

ISMR - Sistema basado en Deep Learning para la transcripción y extracción de conocimiento en entrevistas médico-paciente

Deep Learning-based system for transcription and knowledge extraction from doctor-patient interviews

Pedro José Vivancos-Vicente¹, José Antonio García-Díaz², Fernando Molina¹,
Juan Salvador Castejón-Garrido¹, Rafael Valencia-García²

¹VÓCALI Sistemas Inteligentes S.L., Parque Científico de Murcia,

Carretera de Madrid km 388. Complejo de Espinardo, 30100 Murcia, España

²Facultad de Informática, Universidad de Murcia, Campus de Espinardo, 30100 Murcia, España

{pedro.vivancos, fernando.molina, juans.castejon}@vocali.net

{joseantonio.garcia8, valencia}@um.es

Resumen: Las entrevistas médico-paciente suponen una actividad clave en la medicina donde el médico y el paciente intercambian información clínica que debe registrarse dentro la historia clínica electrónica (HCE). Sin embargo, esta tarea se ve complicada por el exceso de trabajo administrativo que ahora el médico tiene que realizar delante del ordenador. Este proyecto financiado por el CDTI con fondos FEDER consiste en la creación de una plataforma que asista a los médicos y pacientes durante las consultas y que permita a los médicos poder centrarse en el paciente y ofrecerle un trato más personal y exhaustivo mientras que la entrevista está siendo transcrita y procesada de manera inteligente y automática para su posterior consulta. Además, permitirá la extracción de conocimiento médico que permitirá facilitar la inserción de episodios activos en el paciente, la prescripción de medicamentos y la generación de volantes para pruebas diagnósticas.

Palabras clave: Reconocimiento de voz, extracción de conocimiento, reconocimiento de entidades

Abstract: Doctor-patient interviews are a key activity in medicine, in which doctors and patients exchange clinical information that must be recorded in the electronic medical record (EHR). However, the excessive administrative work required to fill out the EHRs takes away valuable time that doctors could spend examining patients. This project, financed by the CDTI with FEDER funds, consists of the creation of a platform that assists doctors and patients during doctor-patient interviews. The aim is to allow doctors to focus on the patient and offer a more personal and comprehensive treatment while the interview is being recorded. Then, the transcription of the interview is processed by extracting medical knowledge that will facilitate the insertion of active episodes in the patient, the prescription of drugs and the generation of flyers for diagnostic tests.

Keywords: Speech recognition, knowledge extraction, named entities recognition

1 Introducción

En España, la demora causada por las abultadas listas de espera y la masificación de algunos servicios se considera un problema relevante, causado por reducciones en el gasto sanitario, el incremento y envejecimiento de la población y últimamente por la crisis del COVID-19. Según el Sistema de Información de Listas de Espera de 2020, un total de 691.508 pacientes se encuentran en un estado

de espera estructural con un tiempo medio de espera de 170 días¹ (datos que son superiores al estudio realizado en 2019). Además, desde la implantación de la historia clínica electrónica (HCE), los médicos se quejan de un exceso de trabajo administrativo, que conlleva a una disminución de la productividad.

¹<https://www.mscbs.gob.es/estadEstudios/estadisticas/inforRecopilaciones/listaEspera.htm>

El médico es quien tiene que realizar todo el proceso interactuando con varios sistemas informáticos diferentes. Un problema añadido es la gran cantidad de protocolos en cada especialidad que lleva a que las interfaces de usuario de este tipo de sistemas sean complejas.

El objetivo principal de este proyecto es la creación una plataforma de transcripción y extracción de conocimiento en entrevistas médico-paciente que asista al médico durante todo el proceso. Gracias a este sistema, el médico podrá entablar conversación con el paciente y/o sus acompañantes despreocupándose en buena medida de la recogida manual de datos. El sistema será capaz de identificar a cada hablante y registrar el audio con marcas de tiempo de cada uno por separado gracias a un dispositivo con un array de micrófonos. Conforme se realiza la transcripción, se utilizarán tecnologías de extracción de conocimiento y de anotación semántica en tiempo real para estructurar el texto escrito en lenguaje natural. Así, este sistema detectará síntomas, medicamentos, enfermedades, episodios anteriores y otros elementos interesantes a ser detectados y registrados durante la consulta. Gracias a esta detección se incorporarán funciones basadas en estándares de interoperabilidad de información clínica como HL7 (Health Level Seven International) que permitan comunicarse con los sistemas de información médicos para poder registrar episodios activos del paciente, generar una receta o un volante para una prueba y así facilitar el trabajo del médico.

2 Arquitectura del sistema

La arquitectura del sistema propuesto se puede observar en la figura 1. Este sistema está dividido en dos grandes componentes. En primer lugar, está el dispositivo de captación de conversaciones, y por otro lado, el servidor, el cuál se subdivide a su vez en tres módulos: el (1) Sistema de transcripción automática de entrevistas médico paciente, el (2) sistema de extracción de conocimiento médico y, el (3) cuadro de mandos (o interfaz de usuario).

A continuación, se detalla brevemente cada uno de estos componentes.

2.1 Dispositivo de captación de entrevistas médico paciente

Este módulo se basa en un dispositivo de captación de conversaciones de entrevistas médi-

co paciente. La arquitectura de este dispositivo está formada por una capa hardware y una capa software. En primer lugar, la capa hardware está formada por un micrófono de array y por un mini-computador que se encargan de la captación de audio de alta calidad. Por un lado, el micrófono de array está dotado de distintos micrófonos para la captura de la señal en tiempo real con calidad y para poder realizar una diarización precisa en entornos con múltiples hablantes. Este dispositivo cuenta con algoritmos específicos para limpiar el audio y para atenuar la reverberación de la sala. Por otro lado, el mini-computador se encarga de ejecutar las capas software, de almacenar los ficheros de audio y de poder comunicarse con servicios externos a través de Internet. En segundo lugar, la capa software dispone de una interfaz por la cual se puede manejar el dispositivo así como los módulos y modelos de transcripción de audio y otros servicios necesarios.

2.2 Servidor

La parte del servidor se encarga de centralizar los procesos necesarios para capturar, transcribir y anotar el audio recogido por el dispositivo de captación. Es importante destacar que, aunque el dispositivo estará dotado del hardware y software suficiente para poder capturar y transcribir la conversación para dar retroalimentación en tiempo real, en el servidor se despliegan recursos más eficaces que requieren un mayor procesamiento computacional. Entre estos servicios se pueden destacar la corrección de la transcripción, la extracción del conocimiento médico y proporcionar una interfaz donde pacientes y médicos puedan acceder y revisar el contenido de las entrevistas.

El servidor está dividido en tres módulos principales: (1) el sistema de transcripción de entrevistas médico-paciente, (2) el sistema de extracción de conocimiento médico, y (3) el cuadro de mandos. Estos módulos se detallan a continuación en las siguientes subsecciones.

2.2.1 Sistema de transcripción

El sistema de transcripción parte de trabajos anteriores de la empresa (Vivancos-Vicente et al., 2020) y es capaz de aplicar un proceso de transcripción automática a documentos de audio con distintos hablantes y sin la necesidad de entrenamiento específico. Estas técnicas están basadas en tecnologías de deep learning. Este módulo devuelve un fichero en for-

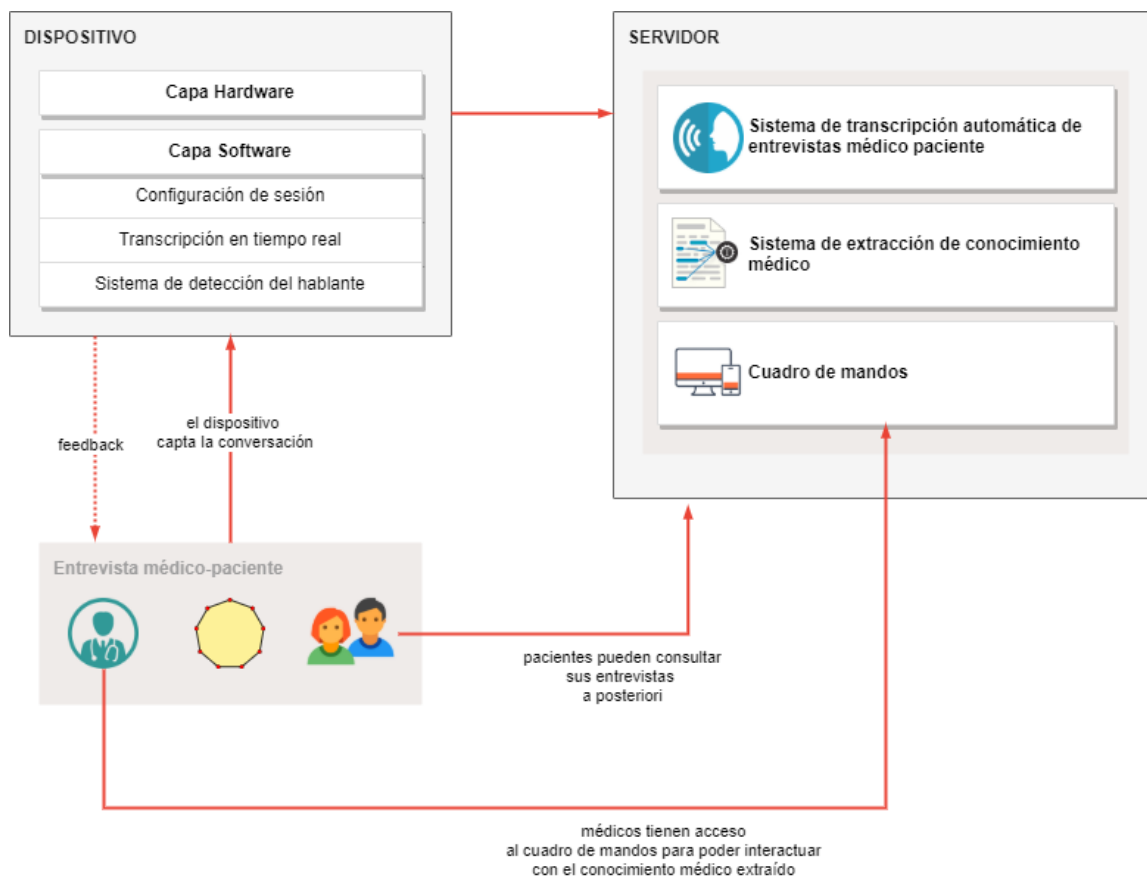


Figura 1: Arquitectura del sistema

mato WebVTT (Web Video Text Tracks Format), que es un formato estándar de subtítulos de vídeos en la web. Como novedad importante, este módulo incluye tecnologías de deep learning para la recuperación de puntuación en las transcripciones obtenidas pudiendo introducir signos de puntuación como comas, puntos e interrogaciones. Este módulo está basado en trabajos anteriores del grupo de investigación (Bravo-Candel et al., 2021).

2.2.2 Sistema de extracción de conocimiento

Este sistema es capaz de extraer y anotar texto escrito en lenguaje natural con conceptos médicos a partir del uso de ontologías y tecnologías de reconocimiento de entidades y anotación semántica. De las entidades se extraen conceptos y acciones, además de predicados y referencias para unir los conceptos anteriores. Este sistema es capaz de identificar relaciones básicas como subtipos, de igualdad, de implicación, de pertenencia así como otras relaciones para delimitar el marco contextual y temporal. Para ello, se están evaluando modelos basados en redes

neuronales recurrentes bidireccionales del tipo Gated Recurrent Units (BiGRU) a partir de conjuntos de vectores entrenados con la versión española de BERT (Cañete et al., 2020). El resultado es un sistema de anotaciones en formato BRAT. Este módulo se basa en trabajos anteriores del grupo de investigación (Rodríguez-García et al., 2014), (García-Díaz, Cánovas-García, y Valencia-García, 2020).

La información extraída se puede exportar a formatos estándar para facilitar la interoperabilidad con sistemas externos basados en HIS/HCE. Cabe destacar que no es lo mismo extraer conocimiento médico de un informe técnico que de una conversación médico-paciente, aunque ésta sea de carácter formal. En primer lugar, porque el significado puede variar según hable el médico o el paciente y, en segundo lugar, aunque aparezca un determinado término en la conversación, como podría ser un síntoma o una enfermedad, éste no quiere decir que el paciente lo sufra. Estos aspectos hacen necesario que este sistema de extracción de conocimiento médico sea capaz de entender conceptos temporales co-

mo eventos pasados, presentes o futuros, de forma que pueda registrar, de manera más o menos precisa, la temporalización de síntomas, enfermedades o causas, entre otros. Para ello, se han empleado reconocedores de entidades, corpus anotados como los analizados en (Jiménez-Zafra et al., 2020) y el uso lexicones.

2.2.3 Cuadro de mandos

Este módulo consiste en una plataforma software que permite la gestión del conocimiento generado por el sistema por parte de los médicos y pacientes. Los médicos como los pacientes disponen de un dashboard, que permite hacer uso del conocimiento adquirido automáticamente. El médico puede realizar consultas para localizar partes de la conversación a partir de la transcripción de la conversación, buscando coincidencias fonológicamente similares. Una vez accede a una transcripción, el médico puede consultar la conversación y realizar anotaciones sobre los conceptos identificados. Estos conceptos son accionables y permiten, según la naturaleza del concepto, generar documentos en HL7 para la integración con sistemas externos. Esto permitirá, por ejemplo, obtener información para registrar episodios activos del paciente en el momento de la consulta, generar la información de recetas a partir de la prescripción médica o generar la información para emitir volantes para pruebas diagnósticas.

3 Trabajo futuro

El proyecto se encuentra en la última anualidad del proyecto. Hasta el momento se ha desarrollado el dispositivo de captación de entrevistas y el sistema de transcripción. Las tecnologías que han dado buenos resultados para la transcripción se basan en modelos DNN-HMM (modelos ocultos de Markov con redes neuronales profundas) (Ravanelli y Omologo, 2017).

El sistema de extracción de conocimiento está todavía en desarrollo y se están generando distintos recursos y herramientas que permitan la identificación y extracción del conocimiento dentro de la transcripción de la entrevista. Por otro lado, se está estudiando distintos estándares para interoperabilidad clínica basados en HL7 para, entre otras cosas, la inserción de información dentro de la historia clínica del paciente, la generación semi-automática de recetas y de volantes de pruebas médicas. Por último, se realizará el

dashboard final así como la integración de los módulos en un prototipo final de la plataforma global. En este sentido, se planificarán distintas pruebas de campo para comprobar la calidad del sistema completo, y así como se estudiará la integración del sistema con sistemas de información externos.

Agradecimientos

Este proyecto ha sido financiado por el CDTI con fondos FEDER dentro del proyecto con referencia IDI-20200407

Bibliografía

- Bravo-Candel, D., J. López-Hernández, J. A. García-Díaz, F. Molina-Molina, y F. García-Sánchez. 2021. Automatic correction of real-word errors in spanish clinical texts. *Sensors*, 21(9).
- Cañete, J., G. Chaperon, R. Fuentes, J.-H. Ho, H. Kang, y J. Pérez. 2020. Spanish pre-trained bert model and evaluation data. En *PML4DC at ICLR 2020*.
- García-Díaz, J. A., M. Cánovas-García, y R. Valencia-García. 2020. Ontology-driven aspect-based sentiment analysis classification: An infodemiological case study regarding infectious diseases in latin america. *Future Gener. Comput. Syst.*, 112:641–657.
- Jiménez-Zafra, S. M., R. Morante, M. Teresa Martín-Valdivia, y L. A. Ureña-López. 2020. Corpora annotated with negation: An overview. *Computational Linguistics*, 46(1):1–52.
- Ravanelli, M. y M. Omologo. 2017. Contaminated speech training methods for robust dnn-hmm distant speech recognition. *arXiv preprint arXiv:1710.03538*.
- Rodríguez-García, M. Á., R. Valencia-García, F. García-Sánchez, y J. J. S. Zapter. 2014. Ontology-based annotation and retrieval of services in the cloud. *Knowl. Based Syst.*, 56:15–25.
- Vivancos-Vicente, P. J., J. A. García-Díaz, Á. Almela, F. Molina, J. S. Castejón-Garrido, y R. Valencia-García. 2020. Transcripción, indexación y análisis automático de declaraciones judiciales a partir de representaciones fonéticas y técnicas de lingüística forense. *Proces. del Leng. Natural*, 65:109–112.