



El aprendizaje profundo o deep learning que caracteriza a las redes neuronales se basa en aprender a partir de la experiencia. Foto: Freepik.

Inteligencia artificial para frenar una nueva ola del COVID-19 en los hospitales

El Hospital Clínic y el Barcelona Supercomputing Center trabajan en un proyecto impulsado por la Secretaría de Estado de Digitalización e Inteligencia Artificial para predecir la evolución de los pacientes de esta enfermedad

Noelia Carrión

¿Puede un ordenador adelantarse y saber cuántos enfermos hospitalizados tendrán que ingresar en la uci antes incluso de que el paciente presente síntomas? Aunque parezca difícil de creer, investigadores del Hospital Clínic de Barcelona y el Barcelona Supercomputing Center (BSC) ya han demostrado que es posible.

Con la vista puesta en una tercera ola del COVID-19, y en las futuras que puedan llegar, ambos equipos han iniciado un proyecto conjunto que permita a los médicos conocer cuál será la evolución del enfermo una vez ingresa en el hospital.

A partir de una fuente de datos de cerca de 3.000 pacientes del hospital y centenares de variables seleccionadas por los médicos del Clínic, el BSC desarrolla modelos predictivos basados en redes neuronales para determinar a tiempo tres posibles escenarios: el desenlace (alta o fallecimiento del paciente),

si es necesario ingreso en la uci o si se requiere asistencia respiratoria. De esta forma “puedes organizar mejor los recursos”, afirma Xavier Pastor, responsable de Informática Médica del Hospital Clínic de Barcelona.

El doctor Pastor recalca la angustia que pasaron los responsables del centro durante la etapa más aguda de la pandemia, llegando a formar un comité ético “para marcar qué pacientes eran candidatos para entrar en la uci y cuáles a pesar de necesitarlo clínicamente había que excluirlos por otras razones” e, incluso, montando hospitales de campaña.

De esta manera, podrían evitarse situaciones tan dramáticas como las que se vivieron al comienzo de esta crisis sanitaria, ya que el objetivo es conseguir una aplicación que el hospital pueda usar periódicamente con la información de los hospitalizados en ese momento y “tenga capacidad predictiva días antes del desenlace o del ingreso en uci”, se-

ñala Marta Villegas, corresponsable del área de minería de textos del BSC.

El text mining o minería de textos, clave para el modelo de predicción

Para desarrollar este modelo, el hospital facilita dos tipos de datos: datos estructurados y no estructurados. Los primeros hacen referencia a “la edad, el sexo del paciente, los resultados de laboratorio o los diagnósticos que se le han ido asignando al paciente a lo largo de su período de hospitalización”, añade la investigadora.

Sin embargo, el valor añadido de la iniciativa reside en los segundos, al tratarse de documentos generados por los propios profesionales del hospital, como son el curso clínico o el informe de alta del paciente. En este caso, el equipo del BSC aplica técnicas de minería de textos para sacar todo aquello que los facultativos consideran relevante como, por ejemplo, si el paciente es fumador o diabético e, incluso, la medicación que toma.



El BSC es el administrador de MareNostrum, el superordenador más potente de España y el tercero en Europa. En la imagen, ejemplo de un supercomputador. Foto: Pexels.

Una información que sin esta herramienta resultaría imposible de conseguir de manera automatizada. “Lo que hace el *text mining* es convertir un word en un excel, donde el excel tiene las variables que tú querías extraer sistemáticamente de ese word”, explica la investigadora del BSC. Una vez conseguido este paso, “lo puedes manipular, mientras que si está en un texto escrito solo puede subrayarlo”. Sería como convertir letras en números.

El equipo de Marta Villegas cuenta, de hecho, con una amplia experiencia aplicando estas técnicas de análisis y procesamiento de datos en el campo de la biomedicina. Entre los proyectos con los que colaboran se encuentra *ICTUS-net*, una iniciativa europea en la que participan hospitales de España, Francia y Portugal. La unidad de Text Mining del BSC lidera el desarrollo de herramientas de minería de textos para obtener información relevante que permita evaluar la calidad asistencial a partir de informes de alta hospitalaria de pacientes con esta enfermedad.

Sin embargo, en el caso de la iniciativa junto al Clínic, no es una tarea fácil porque “cada hospital a lo mejor ha usado una terminología diferente o las unidades de medida están expresadas de manera diferente”, señala la investigadora. Además, el hecho de contar con textos tanto en catalán como en español indistintamente, los errores sintácticos o el propio contexto añaden complejidad a la tarea. “Yo puedo identificar un concepto, pero este concepto lo puedo estar afirmando o negando, a mí me puede interesar seleccionar lo que estoy afirmando y no, por ejemplo, referencias al pasado”, señala Xavier Pastor.

El proyecto está promovido por la Secretaría de estado de digitalización e inteligencia artificial (Sedia), y cuenta con diversos planes que buscan impulsar estas técnicas de procesamientos de documentos “para que la industria de

las tecnologías de la lengua española y todas las lenguas oficiales pueda desarrollarse lo máximo posible”, apunta Marta Villegas. A pesar de que el español es la cuarta lengua más hablada en el mundo, la investigadora recuerda que “con el inglés puedes hacer vida digital, pero en los demás hay una diferencia abismal. Estamos a años luz”.

Por eso, la iniciativa pone a prueba todas estas nuevas tecnologías de análisis de datos con las que aún están familiarizándose los investigadores. En el hospital Clínic ya trabajaban con textos clínicos, pero, según el Dr. Pastor, ahora se ha dado un paso más porque no solo buscan diagnósticos, sino conceptos más complejos como “signos y síntomas clínicos que son previos. Nos estamos yendo a identificar situaciones concretas del paciente y eso es todo un reto”.

Entrenamiento con redes neuronales

Lo investigadores del BSC ya han creado un primer prototipo del algoritmo de predicción gracias a una fuente de datos en abierto con información de 2.000 pacientes del grupo HM Hospitales de Madrid. Pero ¿cómo se consigue desarrollar un modelo que pronostique la evolución de un paciente hospitalizado?

El método, aunque no lo parezca, es más sencillo de lo que uno podría imaginarse. A partir del conjunto total de los datos, “el 80% normalmente lo reservas para entrenar los modelos porque ya sabes para cada uno de estos pacientes cuál ha sido su desenlace. Entones entrenas el sistema y le vas pasando el paciente uno, desenlace uno y así la red neuronal va aprendiendo y corrigiéndose”, explica Marta Villegas.

Con el 20% restante se valida el algoritmo desarrollado para comprobar que su nivel de acierto es el óptimo. Pero no basta con entrenar el modelo una vez, el aprendizaje profundo o *deep learning*

que caracteriza a las redes neuronales se basa en aprender a partir de la experiencia, por lo que necesita repetir este proceso miles de veces. Solo así, la máquina puede llegar a distinguir qué es relevante. Para ello, se necesita también de una gran capacidad de cómputo, algo que el BSC ha superado sin problemas al ser administrador de MareNostrum, el superordenador más potente de España y el tercero en Europa.

Aumentar la colaboración de los hospitales

Lo investigadores del BSC se encuentran satisfechos con estos primeros resultados, pero se adentran en una nueva fase en la que deben poner a prueba el algoritmo de predicción con los datos facilitados por el hospital Clínic de Barcelona. Por otro lado, según Xavier Pastor, hay que tener en cuenta que son “datos recogidos en un momento determinado en unas circunstancias determinadas”, por lo que “si las circunstancias hacen que cambie el significado de los datos, ya tenemos un problema de desviación del modelo”.

Ambos coinciden, eso sí, en que para hablar de redes neuronales o procesamiento del lenguaje natural es fundamental aumentar los datos de los que disponen. Para ello, el BSC ya ha firmado un convenio de colaboración con el Hospital 12 de octubre de Madrid y se encuentra en conversaciones con el Hospital Universitario Virgen del Rocío de Sevilla, porque el propósito final es “que los resultados reviertan en todos los hospitales”, afirma Marta Villegas. Para la investigadora, el proyecto ha evidenciado, además, “la necesidad de tener datos interoperables y que en caso de emergencia sanitaria el acceso a los datos clínicos sea algo ágil”.

Las previsiones, de momento, son esperanzadoras. El equipo de desarrollo del BSC calcula que en unos tres meses tendrían una primera prueba de concepto con los datos aportados por el Clínic de Barcelona, teniendo en cuenta que “ellos tienen que poner recursos en un momento en el que las prioridades del día a día son las que son”, recuerda Marta Villegas.

Un reflejo más de la labor de los trabajadores de los hospitales de España, que no solo están librando la batalla contra el COVID-19 desde las salas de urgencias o las ucis, sino, también, desde las pantallas de sus ordenadores.