

Extracción de información de Historias Clínicas Electrónicas escritas en español para realizar inteligencia epidémica

Javier Petri¹, Pilar Barcena Barbeira², and Viviana Cotik^{1,3}

¹ Departamento de Computación, FCEyN, Universidad de Buenos Aires, Argentina

² Programa de Innovación Tecnológica en Salud Pública, Departamento de Salud Pública, Facultad de Medicina, UBA, Argentina

³ Instituto de Investigación en Ciencias de la Computación (ICC), CONICET-UBA, Argentina {vcotik}@dc.uba.ar

Abstract. La detección automática de síntomas a partir de historias clínicas electrónicas es una fuente valiosa para los sistemas de vigilancia basados en eventos. En este estudio, desarrollamos herramientas para detectar automáticamente síntomas asociados a enfermedades febriles en historias clínicas electrónicas redactadas en español. Para ello, utilizamos un corpus personalizado que incluye 6.228 reportes de salud anotados por expertos y aproximadamente 1 millón de reportes no anotados. Nuestra estrategia consistió en ajustar modelos de reconocimiento de entidades nombradas de última generación, incluyendo modelos BiLSTM-CRF y modelos basados en transformers como RoBERTa. Nos enfocamos en modelos adaptados al dominio y a la tarea para mejorar el rendimiento: los primeros fueron pre-entrenados en corpus biomédicos, mientras que los segundos fueron pre-entrenados adicionalmente en nuestros reportes de salud no anotados. A pesar de las limitaciones computacionales, nuestros modelos demostraron resultados prometedores. En particular, RoBERTa-Clinico, un modelo basado en transformers adaptado a la tarea y pre-entrenado sobre nuestro corpus no anotado, obtuvo el mejor rendimiento para el micro recall (79,30) y un F1 micro de 70,83, cifras comparables a estudios similares. De este modo, contribuimos al limitado cuerpo de trabajos en BioNLP en español.

Keywords: Reconocimiento de entidades nombradas · BioNLP · historias clínicas electrónicas en español · detección automática de síntomas · vigilancia basada en eventos.

Publicado en: Advances in Artificial Intelligence – IBERAMIA 2024: 18th Ibero-American Conference on AI, Montevideo, Uruguay, November 13–15, 2024, Proceedings. Springer-Verlag, Berlin, Heidelberg, 424-436.
https://doi.org/10.1007/978-3-031-80366-6_35