



FACULTAD DE CIENCIAS NATURALES Y EXACTAS
DEPARTAMENTO DE CIENCIA DE LA
COMPUTACIÓN

**Trabajo de Diploma en
Opción al Título de
Licenciado en Ciencia de la
Computación**

Título:
Aprendizaje Profundo para el
Perfilado de Usuarios en Redes
Sociales

Autor: Roberto Labadie Tamyao

Tutores: Dr. Daniel Castro Castro

M.Sc. Reynier Ortega Bueno

Curso 2020-2021

Contenido

Introducción	I
Problemática	3
Hipótesis	3
Objetivos	4

Introducción

Internet se define como la interconexión mundial de redes individuales operadas por el gobierno, la industria, el mundo académico y las partes privadas. En cuestión de muy pocos años, Internet se consolidó como una plataforma muy poderosa que ha cambiado para siempre la forma de comunicarse. Esta manera tan simple y accesible de compartir datos, ha propiciado un desplazamiento de los entes sociales hacia el uso casi exclusivo de este medio.

Un rol fundamental dentro de este universo comunicativo lo juegan las redes sociales donde de acuerdo al sitio DataReportal¹ a finales de 2020 existían más de 3.8 mil millones de usuarios de los 4.5 mil millones de personas conectadas a Internet, lo cual implica un crecimiento desmedido de la cantidad de datos multimodales que se genera a diario.

El mundo del Big Data(Riahi and Riahi, 2018) en el cual nos encontramos inmersos reporta una situación beneficiosa para el desarrollo de procesos sociales, en cuanto a la cantidad de información brindada por los usuarios . Estos procesos van desde estudios de marketing donde la retroalimentación a partir de opiniones permiten dirigir de manera efectiva la promoción hacia grupos sociales específicos, hasta aplicaciones en el área de la Interacción Humano-Computadora (*Human-Computer Interaction* HCI), donde el conocimiento de las características físicas y psicológicas de las personas permite personalizar la interfaz de comunicación. Sin embargo un cúmulo de información tan significativo, se hace imposible de tratar en un espacio de tiempo razonable a menos que se haga de manera automatizada.

Diversos estudios se han dirigido al diseño de métodos para tratar tal cantidad de información y realizar inferencias a partir de la misma, a la vez que han contribuido al desarrollo en ramas tales como el Procesamiento del Lenguaje Natural (*Natural Language Processing* NLP) y Visión por Computadoras (*Computer Vision* CV) en el área de la Inteligencia Artificial (*Artificial Intelligence* AI).

Dentro del NLP, la tarea de Perfilado de Autores (*Author Profiling* AP) (Rosso et al., 2019; Rosso and Rangel Pardo, 2020) se encarga específicamente de realizar un análisis de la información textual elaborada por una persona que permita establecer atributos y patrones de comportamiento para caracterizarla en cuanto sexo, rango de edades o rasgos personales (e.g si la persona es extrovertida o no, ideología política). Sin embargo, para el AP el hecho de que la información recuperada de estos textos varíe enormemente en términos de su formato aun cuando proviene de la misma persona, sumado a que las secuencias de textuales constituyen información digital no estructurada, hacen desafiante el proceso de analizarla y clasificarla automáticamente.

Inicialmente este tipo de tareas se desarrollaban sobre contenido generado en textos formales, periódicos, cartas o revistas, sin embargo determinar el perfil de una persona mediante el análisis de su cuenta en una red social ha tomado un gran auge en los últimos años (Rangel

¹<https://datareportal.com/>

et al., 2018; Rangel and Rosso, 2019; Cimino et al., 2020)

Las redes sociales además de haber logrado acelerar de comunicación entre las personas así como reunir varias formas del pensamiento individual en un mismo espacio, se han convertido en un medio donde se desarrollan sucesos negativos tales como, la divulgación de discursos de odio (), *bullying* o de noticias falsas (Rangel et al., 2020). Este hecho introduce nuevos retos al AP con tareas que involucran la detección de características altamente subjetivas como la ofensa, la toxicidad en el lenguaje y otros patrones psicológicos de comunicación que hacen más complejo el perfilado en relación a tareas tradicionales.

La mayor parte de los avances en el desarrollo de sistemas de AP han sentado sus bases dentro del ambiente académico, reuniendo estudios tanto de la Ciencia de la Computación como de la Lingüística. Algunos de los foros más importantes donde se han compartido tareas de este tipo son PAN² y EVALITA³ los que han dirigido últimamente al análisis del genero textual de micro-blogging empleado en medios sociales como Twitter⁴ con un enfoque multilingüe.

Tradicionalmente se han empleado dos tipos de acercamientos que han probado empíricamente ser efectivos para tratar el AP en redes sociales: los *basados en estilo* y los *basados en contenido*. Las propuestas basadas en estilo se refieren al hecho de analizar como los autores se expresan cuando escriben, en cambio la basada en contenido se apoyan en el área temática del texto analizado. La mayor contribución de varios trabajos se ha basado en la selección de atributos que permitan medir el estilo autor y el contenido simultáneamente mediante el empleo de métodos de *Machine Learning* (ML).

Un proceso fundamental en este tipo de propuestas es la selección de rasgos para representar vectorialmente los textos, este a su vez es un punto crítico en el cual pueden ser eliminados elementos que ayuden al modelo a discernir entre una clase u otra, en el caso de tareas de clasificación, pero también se puede introducir información ruidosa para el modelo.

Con relación a los métodos de ML, la mayoría de los trabajos desarrollados emplean métodos tradicionales como Regresión Logística (Logistic Regression LR) (Valencia-Valencia et al., 2019), Máquina de Vectores de Soporte (*Support Vector Machines* SVM) (Pizarro, 2019) y Bosques Aleatorios (*Random Forest* RF) (Johansson, 2019). En estos trabajos la información textual ha sido representada a través de rasgos (*features* de ahora por delante) léxicos y a nivel de carácter, empleando modelaciones con n-gramas y de representación de Bolsas de Palabras (*Bag of Words* BoW) (Pizarro, 2019; Valencia-Valencia et al., 2019). Otros trabajos incluyen features de estilo contruidos a partir del análisis del uso de letras mayúsculas, signos de puntuación, preposiciones, adverbios entre otros y modelos más recientes incluyen el uso de *embeddings* de palabras (Joo and Hwang, 2019; López-Santillán et al., 2019).

En los últimos años dentro del Machine Learning ha tenido un auge enorme el Aprendizaje Profundo (*Deep Learning* DL) en las áreas de CV y NLP, estableciendo nuevos estados del arte en la mayoría de sus tareas (Alom et al., 2019). Este auge estuvo condicionado por las capacidades de procesamiento de las nuevas computadoras y la cantidad de datos disponibles. Los modelos de DL han mostrado una gran habilidad para aprender representaciones de features con un alto nivel de abstracción que van más allá de lo actualmente interpretable por los seres humanos. Estas representaciones captan elementos que son posiblemente omitidos mediante la extracción manual y que facilitan proceso de inferencia de los propios modelos de

²<https://pan.webis.de/>

³<http://www.evalita.it/>

⁴<https://twitter.com/>

DL y de los métodos más tradicionales de ML.

Dentro del AP también se ha introducido el Aprendizaje Profundo con resultados alentadores, haciendo uso principalmente de Redes Neuronales Convolucionales (CNN) (Petrík and Chudá, 2019), Redes Neuronales Recurrentes (RNN) (Dias and Paraboni, 2019) y más recientemente se han empleados mecanismos de atención y modelos pre-entrenados de arquitecturas *Transformers* (Iyer and Vosoughi, 2020). Sin embargo la mayoría de estos modelos alcanzan mejores resultados al analizar la tarea a la que se orienta el perfilado cuando se emplean para clasificar mensajes individuales en lugar del perfil completo, e.g., se desempeñan mejor en la tarea de detección de odio en un tweet que en la detección de un perfil que tiende a usar un discurso de odio.

Por otro lado, las necesidades que cubren las tareas de AP van más allá del idioma en el que la persona postee un mensaje en una red social o escriba una carta, es por ello que urge la extensión de los modelos de AI propuestos al esquema multilingüe que predomina en redes sociales como Twitter. Este trabajo de tesis se enmarca en el perfilado de autores en redes sociales empleando técnicas de Aprendizaje Profundo para el modelado y la clasificación de los perfiles teniendo en cuenta el enfoque multilingüe propio de este medio.

Problemática

La mayoría de los trabajos existentes empleando DL para resolver tareas de perfilado de autores en redes sociales, tanto tradicionales (e.g., determinar rango de edades, sexo, etc.) como las más recientes (e.g. detección de esparcidos de discurso odio en redes sociales o divulgadores de noticias falsas) ven afectado su desempeño al tratar con las secuencias muy largas que se pueden generar al analizar un perfil completo a la vez, en vez de mensaje a mensaje. La pérdida de información (*information vanishing*) en este tipo de secuencias en términos de relaciones a largo plazo que presentan las estructuras sintácticas del lenguaje, es la causa fundamental de esta dificultad. Por esta razón sería conveniente dividir el problema de “clasificar un perfil” en subproblemas y construir una arquitectura modular, en la que primero se modele el perfil y luego este sea clasificado.

Por otro lado en Cuba existe un limitado tratamiento y aplicación de este tipo de métodos automatizados para llevar acabo tareas de impacto como las mencionadas, ya sea en el área forense o empresarial .

Hipótesis

Dividir el proceso de clasificación de un perfil dentro de una tarea en: (1) codificar los mensajes individualmente y (2) modelar-clasificar el perfil puede ayudar a prevenir la pérdida de información generada del análisis de largas secuencias. Esto sumado a la capacidad de representar la información no estructurada de los métodos de DL puede influir positivamente la precisión de las predicciones. Además proponer un modelo lo suficientemente robusto, podría ampliar camino para el empleo de esos métodos automatizados de AP en nuestro país.

Objetivos

Objetivo General

Diseñar una arquitectura de Aprendizaje Profundo modular para resolver la tarea de Perfilado de Autores en redes sociales bajo diversas subtarear teniendo en cuenta un enfoque multilingüe.

Objetivos Específicos

1. Diseñar una arquitectura que capture rasgos abstractos que permitan clasificar un perfil de usuario de una red social atendiendo tanto a tareas relacionadas con características demográficas de los autores, como con aspectos psicológicos de los mismos.
2. Extender la arquitectura propuesta a un enfoque multilingüe de las tareas evaluadas.
3. Evaluar y analizar el empleo de arquitecturas modulares basadas en Aprendizaje profundo sobre colecciones de datos propuestas en tareas de AP compartidas en las plataformas PAN (2020, 2021) y EVALITA (2020).

Estructura del Trabajo

Este trabajo esta estructurado en tres capítulos además del introductorio y una ultima sección en la que se describen las conclusiones de las modelaciones y experimentos realizados y se proponen caminos a seguir para trabajos futuros. Cada uno de los capítulos se listan a continuación:

- En el Capitulo 1 se presentan conceptos básicos relacionados con el Machine Learning y que están relacionados con el desarrollo de este trabajo.
- El Capitulo 2 describe las tareas de AP en las que se evaluarán las arquitecturas propuestas, así como los datos anotados sobre los cuales se basan los procesos de entrenamiento y evaluación.
- En el capitulo 3 se exponen los experimentos realizados en el proceso de ajuste de cada uno de los modelos y sus módulos sobre los lenguajes estudiados.

Referencias Bibliográficas

- Md Zahangir Alom, Tarek M. Taha, Chris Yakopcic, Stefan Westberg, Paheding Sidike, Mst Shamima Nasrin, Mahmudul Hasan, Brian C. Van Essen, Abdul A. S. Awwal, and Vijayan K. Asari. 2019. [A State-of-the-Art Survey on Deep Learning Theory and Architectures](#). *Electronics*, 8(3).
- Andrea Cimino, Felice Dell’Orletta, and Malvina Nissim. 2020. TAG-it@ EVALITA 2020: Overview of the Topic, Age, and Gender Prediction Task for Italian. In *Proceedings of the 7th evaluation campaign of Natural Language Processing and Speech tools for Italian (EVALITA 2020)*, Online. CEUR. org. CEUR Workshop Proceedings (CEUR-WS.org). Evaluation Campaign of Natural Language Processing and Speech Tools for Italian, EVALITA 2020 ; Conference date: 17-12-2020.
- Rafael Felipe Sandroni Dias and Ivandré Paraboni. 2019. [Combined CNN+RNN bot and gender profiling](#). In *Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, Lugano, Switzerland, September 9-12, 2019*, volume 2380 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Aarish Iyer and Soroush Vosoughi. 2020. [Style Change Detection Using BERT—Notebook for PAN at CLEF 2020](#). In *CLEF 2020 Labs and Workshops, Notebook Papers*. CEUR-WS.org.
- Fredrik Johansson. 2019. [Supervised Classification of Twitter Accounts Based on Textual Content of Tweets](#). In *Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, Lugano, Switzerland, September 9-12, 2019*, volume 2380 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Youngjun Joo and Incheon Hwang. 2019. [Author Profiling on Social Media: An Ensemble Learning Approach using Various Features](#). In *Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, Lugano, Switzerland, September 9-12, 2019*, volume 2380 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Jesus López-Santillán, Luis Carlos González-Gurrola, Manuel Montes-y-Gómez, Graciela Ramírez Alonso, and Olanda Prieto-Ordaz. 2019. [An Evolutionary Approach to Build User Representations for Profiling of Bots and Humans in Twitter](#). In *Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, Lugano, Switzerland, September 9-12, 2019*, volume 2380 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Juraj Petrík and Daniela Chudá. 2019. [Bots and gender profiling with convolutional hierarchical recurrent neural network](#). In *Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, Lugano, Switzerland, September 9-12, 2019*, volume 2380 of *CEUR Workshop Proceedings*. CEUR-WS.org.

- Juan Pizarro. 2019. [Using N-grams to detect Bots on Twitter](#). In *Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, Lugano, Switzerland, September 9-12, 2019*, volume 2380 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- Francisco Rangel, Anastasia Giachanou, Bilal Ghanem, and Paolo Rosso. 2020. [Overview of the 8th Author Profiling Task at PAN 2020: Profiling Fake News Spreaders on Twitter](#). In *CLEF 2020 Labs and Workshops, Notebook Papers*. CEUR-WS.org.
- Francisco Rangel, Manuel Montes-y-Gómez, Martin Potthast, and Benno Stein. 2018. [Overview of the 6th Author Profiling Task at PAN 2018: Cross-domain Authorship Attribution and Style Change Detection](#). In *CLEF 2018 Evaluation Labs and Workshop – Working Notes Papers, 10-14 September, Avignon, France*. CEUR-WS.org.
- Francisco Rangel and Paolo Rosso. 2019. [Overview of the 7th Author Profiling Task at PAN 2019: Bots and Gender Profiling](#). In *CLEF 2019 Labs and Workshops, Notebook Papers*. CEUR-WS.org.
- Youssra Riahi and S. Riahi. 2018. Big Data and Big Data Analytics: Concepts, Types and Technologies. *International Journal of Research and Engineering*, 5:524–528.
- Paolo Rosso, Martin Potthast, Benno Stein, Efstathios Stamatatos, Francisco Rangel, and Walter Daelemans. 2019. [Evolution of the pan lab on digital text forensics](#). In Nicola Ferro and Carol Peters, editors, *Information Retrieval Evaluation in a Changing World: Lessons Learned from 20 Years of CLEF*, pages 461–485. Springer International Publishing, Cham.
- Paolo Rosso and Francisco Rangel Pardo. 2020. [Author Profiling Tracks at FIRE](#). *FIRE 10th Anniversary, SN Computer Science*, 1.
- Alex I. Valencia-Valencia, Helena Gómez-Adorno, Christopher Stephens Rhodes, and Gibran Fuentes Pineda. 2019. [Bots and Gender Identification Based on Stylometry of Tweet Minimal Structure and n-grams Model](#). In *Working Notes of CLEF 2019 - Conference and Labs of the Evaluation Forum, Lugano, Switzerland, September 9-12, 2019*, volume 2380 of *CEUR Workshop Proceedings*. CEUR-WS.org.