

Detección automática de anunciantes en pautas publicitarias on-line

Guillermo Rodríguez¹, Leonardo Da Rocha Araujo¹, Santiago Vidal¹,

Claudia Marcos^{1,2}, Facundo Zabala³, Martin Caraoghlanian³

¹ ISISTAN (UNICEN-CONICET), Tandil, Buenos Aires

² CIC - Comisión de Investigaciones Científicas de la Provincia de Buenos Aires

{grodri,svidal,cmarcos}@exa.unicen.edu.ar

leonardo.araujo@isistan.unicen.edu.ar

³ adCuality, Ciudad Autónoma de Buenos Aires

{facundo,martin}@adcuality.net

Resumen. Ante el constante crecimiento de la inversión en publicidad online, es fundamental contar con herramientas que brinden un monitoreo completo, certero, verificado y en tiempo real de los anuncios que las empresas deciden colocar en diferentes sitios web y redes sociales para vender sus productos mediante campañas publicitarias. Por esta razón, la empresa adCuality provee una plataforma digital que permite acceder de forma fácil y segura a información estratégica para agencias de publicidad, medios, centrales de medios, y anunciantes, de modo que puedan verificar que sus campañas publicitarias se están realizando con éxito. Sin embargo, un desafío que hoy tiene adCuality es la falta de mecanismos semi-automáticos que les permitan identificar publicidad de sus propios clientes e inclusive detectar anunciantes de la competencia que invierten el dinero en otros medios. Hoy en día, esta identificación es realizada mediante una clasificación manual por la empresa y representa un costo muy importante en tiempo y esfuerzo. En este contexto, se estableció una vinculación entre el Instituto ISISTAN, la Fundación Sadosky y la empresa adCuality para desarrollar un prototipo que permita la automatización del proceso de clasificación y la optimización del flujo de datos relacionados con publicidad online de la plataforma de monitoreo en tiempo real.

1 Caracterización General del Proyecto

Esta sección presenta las instituciones y empresas participantes del proyecto, la descripción del proyecto, y finalmente, el tipo de interacción entre las partes.

1.1 Instituciones y Empresas Participantes

Instituto ISISTAN: El Instituto Superior de Ingeniería del Software de Tandil (ISISTAN) es una Unidad Ejecutora (UE) de doble dependencia UNICEN-CONICET. Los integrantes del proyecto poseen experiencia y conocimiento en

temáticas de: arquitecturas Web, refactoring de código fuente, machine learning, deep learning, y Big Data.

adCuality: adCuality (<https://adcuality.com>) es una empresa que proporciona servicios de visualización de datos, búsqueda, social media y monitorización. Existen 3 tipos de servicios: 1 - adCuality for Publishers; aumenta ventas de anuncios informando dónde los anunciantes gastan su dinero con gran precisión en tiempo real. 2 - adCuality for Agencies; maximiza el negocio con información estratégica. 3 - adCuality for Advertisers; ayuda a revelar la estrategia de marketing de los competidores. adCuality ha realizado proyectos de software para empresas nacionales e internacionales, particularmente en Brasil, Chile, Costa Rica, Uruguay y Estados Unidos, entre otros.

1.2 Descripción del proyecto:

El proyecto tiene una duración de 6 meses. Inicialmente, se llevaron a cabo reuniones entre ambas partes. Dichas reuniones estuvieron destinadas a definir la metodología de trabajo y a planificar las actividades a desarrollar. Las actividades acordadas fueron: (i) desarrollar un prototipo para clasificar automáticamente la pauta publicitaria capturada; (ii) medir precisión, y ahorro de costos y tiempos del prototipo de clasificación, desarrollado por el equipo de ISISTAN, sobre un conjunto de pauta publicitaria capturada por adCuality.

El grupo de investigación está formado por 4 integrantes del ISISTAN: 3 docentes-investigadores - Dra. Claudia Marcos, Dr. Santiago Vidal, y el Dr. Guillermo Rodríguez; y por un becario doctoral – Leonardo Da Rocha Araujo. Por el lado de adCuality, el grupo está integrado por 2 personas: Facundo Zabala (Product Manager) y Martin Caraoghlanian (CEO adCuality).

El objetivo general del proyecto es desarrollar un prototipo que permita la automatización del proceso de clasificación y la optimización del flujo de datos relacionados con publicidad online de la plataforma de monitoreo en tiempo real de adCuality. Los objetivos específicos que se identificaron son: (i) investigar técnicas de Deep Learning que permitan clasificar automáticamente pautas publicitarias; (ii) y diseñar e implementar un prototipo basado en dichas técnicas.

1.3 Tipo de Interacción

La interacción entre ambas partes es de diferente índole: colaboración en I+D, asistencia técnica de investigadores a empresas, y disponibilización de un prototipo a la empresa.

2 Detalles de Ejecución del Proyecto

Esta sección describe las actividades realizadas, hitos y cumplimiento de los principales objetivos. Finalmente, se detallan los principales desafíos encontrados.

2.1 Actividades Realizadas

Hasta el momento, se ha avanzado con la realización de las siguientes actividades:

- *Actividad 1:* Definir y modelar el problema de clasificación automática de pauta publicitaria como un problema de Deep Learning.
- *Actividad 2:* Diseñar e implementar un prototipo de solución, basada en técnicas básicas de Deep Learning, para la clasificación automática de pauta publicitaria.
- *Actividad 3:* Realizar una evaluación inicial de la eficiencia del prototipo en un ambiente de prueba con pauta publicitaria capturada por adCuality en una ventana de tiempo de 30 días.

2.2 Hitos y Cumplimiento de los Objetivos

Se propusieron los siguientes hitos con sus entregables:

- **Etapa 1:** Informe técnico, y diseño de un prototipo.
- **Etapa 2:** Prototipo de solución, validación con pruebas de concepto.
- **Etapa 3:** Herramienta final, validación con datos reales e informe.

2.3 Principales Desafíos

El problema de clasificación automática de publicidad no presenta una solución trivial. En primer lugar, la pauta clasificada manualmente es escasa en proporción a la pauta capturada. En ese sentido, es necesario contar con un etiquetado manual de pauta para la implementación de las técnicas elegidas, que insume tiempo y costo. En segundo lugar, es muy importante para la empresa identificar para cada publicidad, el anunciante que la originó (ej.: Garbarino) y el producto que se publicita (ej.: heladera). En este sentido, el desafío consiste en detectar el producto de la pauta.

3 Resultados del Proyecto

Esta sección resume los resultados obtenidos en cada actividad. Cabe aclarar que el proyecto aún se encuentra en desarrollo.

3.1 Resultados de cada Actividad

- *Resultados de Actividad 1:* Se definió una metodología de trabajo. Se analizaron técnicas existentes en la literatura para el procesamiento de imágenes, como redes neuronales convolucionales (CNN) [1], Redes Neuronales Generativas Adversarias (GANs) [2] y Redes Neuronales Recurrentes (RNN) [3].
- *Resultados de Actividad 2:* Se modeló el problema y se seleccionó un conjunto de técnicas básicas de Deep Learning candidatas para resolverlo. Particularmente, se utilizó la red pre-entrenada de Facebook conocida como Retinanet, combinado con las técnicas de correlación cruzada normalizada y *template matching* [4].

- *Resultados de Actividad 3:* Se evaluó el prototipo en condiciones de laboratorio. Como hay poca pauta clasificada manualmente en proporción a lo que se captura, se obtuvo una precisión del 85.18%, la cual no es suficiente para la empresa. En base al feedback obtenido, se están analizando técnicas de Deep Learning más específicas para nuestro problema como *one-shot learning* [5] y *few-shot learning* [6] para poder aumentar la precisión.

3.2 Evaluación de los Resultados y Lecciones Aprendidas

El prototipo desarrollado para la clasificación automática de la pauta publicitaria es un primer paso que contribuye a la optimización de los procesos de validación de pauta de adCuality. A partir de un prototipo base, se plantea una fase posterior de ajuste del mismo. La perspectiva a largo plazo es poder integrar el prototipo a la aplicación Web que tiene actualmente la empresa. En este sentido, la empresa posee un objetivo estratégico de mejorar sus servicios a través de la validación automática de pauta publicitaria.

4 Balance y Conclusiones

La vinculación entre miembros del ISISTAN y adCuality se está desarrollando de una manera muy positiva y enriquecedora. La comunicación entre ambas partes es fluida, interactuando a medida que se obtienen avances en el proyecto, o cuando se encuentra algún impedimento, o es necesario la toma de decisiones sobre diferentes alternativas. Los primeros resultados obtenidos en términos de precisión son prometedores, pero aún quedan estrategias para explorar y alcanzar valores de precisión que sean competitivos para la empresa.

References

1. Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep convolutional neural networks." *Advances in neural information processing systems* 25 (2012): 1097-1105.
2. Goodfellow, Ian; Pouget-Abadie, Jean; Mirza, Mehdi; Xu, Bing; Warde-Farley, David; Ozair, Sherjil; Courville, Aaron; Bengio, Yoshua (2014). *Generative Adversarial Networks*. *Proceedings of the International Conference on Neural Information Processing Systems (NIPS 2014)*. pp. 2672–2680.
3. Schuster, Mike, and Kuldip K. Paliwal. "Bidirectional recurrent neural networks." *IEEE transactions on Signal Processing* 45.11 (1997): 2673-2681.
4. Jiao, L., & Zhao, J. (2019). A survey on the new generation of deep learning in image processing. *IEEE Access*, 7, 172231-172263.
5. Liu, C., Xu, C., Wang, Y., Zhang, L., & Fu, Y. (2020). An Embarrassingly Simple Baseline to One-shot Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (pp. 922-923).
6. Larochelle, H. (2020). *Few-Shot Learning*. *Computer Vision: A Reference Guide*, 1-4. Springer.