir con ciertos objetivos específicos que podemos enumerar:

- Evaluar y seleccionar diferentes métricas para bases de datos y repositorios con información sobre los sistemas conversacionales basados en lenguajes extranjeros.
- Desarrollar un enfoque híbrido (análisis de videos y audio) utilizado para la obtención de datos para los procesos de reconocimiento de emociones.
- Mejorar la respuesta en la interacción humano-computadora, a través del empleo de computación afectiva para el aprendizaje de lenguas extranjeras en los sistemas conversacionales actuales.
- Implementar algoritmos de computación afectiva que puedan ejecutarse en tiempo real.
- Diseñar una app de asistencia conversacional basada en el idioma español y con aplicación de computación afectiva

#### 4. FORMACION DE RECURSOS HUMANOS

El equipo de trabajo se encuentra formado por tres investigadores, dos doctores y un alumno del Doctorado en Ciencias de la Ingeniería de la UNPSJB

Esta línea de investigación proporcionará un marco propicio para la iniciación de estudios de posgrado del alumno.

Dos integrantes de esta línea de investigación dirigen tesis de grado y posgrado en temáticas afines.

#### 5. REFERENCIAS

Cambria, E., Das, D., Bandyopadhyay, S., & Feraco, A. (2017). *Affective Computing and Sentiment Analysis. A Practical Guide to Sentiment Analysis. Socio-Affective Computing*, 5. [https://doi.org/10.1007/978-3-319-55394-8\\_1](https://doi.org/10.1007/978-3-319-55394-8_1)

Choi, J. I., Ahmadvand, A., & Agichtein, E. (2019, November). Offline and Online Satisfaction Prediction in Open-Domain Conversational Systems. *Conversational Systems*. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM '19)*, 1281–1290. <https://doi.org/10.1145/3357384.3358047>

Hamed Jelodar, H., & Wang, Y. (2020, 10 24). Deep Sentiment Classification and Topic Discovery on Novel Coronavirus or COVID-19 Online Discussions: NLP Using LSTM Recurrent Neural Network Approach. *IEEE J Biomed Health Inform*,

10, 2733-2742. [10.1109/JBHI.2020.3001216](https://doi.org/10.1109/JBHI.2020.3001216)

Lou, C., Liang, B., Gui, L., He, Y., Dang, Y., & Xu, R. (2021, 07). Affective Dependency Graph for Sarcasm Detection. *SIGIR '21: Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 44, 1844-1849. <https://doi.org/10.1145/3404835.3463061>

Majumder, B., Gupta, A., Surana, H., & Vajjala, S. (2020). *Practical Natural Language Processing: A Comprehensive Guide to Building Real-World NLP Systems*. O'Reilly Media.

Moschona, D. S. (2020). An Affective Service based on Multi-Modal Emotion Recognition, using EEG enabled Emotion Tracking and Speech Emotion Recognition. *2020 IEEE International Conference on Consumer Electronics - Asia (ICCE-Asia)*, 1-3. [10.1109/ICCE-Asia49877.2020.9277291](https://doi.org/10.1109/ICCE-Asia49877.2020.9277291).

Poria, S., Cambria, E., Bajpai, R., & Hussain, A. (2017). A review of affective computing: From unimodal analysis to multimodal fusion. *Information Fusion*, 37, 98–125. <https://doi.org/10.1016/j.inffus.2017.02.003>

Rodriguez, G., Jofré, N., Alvarado, Y., Fernandez, J., & Guerrero, R. (2021, Abril 15). INTERACCIÓN HUMANO-COMPUTADORA: ESTRATEGIAS HACIA LA EMPATÍA COMPUTACIONAL. *XXIII Workshop de Investigadores en Ciencias de la*

Computación, 270 - 274. ISBN: 978-987-24611-3-3

Shamim Hossain, M., & Muhammad, G. (2019, September). Emotion recognition using deep learning approach from audio–visual emotional big data. *Information Fusion*, 49, 69-78. 10.1016

Wang, Y., Wu, J., & Hoashi, K. (2019). Multi-Attention Fusion Network for Video-based Emotion Recognition. *International Conference on Multimodal Interaction (ICMI '19)*, 595–601. <https://doi.org/10.1145/3340555.3355720>