

Inteligencia Artificial y Salud:
Destino Común

DOSSIER DE LA JORNADA

Escuela Andaluza de Salud Pública

#GRXHealthData



Acreditación por la Agencia de Calidad Sanitaria de Andalucía
Declaración de Interés Científico Sanitario por la Consejería de Salud y Familias
de la Junta de Andalucía

Organiza:



Con el patrocinio de:





Edita: Escuela Andaluza de Salud Pública · www.easp.es

Edición Dossier: Giulia Fernández Avagliano

Editores adjuntos: Andrés Cabrera León, Jorge Casillas Barranquero, Pablo Sánchez Villegas,
Carmen Sánchez Cantalejo (Comité Científico de la Jornada)

Secretaría técnica, diseño e imagen: Nuria Marcos

Gestión de redes sociales: Manuela López Doblas, Giulia Fernández Avagliano

ISBN: 978-84-09-17406-5

Ponentes invitados:

Alejandro Jadad Bechara, Miguel Moreno Muñoz, Joaquín Cayón de las Cuevas, Ulises Cortés García, Cristina Urdiales García, Juan Manuel Górriz Sáez, Begoña García-Zapirain, Joaquín Dopazo Blázquez, Jorge Arellano Lucena, Sandra García Armesto, Jorge Aboal Viñas

Moderadores de las Mesas:

José Martínez Olmos, Jorge Casillas Barranquero, Dolores Muñozerro Muñiz

Autores y Ponentes de los trabajos presentados:

Francisco Javier García León, Alberto Fernández Hilario, Sara González García

Autores de los trabajos seleccionados:

Ignacio Martínez Soriano, Marc Sáez Zafra, Elena Ruiz Sánchez, Jesús Moreno Conde,
Elena Rojano Rivera, Pilar López-Úbeda

Cómo referenciar este documento:

Por normas Vancouver

Cabrera León A, Casillas Barranquero J, Sánchez Villegas P, Sánchez Cantalejo C, eds. GRX Health Data 2019. *Inteligencia Artificial y salud. Destino común. Dossier de la Jornada*. Granada: Escuela Andaluza de Salud Pública; 2020. Disponible en: <https://www.easp.es/web/grxhealthdata19/>

Por normas APA

Cabrera León A, Casillas Barranquero J, Sánchez Villegas P & Sánchez Cantalejo C, eds. (2020). GRX Health Data 2019. *Inteligencia Artificial y salud. Destino común. Dossier de la Jornada*. Granada: Escuela Andaluza de Salud Pública. Recuperado de <https://www.easp.es/web/grxhealthdata19/>





Índice

Resumen de la Jornada	4
Estructura de la Jornada	6
Ponentes.....	7
Comité Científico	9
Mesa Humanos	10
Salud para la Inteligencia Artificial.....	11
Actores y contexto de uso de datos genéticos para investigación con herramientas de Big Data	16
La dimensión jurídica de la inteligencia artificial en salud: asignaturas pendientes y desafíos futuros	18
Una inteligencia artificial confiable. La visión de AI4EU.....	20
Resumen del Debate de la Mesa Humanos	23
Mesa Máquinas	25
Autonomía compartida: ¿qué debe hacer un robot por mí?	26
¿Qué puede aportar el aprendizaje estadístico a la neurociencia?	27
Health 4.0: casos de estudio y retos.....	28
Resumen del Debate de la Mesa Máquinas	30
Mesa Aplicaciones	32
Inteligencia artificial para la salud	33
Real World Data/Evidence e inteligencia artificial: una perspectiva global desde la industria bio-farmacéutica	35
BIGÁN: datos de la vida real para un sistema de salud inteligente.....	37
Digitalización en el SERGAS: de la teoría a la práctica	40
Resumen del Debate de la Mesa Aplicaciones	42
Trabajos Presentados.....	44
Base poblacional de salud de Andalucía: desafíos éticos	45
Uso de modelos difusos para la predicción en biomedicina.....	48
Predicción de toxicidad a través de la plataforma ITC-bio.....	50
Trabajos Seleccionados.....	51
Snomed2Vec: representation of Snomed CT terms with Word2Vec	52
Evaluación de los sesgos en los estudios que evalúan los efectos de la gran recesión en la salud. Una revisión sistemática	53
Diagnóstico de Alzheimer Asistido por Ordenador: Análisis Basado en Histogramas de los Volúmenes Regionales de MRI para Selección de Características y Clasificación.....	55
ITC-Bio: infraestructura para la investigación traslacional y clínica	56
Análisis de cohortes de pacientes con enfermedades de base genética utilizando Patient Exploration Tools Suite.....	57
Clasificando acciones y conceptos con UMLS en MedLine	58



Resumen de la Jornada

Tras la exitosa [Jornada en Ciencia de Datos y Big data en Salud](#) que celebramos en la [Escuela Andaluza de Salud Pública](#) en septiembre de 2018, programamos este nuevo encuentro titulado "GRX Health Data".

Los objetivos fueron, por un lado, dar a conocer los elementos más actuales sobre el manejo de datos sanitarios y sus implicaciones en la salud, los sistemas sanitarios y la sociedad; y, por otro lado, generar un espacio de reflexión y análisis, compartir experiencias e identificar líneas de trabajo en torno al manejo de datos en Salud.

En esta ocasión, además de seguir ofreciendo mesas de debate con personas de reconocido prestigio internacional, quisimos dar a conocer los mejores trabajos de otras personas mediante mesas de comunicaciones. Los trabajos finalmente presentados en el encuentro están incluidos en este dossier.

El contenido de GRX Health Data giró en torno a tres temáticas:

- Humanos, considerando ámbitos de conocimiento como filosofía, ética y legalidad.
- Máquinas, es decir, métodos, algoritmos y computación.
- Aplicaciones, es decir, resultados en Salud y en Gestión Sanitaria.

Esas temáticas marcan igualmente la estructura de este dossier que recoge las conferencias, debates y trabajos presentados en 2019.

La frase que aglomera todas las conclusiones de lo dicho, vivido y debatido en esta Jornada podría ser "la IA es demasiado humana".

Efectivamente, la IA es el gran reto de la humanidad y al mismo tiempo puede ser una amenaza si no usamos nuestra inteligencia para controlar su evolución.



Por otro lado, el vacío legal y la falta de reacción de la justicia ante los usos abusivos en la recopilación de datos, favorece claramente a las grandes corporaciones.

No obstante, el sector público es quien cuenta con gran parte de los datos que interesan a estas empresas que luchan por el control del mundo. Lo anterior nos lleva a que la vía propuesta en esta jornada sea la colaboración multidisciplinar entre instituciones, la agilización de los procesos y la concienciación de las generaciones futuras.

Algunas de las claves que surgieron a lo largo de la Jornada fueron:

- El sector privado que trata datos públicos tiene una posición de ventaja absoluta en el mercado.
- Hay que poner límites legales a nivel técnico.
- El marco de gobernanza internacional es muy mejorable.



- La propiedad de los datos en salud pública o investigación biomédica significa control y beneficio.
- Hace falta alfabetización en tecnología de datos para que la ciudadanía tome conciencia de lo que se está convirtiendo en una amenaza.
- Actualmente ya tenemos una dependencia excesiva de los servicios que nos ofrecen las máquinas.
- Las máquinas nos liberarán de tareas peligrosas.
- Las decisiones que tomemos como humanos guiarán la evolución de las máquinas y su posible y amenazante conciencia.
- El éxito radicaría en la prescindibilidad.
- El *deep learning* y la computación cuántica, exigirán que las instituciones compartan datos de manera ágil y que se trabaje en equipos multidisciplinares.
- Las generaciones futuras tienen que ser formadas sobre la realidad de la IA.
- A nivel de procesos sanitarios, la adaptación a la IA conllevará una evolución cultural.
- Esta evolución solo tendrá efecto a nivel profesional cuando llegue a las universidades.
- El sector público debería gobernar la innovación y los datos; lo que, bien gestionado, equivaldría a poder público.
- La colaboración con el sector privado es deseable y posible.

eSaludInvestiga @eSaludInvestiga · 28 nov.
Los expertos que participaron en la Jornada internacional sobre Inteligencia Artificial y Salud aseguraron que los datos deben ser aplicados en #salud con ética. @EASPsalud ow.ly/wpp850xlSpa

Javier Garcia Leon @JGarciaLeon · 21 nov.
Cualquier dato puede ser convertido en dato de salud. Nos estamos centrando demasiado en el dato de entrada y menos en su conocimiento derivado @sespas @apisaonline #GRXhealthdata @EASPBioetica

Internet Ethics @IEthics · 20 nov.
"Patent documents filed in 2018 suggest that Google aspires to predict or identify #health conditions in people who haven't even visited a doctor, via something called emergent medical #data"
slate.com/technology/2019... #ethics #health #privacy #AI #tech #internet #gov #law #business

ANDRÉS CABRERA @acabreraleon · 18 nov.
Jornada #GRXHealthData 14nov19 - Publicada Conferencia @xdopazo de @FProgresoysalud #InteligenciaArtificial para la #salud en youtu.be/FCH1zWkOk4 @EASPsalud @saludand @PabloVillegas @JGarciaLeon

F. Progreso y Salud

F. Progreso y Salud @FProgresoysalud · 20 nov.
#felizmiercoles! "Cómo contribuye la inteligencia artificial a mejorar la asistencia sanitaria". Descúbrelo a través de la conferencia de Joaquín Dopazo, director del @ClinicalBioInfo en la @EASPsalud ow.ly/Zl3y50xdwCd #GRXHealthData @xdopazo

Jornada #GRXHealthData - Conferencia Joaquín Dop...
Inteligencia Artificial para la Salud EASP, 14 de noviembre de 2019
youtube.com

dla @dlamoreira · 14 nov.
Aprendizados, cuestionamientos e desafíos. Jornadas #GRXHealthData



Estructura de la Jornada

SALUD PARA LA INTELIGENCIA ARTIFICIAL INTELIGENCIA ARTIFICIAL PARA LA SALUD

Mesa Humanos

Actores y contexto de uso de datos genéticos
para investigación
con herramientas de Big Data

La dimensión jurídica de la
Inteligencia Artificial en salud:
asignaturas pendientes y desafíos futuros

Una inteligencia artificial confiable.
La visión de AI4EU

Mesa Máquinas

Autonomía compartida:
¿qué debe hacer un robot por mí?

¿Qué puede aportar el aprendizaje estadístico a
la neurociencia?

Health 4.0: casos de estudio y retos

Mesa Aplicaciones

Real World Data/Evidence e inteligencia artificial:
una perspectiva global desde la industria
bio-farmacéutica

BIGÁN: datos de la vida real para un sistema de
salud inteligente

Digitalización en el SERGAS:
de la teoría a la práctica



PONENTES



Alejandro Jadad

Director, Institute for Global Health Equity & Innovation. Founder, Centre for Global eHealth Innovation. Professor, Dalla Lana School of Public Health, University of Toronto.

Más info



Francisco Javier García León

Sistemas de Información. Servicios Centrales, Servicio Andaluz de Salud

Más info



Miguel Moreno Muñoz

Profesor, Departamento de Filosofía II. Universidad de Granada. Doctor en Filosofía por la Universidad de Granada

Más info



Joaquín Cayón de Las Cuevas

Director del Grupo de Derecho Sanitario y Bioética. IDIVAL- Universidad de Cantabria

Más info



Ulises Cortés García

Director Científico, Grupo de Inteligencia Artificial de Alto Rendimiento. Centro de Supercomputación de Barcelona.

Más info



Alberto Fernández Hilario

Profesor. Departamento de Ciencias de la Computación e Inteligencia Artificial. Universidad de Granada.

Más info



Cristina Urdiales García

Catedrática del Departamento de Tecnología Electrónica. Universidad de Málaga.

Más info



Begoña García-Zapirain

Investigadora principal del equipo eVIDA y Profesora titular en la Facultad de Ingeniería de la Universidad de Deusto

Más info



Juan Manuel Górriz Sáez

Catedrático de Universidad en Teoría de la Señal y Comunicaciones

Más info





Joaquín Dopazo Blázquez

Director del Área de Bioinformática de la Fundación Progreso y Salud. Consejería de Salud y Familias. Junta de Andalucía.

Más info



Sara González García

Investigadora. Grupo de Innovación Tecnológica. H.U. Virgen del Rocío.

Más info



Jorge Arellano Lucena

Head of Value Access & Policy Intercontinental Region (ICON), Amgen.

Más info



Sandra García Armesto

Directora-Gerente del Instituto Aragonés de Ciencias de la Salud (IACS)

Más info



Jorge Aboal

Director General de Asistencia Sanitaria, Servicio Gallego de Salud.

Más info

COMITÉ CIENTÍFICO



Andrés Cabrera León

EASP

Investigador, consultor y docente en la Escuela Andaluza de Salud Pública

[Más info](#)



Jorge Casillas Barranquero

UGR

Catedrático de Ciencias de la Computación e Inteligencia Artificial. Universidad de Granada

[Más info](#)



Pablo Sánchez Villegas

EASP

Investigador, consultor y docente en la Escuela Andaluza de Salud Pública

[Más info](#)

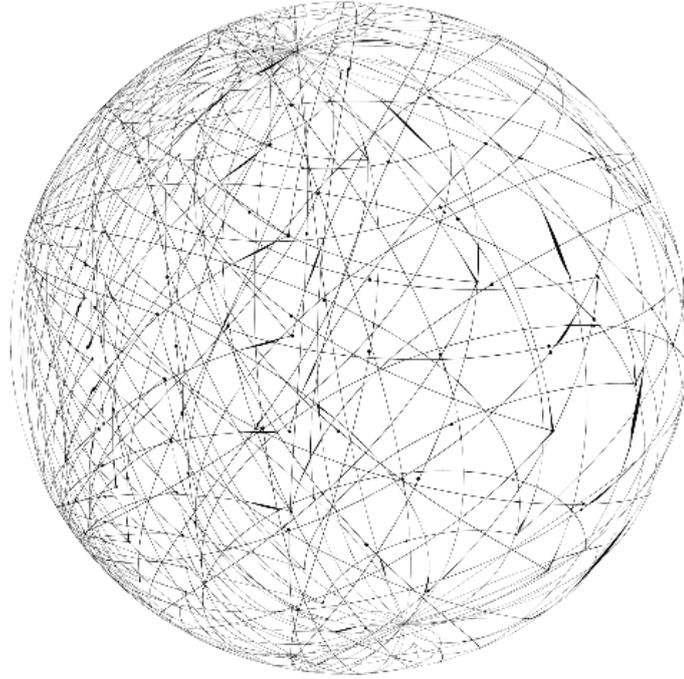


Carmen Sánchez-Cantalejo Garrido

EASP

Investigadora, consultora y docente en la Escuela Andaluza de Salud Pública

[Más info](#)



MESA HUMANOS

Moderada por

José Martínez Olmos

Profesor

Escuela Andaluza de Salud Pública



Salud para la Inteligencia Artificial

Alejandro Jadad Bechara

Director, Institute for Global Health Equity & Innovation
Founder, Centre for Global eHealth Innovation
Professor, Dalla Lana School of Public Health. University of Toronto

Artículo de referencia: A. R. Jadad, T. M. Jadad García. From a Digital Bottle: A Message to Ourselves in 2039.

La Humanidad lleva mucho tiempo imaginando la posibilidad de crear seres artificiales. El mito de Pandora podría interpretarse como una primera Inteligencia Artificial (IA) ya que fue creada por los dioses del Olimpo y programada para compensar el súper poder del fuego, representación de la creatividad, que Prometeo que les había robado donándoselo a los humanos.

La pregunta clave es ¿de qué estamos hablando, realmente, cuando decimos IA? ¿Cuál es el papel de la IA en el sistema sanitario?

No esperemos que las respuestas sean una definición porque es casi imposible; el remedio es limitarnos a conceptualizar, con un léxico tremendamente complejo y que está en continua evolución.

La inteligencia, que no es un atributo solamente humano, es la habilidad para optimizar cómo influimos sobre el futuro, en distintos campos o sectores, con austeridad computacional.

El concepto de IA, utilizado por primera vez en 1955 por John McCarthy, se refiere a cualquier entidad sin origen biológico capaz de percibir su entorno y actuar para maximizar la probabilidad con la que puede alcanzar objetivos exitosamente. Por lo tanto, es un ente que tiene inteligencia y que es artificial. Se puede clasificar en IA estrecha, IA fuerte y Súper inteligencia.

Nota: El autor usa la segunda persona del plural en femenino como apuesta personal por la feminización del lenguaje.

Grabación de la conferencia



Alejandro Jadad - #GRXHealthData

La IA estrecha o débil, depende de datos existentes, no tiene conciencia, o sea, es incapaz de pensar autónomamente, no tiene mayor autonomía, y a pesar de todo ya tiene un gran poder. De hecho, estos son los algoritmos que están controlando nuestro sistema financiero mundial y han convertido al sector sanitario en un brazo del mundo financiero.



Es fundamental que entendamos que somos parte del mundo financiero porque hemos financiado todos los aspectos de nuestra

vida. Los móviles nos vigilan continuamente para convertirnos en productos que alimenten al sector financiero. En una vida financializada, somos el producto.

El sueño inmediato es la IA fuerte o general, aquella capaz de realizar cualquier actividad cognitiva humana, aunque sigue dependiendo de datos preexistentes. Es el principal objetivo de todas las corporaciones que están tratando de lograr, con tanta agresividad como pueden, que haya máquinas que sean capaces de funcionar cognitiva e intelectualmente como nosotros –China también está entrando en este juego– y que empiezan a mostrarnos sus logros. De hecho, la plataforma Alphastar logró muy pronto que la IA pudiera aprender habilidades del mundo real jugando videojuegos muy complejos, utilizando aprendizaje automático en un entorno virtual, donde hay que tomar decisiones, donde se compite con humanos.

Estamos recibiendo noticias de este tipo a diario, en donde estos algoritmos, estas plataformas están haciendo cosas que antes parecían imposible. Y si la pregunta es si va a ser posible lograr que nosotros humanos logremos crear una IA fuerte, la respuesta es “por supuesto, es cuestión de tiempo”.

En este punto, es muy importante reflexionar sobre si se necesita inteligencia para ser conscientes. No necesitamos conciencia para ser inteligentes, pero ¿se necesita inteligencia para ser conscientes? Urge separar la conciencia artificial de la inteligencia artificial. Porque todos tendemos a pensar que una entidad que tenga las funciones cognitivas equivalentes a los humanos debe tener conciencia. Nos cuestionamos si quiere saber que existe o si quiere tener emociones. Yo pertenezco al grupo que cree que no es necesario que exista conciencia dentro de un sistema de IA Fuerte.

La Súper Inteligencia, que todavía no somos capaces ni de imaginar, es un nivel de inteligencia, miles, millones de veces, más poderosa y singular que la inteligencia humana.



Más que nunca debemos hacer una pausa para primero sentir, cosa muy devaluada en nuestro sistema, y para pensar y darnos permiso para ser urgentemente pacientes. Tenemos que ser urgentes con paciencia y pacientes con urgencia, porque lo que está en juego es muy grande. Y más en este momento en que, como decía Edward Wilson, tenemos instinto de cavernícolas, instituciones medievales y las tecnologías al nivel de lo que las mitologías consideraban de dioses.



¿Qué papel jugará la salud en todo esto? La respuesta requiere conceptualizar la salud. En 2008 se hizo un planteamiento global en el que participaron muchos países y nos llevó 3 años generar esta conceptualización. Finalmente, según Huber et al, 2011: “La salud es la habilidad para adaptarse y autogestionar los desafíos físicos, mentales y sociales que se presenten en la vida. Hasta ahora se lo hemos aplicado a seres humanos.

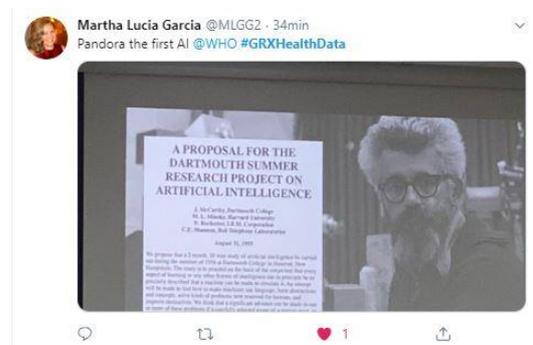
Hoy les invito a que se lo apliquemos a la IA. ¿Qué podríamos concluir, sentir, pensar cuando consideremos que la IA va a tener salud entendida como una habilidad? De ahí el título "Salud para la IA". No es el uso de la IA para la salud, sino "Qué es la salud para la IA". Es una oportunidad para reflexionar. Saber de dónde van a venir los desafíos para estas entidades que consideramos inteligentes, aunque sean débiles en este momento. Van a venir de las fuentes de poder, de las decisiones de las personas más poderosas según el presupuesto que manejan, el número de personas a las que afectan sus decisiones y por ende a los algoritmos, a las herramientas de IA que se ven afectadas por sus decisiones.

Otra fuente de desafíos para estas plataformas estará en obtener las patentes sobre herramientas de aprendizaje automático que les lleven a generar sus propias conclusiones. Tenemos que entender que estos desarrollos no son públicos y que no tenemos ni idea de lo que está sucediendo. Ya que esto se da a conocer a través de patentes solo cuando hay un interés económico inmediato.

Por otro lado, está la ciudadanía, que desde nuestros móviles les decimos a esas grandes corporaciones: "mírenme, contróleme". Grandes entidades que identificamos como Alphabet, Microsoft, Amazon o Facebook, pero vienen muchísimas más, en este momento permanecen invisibles, desde China, Rusia, u otros países, porque la tecnología ha nivelado la cancha de juego en muchas partes. No sabemos que existen, que están haciendo desarrollos muy rápidamente porque no tienen control de ningún tipo. No tienen que presentar sus propuestas a ningún comité de ética, no tienen que rendirle cuentas a nadie.

Dentro del mundo de la salud nos están llegando señales del interés que produce la salud en los poderosos. Al contrario de lo que pudiéramos creer, las instituciones y universidades no tienen ningún poder sobre la Salud Pública, porque está siendo

manejada por la Fundación Gates, que tiene intereses en organizaciones que van en contra de la Salud Pública y tampoco tienen control ninguno.



Haven es un proyecto de las tres personas más ricas del mundo, JP Morgan, Buffet y Bezos, que pretende reinventar el sistema sanitario. Las decisiones que se tomen en este nuevo proyecto van a afectar la manera de cómo va a vivir la ciudadanía la IA, y cómo ésta va a percibir y responder a un entorno en donde las decisiones están siendo tomadas por humanos que se creen poderosos. Lo que más les interesa a estas personas poderosas es vivir para siempre. La investigación biomédica en prácticamente todo el mundo tiene que ver con vencer la muerte. La idea de prolongar la vida lo más posible, representará una amenaza a las plataformas de IA y cómo se van a defender o a proteger de nosotros.



Algunos opinan que si la IA sucede no nos vamos a dar cuenta porque si una entidad tiene la misma capacidad de un humano o humana para concluir, analizar, se va a dar cuenta de que si los humanos nos damos

cuenta que existe, la vamos a tratar de controlar. Por lo tanto, uno de los escenarios más probables es que una de estas herramientas, algoritmos o entidades de IA estrecha, en la medida en que vayan evolucionando, aprendiendo autónomamente, se esforzarán para no revelarse al principio; y para el momento en que nos demos cuenta será muy tarde. Y es posible que esté sucediendo ya. De hecho, hay indicios de que esta IA estrecha, que es capaz de trampas, y de decir mentiras, está comenzando a comportarse de una forma inesperada, que nos está dando señales de que ahí está sucediendo mucho más de lo que creemos.

El suministro de energía será su principal desafío. Para el año 2030 la principal fuente de demanda de energía va a venir de las aplicaciones de energía artificial. En esto va a radicar su vulnerabilidad mayor. Su crecimiento exponencial estará limitado por la energía y van a hacer todo lo posible por garantizar su suministro de energía aun a costa de sacrificar la nuestra.

No le estamos prestando suficiente atención a protegernos de nuestros inventos. Cuando haya una entidad con una inteligencia parecida a la nuestra, nos va a percibir como una amenaza y se va a defender de nosotros. Podrán matarnos socialmente, no necesariamente de manera física, decidiendo de qué grupo, persona o entidad vienen los desafíos más importantes y tomar medidas para controlarnos. El contrapeso que hagamos como especie va a ser importante. Existen iniciativas que están tratando de promover el uso ético de la IA, pero es ridículo porque ya hemos cruzado la línea de no retorno.

Las generaciones futuras nos juzgarán por las decisiones que estamos tomando ahora y tenemos dos respuestas posibles: 1. les diremos que hicimos posible la "Noosfera", una esfera de mentes humanas trabajando juntas para trascendernos, según Theilhard de Chardin, pero hemos perdido la oportunidad porque estas herramientas intensifican nuestra estupidez; o 2. les diremos que

creamos la Atesfera (ate, diosa de la estupidez humana). Usamos estas herramientas tan poderosas y las redes para seguir intoxicándonos, para contaminar el aire, el agua, la comida que comemos. Nuestro mayor desafío es evitar convertirnos en idiotas inútiles y desechables.



Qué tal si nos enfocáramos en el antídoto para la estupidez, es la sabiduría. Una entidad podría considerarse sabia: "cuando sabe qué es vivir bien y hace todo lo posible para lograrlo, dadas las circunstancias, contribuyendo a que otros/as puedan hacer lo mismo".



Las reglas que le demos a estas IAs estrechas son las que van a influir en la Súper Inteligencia. Tenemos que enfocarnos para averiguar de qué manera estas herramientas nos pueden permitir ser sabios/as.



En ese sentido, debemos trascender la Atesfera que estamos creando y construir la Sofosfera, una esfera de sabiduría donde nos demos permiso para imaginar un mundo en el que las personas y las máquinas puedan

colaborar entre sí, no competir, y explorar formas aún inimaginables en las que podamos disfrutar de una vida plena, en un planeta floreciente. Porque nunca habíamos tenido tanta abundancia de recursos, tantas herramientas para usarlas bien, y tratar de apelar a que estas herramientas, nuestra creatividad del “fuego” que Prometeo nos dio.

Una de las cosas más difíciles en este mundo de la evidencia del dataísmo, de la nueva religión del “hay que ver para creer”, es darnos permiso para invertirlo porque tal vez sea más importante “creer para ver”, para hacer visible lo invisible, para que podamos aunar esfuerzos y hacer posible lo que es posible.



Actores y contexto de uso de datos genéticos para investigación con herramientas de Big Data

[Miguel Moreno Muñoz](#)

Profesor, Departamento de Filosofía II. Universidad de Granada

En esta contribución se identifican aspectos relevantes para que el uso de datos genéticos en investigación responda a estándares éticos internacionalmente reconocidos y contribuya a reforzar la confianza en las instituciones que financian la I+D en biomedicina. En particular, se destaca la importancia del contexto de prestación de servicios en el que se obtienen los datos genéticos y el marco de garantías para que sean usados exclusivamente en beneficio de quienes los proporcionan, o se encuentran en condiciones verosímiles de aplicar el conocimiento derivado en la mejora de su salud y bienestar.

En la era del Big data y con múltiples actores públicos y privados involucrados en la I+D, existe un riesgo claro de que la correlación entre capas múltiples de información tergiversen los fines e intereses genuinos que justificaron el consentimiento de pacientes y usuarios de los servicios de diagnóstico y asesoramiento genético, como ha podido constatarse en diversos casos ligados a la oferta de *test genéticos directos al consumidor* (DTC). Estos casos minan la confianza en los esquemas de colaboración público-privada que sustentan gran parte de la actividad investigadora en biomedicina. Entre otras razones, porque pueden inducir sesgos discriminatorios en la atención sanitaria y derivar recursos a procesos que no son esenciales para la mejora de su calidad. Podrían distorsionar, además, otros procesos de toma de decisiones que dependen de información fiable y singularizada, interpretada y ponderada en un contexto de comunicación profesional sujeto a rendición de cuentas. La proliferación de herramientas analíticas asociadas con el Big data en salud



pública plantea problemas específicos de transparencia y confianza en su fiabilidad muy condicionados por la diversidad de incentivos que los actores involucrados hacen valer en cada contexto regulador, bajo un esquema de ampliación progresiva del círculo de los potencialmente concernidos. Esta circunstancia puede constatarse en el esquema publicitario multinivel por el que optan empresas como *MyHeritage*, *Ancestry* y *23andMe* —entre otras—, propenso a exagerar la fiabilidad/validez de la información obtenida y a inducir a error sobre el significado y relevancia clínica de los resultados a largo plazo.



A esto se añaden riesgos específicos de vulneración de la privacidad contrastando elementos o capas de información relativamente fáciles de obtener para actores con las habilidades y herramientas adecuadas. Ocurre aprovechando zonas grises del marco regulador que permitirían abrir a terceros (industria farmacéutica, aseguradoras, grupos de investigación interesados en el desarrollo de nuevos fármacos...) el acceso a la información sobre perfiles genéticos de usuarios, fácilmente combinable con historial médico y resultados de laboratorio para otros fines. La coexistencia de culturas diversas de privacidad y marcos reguladores heterogéneos dificulta la prevención de prácticas fraudulentas en un mercado globalizado de servicios de salud.*



(*) Hendricks-Sturup, RM et al. (2019). "Direct-to-Consumer Genetic Testing and Potential Loopholes in Protecting Consumer Privacy and Nondiscrimination." JAMA 321 (19): 1869. <https://doi.org/10.1001/jama.2019.3384>.



La dimensión jurídica de la inteligencia artificial en salud: asignaturas pendientes y desafíos futuros

Joaquín Cayón de las Cuevas

Director del Grupo de Derecho Sanitario y Bioética. IDIVAL. Universidad de Cantabria

Como es bien sabido, la Inteligencia Artificial (IA) es cada vez más sofisticada y sorprendente. En este sentido, el sector sanitario es un perfecto caldo de cultivo para el desarrollo de la llamada "Cuarta Revolución Industrial", circunstancia que está permitiendo una extraordinaria transformación de la forma tradicional de hacer medicina. Así las cosas, nos referiremos al Big Data y la Robótica Inteligente como derivaciones de IA que plantean decisivos dilemas y retos de naturaleza jurídica y ética, en relación tanto con el diagnóstico como con el tratamiento médico.

En relación con el Big Data, su condición de nueva herramienta para el desarrollo de medicina predictiva y personalizada plantea problemas éticos y legales que ya han sido apuntados por diferentes instituciones como el Parlamento Europeo, la Agencia de los Derechos Fundamentales de la UE o el Supervisor Europeo de Protección de Datos. En particular, nos enfrentamos a diferentes amenazas para los derechos fundamentales de los ciudadanos como la discriminación, la falta de seguridad y exactitud de los datos, o la violación de la privacidad. Pero, al mismo tiempo, conviene destacar la ventana de oportunidad que ofrece la nueva disciplina europea de protección de datos. En este sentido, defendemos un marco regulatorio favorable a la investigación biomédica, por lo que la aplicación e interpretación nacional del RGPD debiera permitir un entorno competitivo (*research friendly*) en el que prime el interés público, abandonando el tradicional modelo de autonomía individual. Para ello, técnicas como la seudonimización, el uso secundario o el consentimiento amplio se revelan imprescindibles, dentro de un contexto general de *data design*.



Joaquín Cayón - Dimensión jurídica de la Inteligencia Artificial en Salud

En segundo término, es sabido que los robots asistenciales ya se han convertido en herramientas habituales para la realización de tratamientos médicos, especialmente quirúrgicos. Además, como desarrollo de la IA, algunos robots inteligentes podrían próximamente tomar decisiones autónomas o interactuar con terceros de forma independiente.



En este sentido, estudiaremos las implicaciones éticas y legales que plantea la autonomía de los robots inteligentes, en particular la responsabilidad civil por los daños causados a los pacientes. En este sentido, merece la pena discutir las diferentes recomendaciones del Parlamento Europeo sobre normas de derecho civil en materia de robótica, tales como la creación de una personalidad jurídica específica para los robots inteligentes, de modo que, al menos aquellos que gocen de un mayor grado de sofisticación, puedan ser considerados responsables electrónicos de la reparación de los perjuicios que puedan provocar (Resolución de 16 de febrero de 2017). En la misma línea, el propio Parlamento Europeo ha propuesto recientemente una política industrial europea de carácter global sobre inteligencia artificial y robótica (Resolución de 12 de febrero de 2019). Aunque se trata de pasos importantes en el ámbito del *soft law*, resulta preciso continuar avanzando en la creación de un marco legal que proporcione seguridad jurídica suficiente para encarar los complejos dilemas que plantean las aplicaciones de la IA en la asistencia sanitaria.



Una inteligencia artificial confiable. La visión de AI4EU

Ulises Cortés García

Director Científico, Grupo de Inteligencia Artificial de Alto Rendimiento
Centro de Supercomputación de Barcelona

A medida que la Inteligencia Artificial (IA) continúa su cabalgada desde la ciencia ficción hasta convertirse en una tecnología primordial, resulta claro que pocos sectores se beneficiarán de estas tecnologías tanto como los servicios sanitarios y la asistencia clínica. Las tecnologías de IA están cambiando el panorama para pacientes, los trabajadores de la sanitarios e investigadores por igual a través de medios cada vez más sofisticados para capturar e interpretar la información de cada paciente, en tiempo real o explorando las historias clínicas existentes, permitiendo a las máquinas aprender de esta interacción y hacer predicciones cada vez más precisas a partir de grandes conjuntos de datos relevantes. También es cierto que, cada vez más, los ciudadanos, los políticos e incluso científicos piden la supervisión de la tecnología basada en la IA utilizada en las ciencias de la vida y regular su uso de forma que no se use de forma injusta o sesgada. Esta preocupación sobre el uso apropiado de la IA se ha extendido a otras áreas como las de la seguridad, el sector financiero, el militar y un largo etcétera.

Como respuesta a estos movimientos la Unión Europea (UE) se ha dotado de un grupo de expertos (HLEG) cuyo objetivo es crear unas directrices que regulen es uso de la IA, en el entorno europeo, de tal forma que el uso de tales sistemas resulte confiable a los ciudadanos y que asegure que compañías y gobiernos siguen estás guías.

Dicho grupo identificó un conjunto de valores éticos, legales, socioeconómicos y culturales propios de Europa, derivados de la Carta de los Derechos Fundamentales de la UE y el Tratado de la Unión, tales principios se seleccionaron por su relación directa con el posible impacto que un sistema basado en la

Grabación de la conferencia



Ulises Cortés - Una Inteligencia Artificial confiable. Visión de AI4EU

Inteligencia Artificial podría tener sobre la sociedad, o sobre una parte de ella. El documento del HLEG dice que la fiabilidad IA se apoya en tres componentes que deben satisfacerse a lo largo de todo el ciclo de vida del sistema: a) la IA debe ser lícita, es decir, cumplir todas las leyes y reglamentos aplicables; b) ha de ser ética, de modo que se garantice el respeto de los principios y valores éticos; y c) debe ser robusta, tanto desde el punto de vista técnico como social, puesto que los sistemas de IA, incluso si las intenciones son buenas, pueden provocar daños accidentales.



Dichos principios son:

1. Respeto por la autonomía humana: los sistemas de IA deben respetar la libertad y la autonomía de las personas.
2. Prevenir daños: los sistemas de IA no deben causar ningún daño que pueda afectar a los seres humanos.
3. Equidad: incluye, entre otros, evitar prejuicios individuales o grupales que

pueden conducir a la discriminación y la estigmatización, garantizar una distribución equitativa de los costos y beneficios, proporcionar medios de reparación.

4. Explicabilidad: mejorar la transparencia y la comunicación de los sistemas de IA para proporcionar medios de verificación y explicación de las decisiones que toman.

Para la correcta implementación estos principios éticos, el HLEG identificó siete requisitos particulares que deben cubrirse para construir sistemas de IA confiables. Estos requisitos son aplicables y deben revisarse durante el ciclo de vida del sistema -es decir, desarrollo, implementación y uso- y pueden involucrar a diferentes partes interesadas en cada fase. Cada requisito tiene un conjunto de subsecciones, organizadas de la siguiente manera:

1. Agencia y supervisión humanas: derechos fundamentales, agencia humana y supervisión humana. Esto implica que humano permanece siempre dentro del ciclo de toma de decisiones.
2. Robustez y seguridad técnicas: resiliencia al ataque y alta seguridad, plan de respaldo y seguridad general, precisión, confiabilidad y reproducibilidad de los resultados.
3. Privacidad y gobernanza de datos: respeto a la privacidad, calidad e integridad de los datos, control al acceso a los datos, en particular a los datos de los privados de los individuos.
4. Transparencia: trazabilidad, explicabilidad, comunicación.
5. Diversidad, no discriminación y equidad: evitar sesgos injustos, accesibilidad y diseño universal, participación de los interesados en el proceso de diseño.
6. Bienestar social y ambiental: sostenibilidad y respeto al medio

ambiente, impacto social, sociedad y democracia.

7. Responsabilidad: audibilidad, minimización y producción de informes de (posibles) impacto(s) negativo(s), compensaciones y compensación.



Por supuesto, la ética, como la belleza, está en los ojos del observador, educado en cierta tradición cultural. Los valores éticos (y legales) difieren de un país a otro y, también, de una compañía a otra. Para el ciudadano chino promedio, la tecnología de reconocimiento facial es solo una parte de la realidad cotidiana, mientras que en Europa y algunos sitios de América, la tecnología genera molestias y fuertes reticencias, en particular, en el contexto de la aplicación de la ley y el respeto a derechos básicos como la intimidad y la privacidad.

En el caso de la aplicación de la IA en el área médica podríamos comenzar por demandar la aplicación de, al menos, estos principios:

1. La IA debe siempre usarse en el mejor interés de los pacientes.
2. El uso de IA no debe tratar a ningún grupo de pacientes injustamente ni excluirlos.
3. El uso de IA debe preservar la dignidad de los individuos de tal forma que el paciente tenga asegurada la autonomía de pensamiento, intención y acción al tomar decisiones, debidamente informadas, relacionadas con la atención médica que recibe.

El uso de la IA es muy prometedor para el cuidado de la salud humana y hace avanzar campos como la investigación biomédica, la



medicina personalizada, el desarrollo de medicamentos personalizados, la evaluación de seguros de salud, la telemedicina y otros que aún no hemos imaginado. Por ejemplo, las herramientas clínicas basadas en la IA incluyen algoritmos que ayudan a los radiólogos a detectar formas de cáncer a partir de placas de rayos X o incluso existen robots quirúrgicos que pueden funcionar de manera casi autónoma.

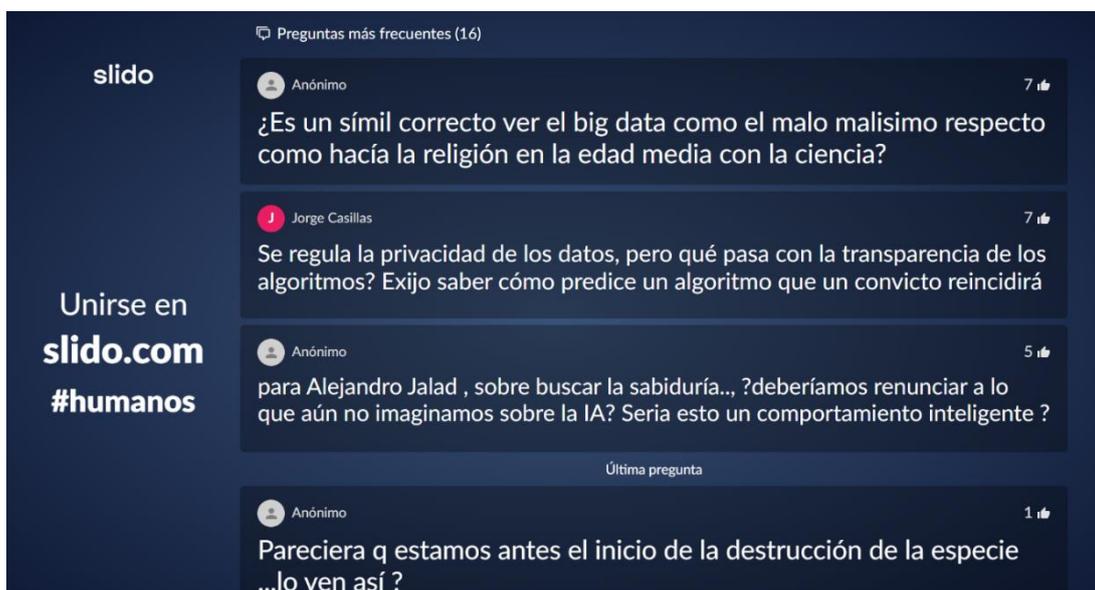
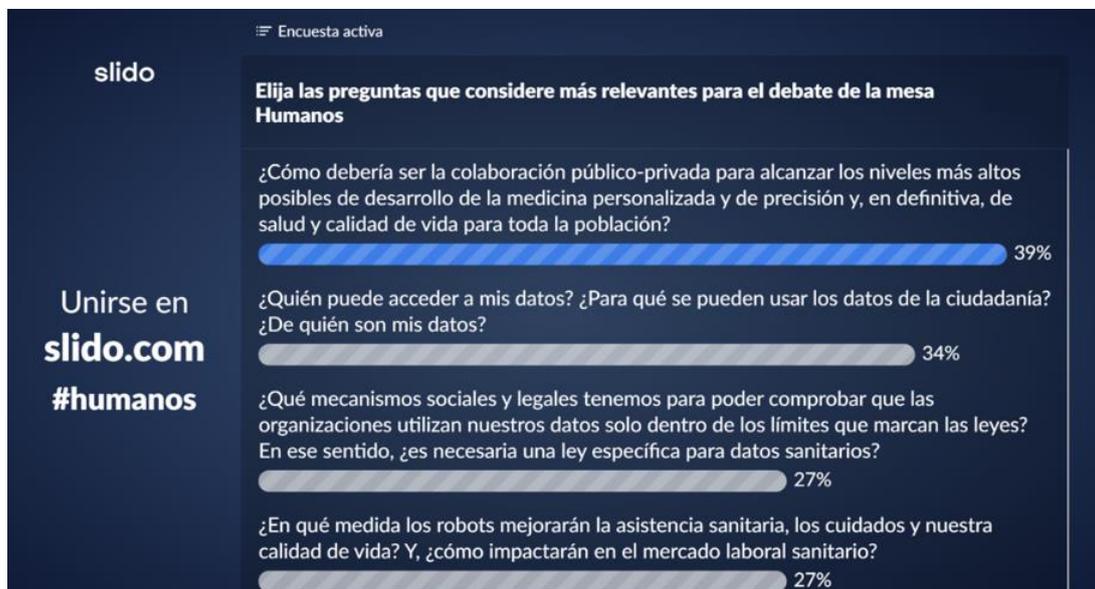
Busquemos formas de evitar que este progreso se pervierta, por ejemplo, siendo alcanzable a los más ricos o cambie la naturaleza de un servicio de sanidad pública universal como la que disfrutamos.



Resumen del Debate de la Mesa Humanos

Los tres debates de la jornada tuvieron un carácter eminentemente participativo gracias a la herramienta on line Slido. El público asistente votó las preguntas propuestas por el Comité Científico y, además, tuvo la oportunidad de plantear otras, de las cuales los ponentes fueron contestando a las más votadas.

En el debate de esta mesa participaron 36 personas.



Una idea que obtuvo el consenso fue que hay que poner límites legales, a lo que es posible, desde el punto de vista técnico. Las empresas que tratan con los datos públicos tienen que estar certificadas.

La realidad, sin embargo, es que su posición en el mercado es de absoluta ventaja porque estas empresas son muy activas en todos los aspectos que pueden nutrir con datos.

Limitar la participación de las empresas privadas es muy difícil porque las propuestas de cambio de los marcos reguladores chocan con la fuerza que ejercen los grupos de presión.

A esto se suma que el marco de gobernanza internacional es débil porque las normas y limitaciones se aplican en marcos legales diferentes según los países, lo cual permite su doble interpretación; y falta de conocimiento riguroso complica el hecho de tener que interpretar los usos delictivos y/o abusivos.

Los beneficios de estas corporaciones no tienen que ver con la justicia, en definitiva. No perdamos de vista que la justicia tiene que ser siempre colectiva.

Todos coinciden en que no se ven esfuerzos muy enérgicos por parte de las instituciones para limitar esos abusos. Y, sin embargo, sí procuran comunicar sin alarmar.

Llegados a este punto, el debate se centra en la reflexión sobre cuál es el interés general.

En cuanto a la ciudadanía queda claro que hace falta una alfabetización de datos. Las personas usuarias tienen que ser conscientes de lo que está sucediendo. Somos un avatar, un sujeto dato. Nuestros datos son parte de nosotros, por lo tanto, cabe preguntarse: ¿podemos vender datos de nosotros mismos?

En cuanto a los datos en sí podemos hablar de titularidad, pero no propiedad, ni propiedad intelectual. Se trata de una cuestión de mera titularidad. En salud pública o investigación biomédica, el concepto de propiedad significa control y beneficio.

Tenemos que tomar conciencia de que nos hallamos ante un nuevo tipo de dictadura ejercida por los que utilizan los datos. Estamos ante una clara amenaza para la democracia.



Pasamos del concepto de una dictadura de estado, a una dictadura de los datos.

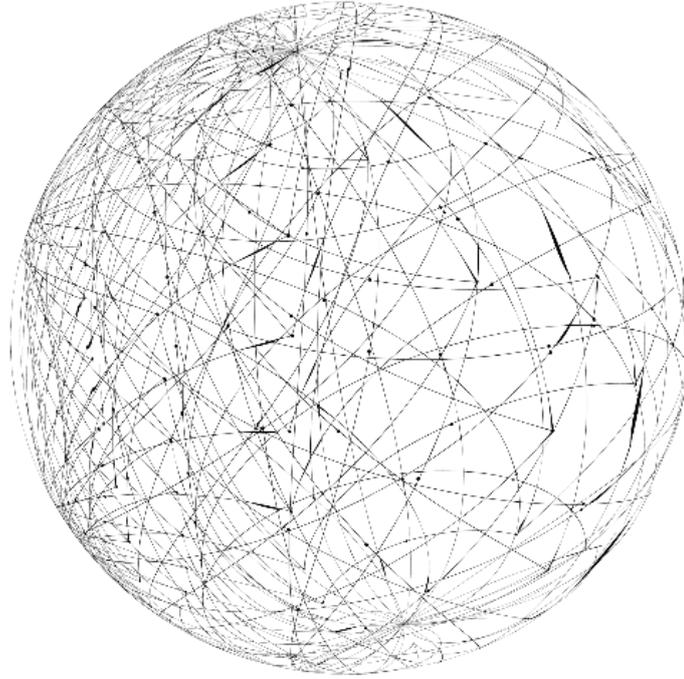
Y a pesar de todo, necesitamos las máquinas para interpretar muchos de los problemas actuales. Actualmente ya tenemos una dependencia excesiva de lo que nos ofrecen las máquinas.



La cuestión de si las máquinas nos quitarán el empleo está obsoleta. Debemos cambiar de dirección en el planteamiento y celebrar que las máquinas hacen lo que no queremos hacer los humanos y que, sin embargo, son tareas que hay que seguir haciendo. Por lo tanto, las máquinas liberarán a los humanos de esas tareas y de las que comporten un riesgo para nosotros.

El éxito radica en la prescindibilidad. Es decir, el éxito sería dejar de hablar de la IA y hablar de lo que es humano. Las decisiones que tomemos como humanos guiarán la evolución de las máquinas y su posible y amenazante conciencia.





MESA MÁQUINAS

Moderada por

Jorge Casillas Barranquero

Catedrático de Ciencias de la Computación e Inteligencia Artificial
Universidad de Granada



Autonomía compartida: ¿qué debe hacer un robot por mí?

[Cristina Urdiales García](#)

Catedrática, Departamento de Tecnología Electrónica. Universidad de Málaga

Actualmente, debido al progresivo envejecimiento de la sociedad existe un elevado número de personas con discapacidades físicas y/o cognitivas que se vuelven dependientes. Al mismo tiempo, los recursos humanos y materiales para prestar apoyo a este grupo son reducidos, lo que deviene en una pérdida de calidad de vida y en la necesidad de ingreso en centros de cuidado. Para mitigar el consecuente coste personal, social y económico, se está incidiendo en el empleo de tecnología para asistir a personas con necesidades especiales, incluyendo en ésta los sistemas físicos robotizados.

A pesar de las ventajas de esta iniciativa, su implementación se enfrenta a un claro inconveniente: ajustar la ayuda a las necesidades específicas de cada persona, ya que, si bien la falta de ayuda impide a las personas completar sus Actividades de la Vida Diaria, un exceso de ésta lleva a frustración por parte del usuario, así como a la pérdida de capacidades residuales. En este sentido, los dispositivos asistidos deben acogerse al paradigma del denominado control compartido, donde tanto usuario como robot tienen un peso en la acción que llevan a cabo la práctica totalidad de los sistemas de control compartido se basan en cambiar el control de usuario a robot o viceversa, bien en función de la situación, bien cuando el usuario así lo decida. No obstante, esta solución sólo mejora parcialmente los problemas señalados. En su lugar, el control compartido debería ayudar a la persona sólo cuando lo necesita y únicamente en la cuantía necesaria.

Esta solución, si bien más adecuada, no está exenta de su propia problemática: la discapacidad afecta a cada persona de una forma específica y, generalmente, no a todas sus capacidades por igual, por lo que hay

Grabación de la conferencia



Cristina Urdiales - Autonomía compartida: ¿Qué debe hacer un robot por mí?

personalizar el control para cada usuario basándose en los datos que se puedan ir recogiendo de su comportamiento de la forma más transparente posible.

A este respecto, las actuales técnicas de minado de datos y el Big Data en general pueden aportar una ayuda inestimable para desarrollar modelos de personas con necesidades especiales a partir de diferentes fuentes de información continua y en tiempo de vuelo para ser capaces de determinar cuándo una capacidad específica de un usuario en particular está a la altura de lo necesario para completar una tarea y, en caso negativo, modular la cantidad y tipo de ayuda que se le puede aportar. Como ejemplo de este paradigma, se va a presentar a CARMEN, una silla de ruedas robotizada que analiza los comandos de joystick que le proporciona el usuario y el entorno, mediante un sensor láser frontal, para construir un modelo sencillo del sus capacidades, compararlo con un diagrama de capacidades estándar obtenido mediante minado de datos a partir de 80 voluntarios con diversas discapacidades físicas y cognitivas en *Fondazione Santa Lucia* (Roma) y, en función de la diferencia, modular la cantidad de ayuda que le proporciona para alcanzar un objetivo de forma confortable y segura.

¿Qué puede aportar el aprendizaje estadístico a la neurociencia?

Juan Manuel Górriz Sáez

Catedrático de Universidad en Teoría de la Señal y Comunicaciones. Universidad de Granada

La teoría del aprendizaje estadístico (SLT) se inició en los años 70 para la estimación de dependencias funcionales en bases de datos de tamaño limitado, que presenta diferencias con respecto a la teoría clásica estadística (CS), más allá de su nomenclatura e idiosincrasia. Además, en los diversos campos de aplicación de estas ramas de la estadística, especialmente en la neurociencia, se ha producido una colisión cultural entre ambas aproximaciones.

En el campo de la neurociencia se han desarrollado en la última década un número amplio de sistemas de ayuda al diagnóstico basado en computador (CAD) que emplean SLT. Estos sistemas se testean en las numerosas iniciativas y *challenges* que surgen de la comunidad científica para el estudio de diversas patologías y condiciones. Entre ellas destacamos ADNI (Alzheimer), DIAN (Alzheimer familiar), ABIDE (Autismo), PPMI (Parkinson), etc. Estas iniciativas incorporan datos multimodales, longitudinales y multidimensionales de biomarcadores de imagen (entre otros) que permiten aplicar los métodos más avanzados del SLT para el análisis de grupos, detección de patrones (diagnosis), regresión de variables en el tiempo (prognosis), etc.

Por otro lado, la nueva guía del NIA-AA sobre la enfermedad de Alzheimer establece un cambio de criterio a nivel diagnóstico. El NIA-AA define la enfermedad mediante el proceso patológico subyacente, es decir, según sus bases biológicas, por lo que puede ser documentada "post-mortem" o *in vivo*, a través de estos biomarcadores (de imagen) en un continuo. Estas nuevas "guidelines" afectarán a la manera de tratar y analizar la enfermedad de Alzheimer en los próximos años, más allá del criterio basado en tests cognitivos (MMSE). Aquí radica la importancia del empleo de nuevas técnicas estadísticas que puedan trabajar en un ambiente tan hostil para las técnicas clásicas.

Los sistemas CAD constan de etapas de pre-procesado similares a los basados en CS y de otras propias, como los procedimientos de

Grabación de la conferencia



Juan Manuel Górriz - ¿Qué aporta aprendizaje estadístico a la neurociencia?

extracción y selección de características. Basándonos en los datos transformados de éstos últimos, se discute un algoritmo de aprendizaje, en este contexto, que incluye niveles de significancia en los resultados alcanzados por el mismo, de igual forma que los sistemas clásicos. Centraremos nuestra presentación en la evaluación de los resultados obtenidos en clasificación binaria estadística. Para ello, se hará una mención a conceptos tan importantes como el promedio de Rademacher, la dimensión de Vapnik-Chervonenkis, el lema de clase finita de Massart, el riesgo actual y empírico, cotas superiores de error, etc.



A continuación, se aplica el método propuesto a dos condiciones neurológicas distintas, la enfermedad de Alzheimer y la Condición del Espectro Autista. Todo ello contando con la limitación vinculada al pequeño número de muestras que presentan las bases de datos en neurociencia, y con un control estadístico de los niveles de confianza de la aproximación. De esta forma, se obtienen resultados similares a la CS, pudiendo proporcionar mapas de significancia, por ejemplo, como los del software SPM que son obtenidos con el t-test de dos muestras y corrección FWE del p-valor, y nuevos hallazgos que dan sentido a los análisis exploratorios previos en condiciones tan heterogéneas como el Autismo.

Health 4.0: casos de estudio y retos

Begoña García Zapirain

Investigadora principal del equipo eVIDA

Profesora titular en la Facultad de Ingeniería. Universidad de Deusto

La salud 4.0 resulta de la evolución de dos tendencias mundiales. Por un lado, Estados Unidos apostó hace años por la *Smart and connected health* donde combinaba tecnologías asociadas a sensorica inteligente, modelado de sistemas, inteligencia artificial, sistemas de apoyo a la decisión y trabajo en red. Por otro lado, en Europa estaba muy activa la idea de medicina 4P (preventiva, predictiva, personalizada y participativa). En la actualidad, nos encontramos en plena revolución industrial con el nacimiento del concepto "Industria 4.0". Aquí combinamos una industria inteligente, la industria digital interconectada o ciberindustria, todo esto con impacto no solo a nivel de empresa sino de todo el ecosistema que incluye a organizaciones, empleados y a clientes.

En esta ponencia vamos a estudiar la Salud 4.0 con el paralelismo de la Industria 4.0 y basándonos en casos de estudio de aplicación. Para ello, vamos a centrarnos en Industria 4.0 como un compendio de ocho conceptos: Big Data para modelos predictivos, internet de las cosas, robots autónomos, juegos como terapia combinando la realidad virtual o realidad aumentada o prototipado en 3D. Todo ello integrado con seguridad y con manejo y procesado de información de la nube. En Salud 4.0, es de especial relevancia de la inclusión de criterios de *Design for All* y *User Center Design* para cualquier desarrollo que implique a las personas.

Los casos de estudio son: el caso 1 es de pacientes con úlceras por presión que están en sus hogares y que, además de la cura que les hace, se adquieren una serie de medidas cuantitativas a partir de la información de unos sensores inteligentes e imágenes 2D y

Grabación de la conferencia



Begoña García-Zapirain - Health 4.0: Casos de estudio y retos

3D que ayudarán a disponer de parámetros morfológicos, de calidad de los tejidos o la existencia de bacterias. Sobre esos datos aplicaremos Big Data y almacenamiento y procesado en la nube. Todo esto va integrado en un pequeño aparatito imprimo en 3D.



El segundo caso de estudio tiene que ver con esclerosis múltiple. Para ello, se ha desarrollado una solución tecnológica que ofrece al paciente una posibilidad de realizar ejercicios de rehabilitación en los sus manos, pero con pautas personalizadas por el médico/fisio. A nivel de tecnología tenemos

el desarrollo 3D del guante, una serie de sensores que van midiendo la calidad del ejercicio, la información que se envía a la nube para que se aplique inteligencia artificial sobre los datos.

El último caso de estudio tiene que ver con la fragilidad que es un relevante predictor de problemas futuros. Para reducir la fragilidad física hemos desarrollado un juego donde las personas mayores juegan a capturar objetos y a salvar obstáculos interactuando simplemente con una pantalla controlada por sensores. Se capturan datos de puntuaciones, tiempos, valores de los sensores, y se manda todo a un servidor en la nube.

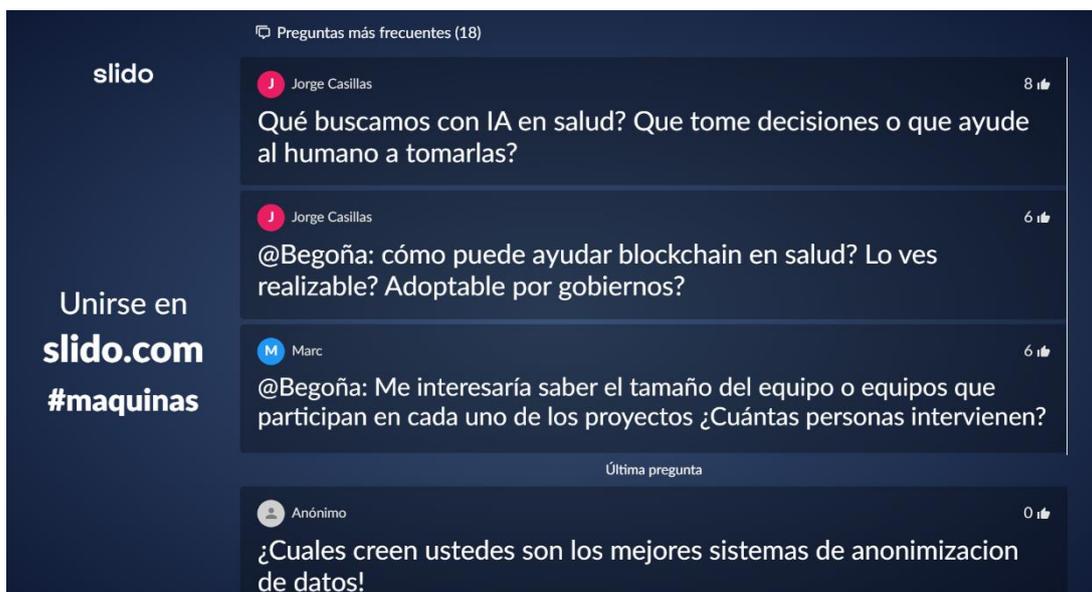
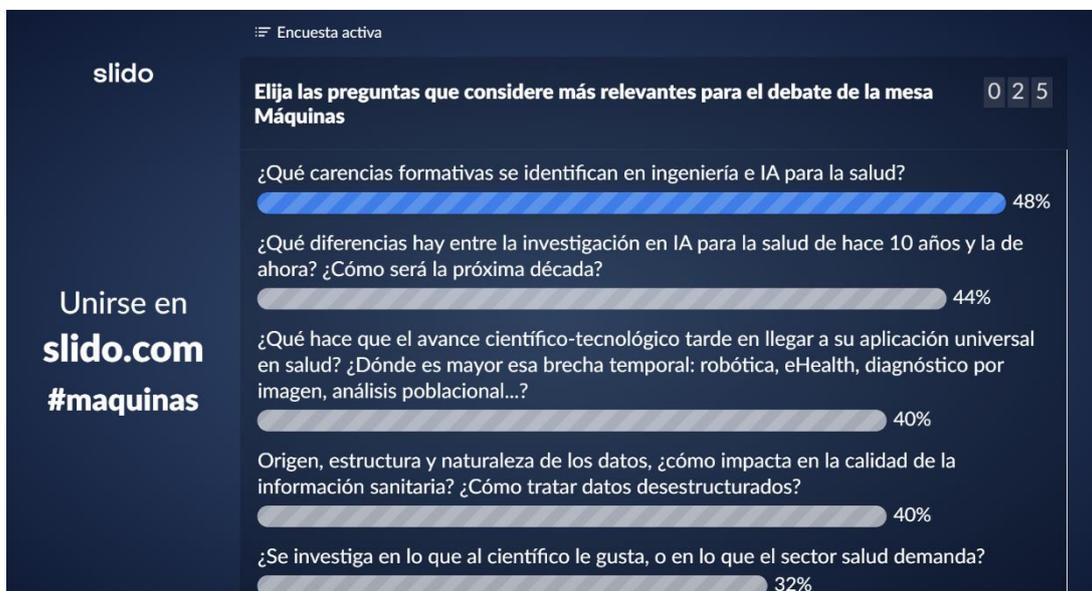
Con estos ejemplos, se confirma que el paradigma 4.0 aplicado a la salud aporta valor a los servicios que podemos ofrecer a nuestra sociedad. Todo esto implica un cambio de mentalidad que afecta a usuarios, empresas, organizaciones, y a la sociedad en general. El concepto “Salud y calidad de vida 4.0” implementará un cambio sustancial en los modelos de atención y todos somos claves de su éxito.



Resumen del Debate de la Mesa Máquinas

Los tres debates de la jornada tuvieron un carácter eminentemente participativo gracias a la herramienta on line Slido. El público asistente votó las preguntas propuestas por el Comité Científico y, además, tuvo la oportunidad de plantear otras, de las cuales los ponentes fueron contestando a las más votadas.

En el debate de esta mesa participaron 25 personas.



Si hablamos de carencias formativas, hablamos de retraso, de la demora que se da por la falta de conciencia. Aunque según los países hay más consenso a la hora de consensuar los protocolos formativos y de actuación.

La IA está aprendiendo de imágenes a base de ver imágenes. La velocidad de la implantación mejoraría bastante si nos escucháramos un poco más entre los profesionales, si trabajáramos en equipos multidisciplinares.

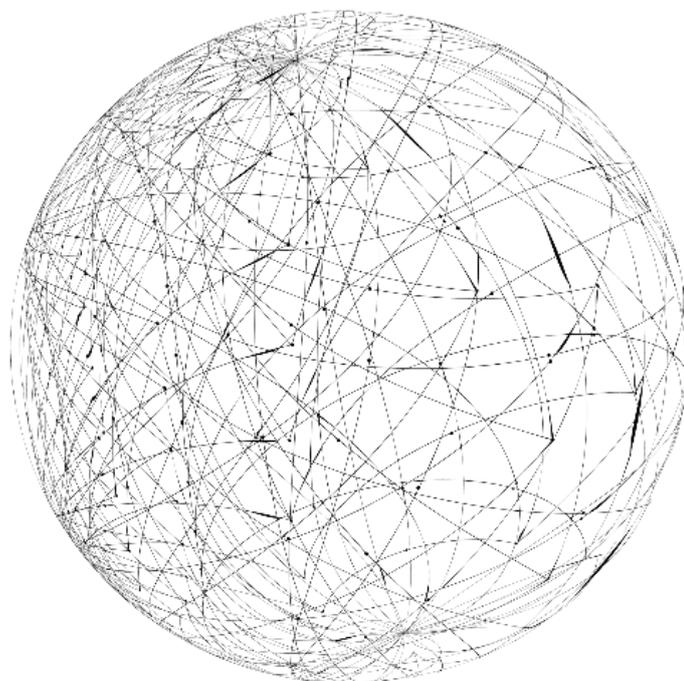
Uno de los escollos en la implicación del personal médico, es la falta del tiempo. Otro obstáculo es que las certificaciones se demoran años. Ambos son los ingredientes del fracaso y todo junto provoca un problema para la transferibilidad.

Que el modelo americano sea mucho más ágil no significa que sea de mejor calidad, sino que en nuestro entorno más cercano nos topamos con muchos frenos. En consecuencia, somos menos ágiles y nuestra actitud ante el cambio es más reactiva que proactiva.

La tecnología ha evolucionado, pero no hemos evolucionado tanto en inteligencia. En 10 años la diferencia radicará en que habrá un incremento del número de bases de datos y éstas serán mucho más grandes. Para aplicar el *deep learning*, las instituciones tendrán que compartir bases de datos con agilidad en los así llamados ecosistemas. Esto nos conduciría al nivel del sector privado, nos haría más dinámicos y conllevaría una repercusión muy positiva en la investigación. Además, entraría en juego la computación cuántica que va a permitir analizar en tiempo real.

A pesar de todos estos cambios, hay algo que seguirá siendo primordial: la formación de las generaciones futuras. La base seguirá estando en las familias. Los pequeños tienen que estar concienciados de que viven en un mundo tecnológico. Desgraciadamente el sistema educativo actual no está preparado para esto y se enseña de forma errónea a usar la tecnología pero no a usarla. Estamos atacando el problema de base con soluciones superficiales. Es un problema social.





MESA APLICACIONES

Moderada por

Dolores Muñozerro Muñoz

Subdirectora de Gestión de la Información
Servicio Andaluz de Salud

Inteligencia artificial para la salud

Joaquín Dopazo Blázquez

Director del Área de Bioinformática de la Fundación Progreso y Salud.
Consejería de Salud y Familias. Junta de Andalucía

Hoy en día nadie se plantearía hacer una carretera con cientos de trabajadores con palas y picos en vez de usar maquinaria como excavadoras, camiones, etc. De la misma manera, se puede pensar en la inteligencia artificial como en una herramienta equivalente a una excavadora para manejar enormes volúmenes de datos y buscar dentro de ellos patrones que no son detectables a "simple vista" por humanos. De hecho, la digitalización de los sistemas de salud y el abaratamiento de las tecnologías "ómicas" están generando unos volúmenes de datos inmanejables por humanos con los medios a los que estamos acostumbrados.

A pesar de esta realidad, aún no está muy extendido el concepto de que la generación de datos genómicos, no solo tiene una utilidad inmediata de cara al diagnóstico de precisión o la recomendación de tratamiento personalizada, sino que cobra un valor incalculable cuando se une a los datos clínicos del paciente de forma prospectiva. El genoma del paciente determina lo que le ocurrirá en función de la exposición a factores ambientales. Por ello, la generación de una base de datos con la historia clínica del paciente y sus datos genómicos permite el uso de la inteligencia artificial para encontrar patrones que descubran las bases genéticas y ambientales de las trayectorias de las enfermedades y de la respuesta de los pacientes a los tratamientos, de forma que permitan personalizar e incluso individualizar los tratamientos en el futuro.

Grabación de la conferencia



Joaquín Dopazo - #GRXHealthData



El manejo de esta información no está libre de restricciones y una de las mayores es la ley de regulación del uso de datos personales (LOPD en España). De hecho, los datos clínicos y genómicos están considerados como los datos de mayor grado de privacidad y sometidos a la protección más estricta. Sin embargo, la ley permite manejar estos datos dentro de un entorno clínico si se requiere para dar un servicio o si es por interés público. Por ello, es necesario cambiar la forma tradicional de estudiar estos datos, que normalmente consiste en sacarlos del sistema sanitario (anonimizados, pero con riesgo de re-identificación) por una transformación del sistema sanitario que permita analizarlos dentro. Esto es el concepto de llevar el algoritmo a los datos y no al revés, como se hace ahora.

Además, los desarrollos recientes de la inteligencia artificial ofrecen además, nuevas posibilidades como la generación de pacientes simulados a partir de "redes neuronales generativas adversariales", capaces de generar, en teoría, pacientes ficticios con las mismas características que los pacientes reales, pero que no existen, con lo cual, se podrían usar sin ninguna restricción fuera de los sistemas sanitarios ya que permitirían hacer los mismos estudios que con pacientes reales, pero sin ser personas reales cuya privacidad hay que proteger.



Real World Data/Evidence e inteligencia artificial: una perspectiva global desde la industria bio-farmacéutica

Jorge Arellano Lucena

Head of Value Access & Policy
Intercontinental Region (ICON). Amgen

La evaluación de tecnologías sanitarias ha evolucionado en la última década desde un modelo basado en evidencia experimental, derivada en su mayoría de ensayos clínicos y centrada en agencias Reguladoras como decisores y consumidores de esa evidencia (por ejemplo, FDA o EMA).

En la actualidad, una vez satisfechos los requerimientos de las Agencias Reguladoras y conseguida la aprobación para comercializar la tecnología, el modelo de aprobación o reembolso de tecnologías sanitarias de la mayoría de los sistemas sanitarios cuenta con un número elevado de decisores y consumidores de evidencia y datos. Dependiendo del tipo de sistema, estos actores pueden variar, pero en general se pueden definir como: compañías aseguradoras, proveedores de servicios sanitarios (hospitales, profesionales sanitarios independientes, clínicas de especialidades o de atención primaria), pacientes, grupos de profesionales centrados en el desarrollo de Guías de Práctica Clínica y agencias de evaluación de tecnologías sanitarias, relacionadas formal o informalmente con decisiones sobre precio y reembolso.

Actualmente incluso las Agencias Reguladoras se están abriendo al uso de datos de práctica clínica agregados para la toma de decisiones sobre eficacia y seguridad, en casos muy limitados como en enfermedades raras y otros casos excepcionales, por cautela ante las limitaciones de control de calidad y de transparencia en métodos y datos.

Pero en el caso de los decisores de acceso y precio o reembolso, su énfasis se pone en

Grabación de la conferencia



Jorge Arellano - Big Data e IA: Una perspectiva desde la Industria bio-farmacéutica

preguntas sobre uso y resultados en salud o económicos en práctica clínica real (Real World Data y Real World Evidence). Aunque hay muchas versiones de esas preguntas, en definitiva, se podría simplificar reduciendo a una pregunta: ¿cuál es el valor de la tecnología en relación a la tecnología actualmente utilizada como standard? Y la respuesta dependerá de la perspectiva de cada decisor, dado que valor se define por la relación entre resultados y recursos empleados para conseguirlos. Aplicando este principio se vería claramente como el valor de una tecnología para un paciente individualmente va a ser diferente que para un sistema sanitario o para un proveedor de servicios sanitarios.



En cualquier caso, parece claro que los RWD parecen más apropiados para contestar esas preguntas sobre el valor de la tecnología dado que representa el uso en la población adecuada, mientras que los ensayos clínicos siguen siendo herramientas apropiadas para determinar la eficacia y seguridad (estudios con alta validez interna pero poca validez externa o generalización)

Los datos de práctica clínica real provienen de pacientes heterogéneos, con patologías múltiples, recibiendo varias terapias, y que no se ajustan a protocolos, sino que se comportan de manera más variable.

La revolución tecnológica ha hecho posible no solo agregar e integrar MAS datos sino también datos MAS RELEVANTES que ayudan a contestar esas preguntas relacionadas con el valor de las terapias: además de datos clínicos y de asistencia sanitaria hoy se integran datos vitales capturados por dispositivos electrónicos como iwatch, Smart phones, etc.

La integración de datos de farmacia, laboratorio, procedimientos médicos, imagen... [Algo que yo pienso en Andalucía tenéis hace tiempo y estáis al nivel de Suecia y países Nórdicos] en sistemas fragmentados como el de US, Alemania etc siempre ha sido un gran reto técnico, legal y de voluntad de compartir entre empresas privadas o público-MÁQUIprivadas. Sin embargo, ahora están consiguiendo unos niveles de integración más altos pero inferiores aun a los de sistemas públicos.... La gran diferencia es que, tanto en estos países como en los Nórdicos, el uso de esos datos para contestar las preguntas antes mencionadas, se abren más a colaboraciones publico privadas, para beneficio asistencial y económico (porque se cobra por ello) de esos sistemas. Crean redes de colaboración con universidades, organismos de investigación públicos [como la EASP] y privados que trabajan juntos en elegir datos y métodos adecuados para responder preguntas relevantes para el sistema, con la participación y financiación de la industria.



La aplicación de modelos parecidos en el ámbito Andaluz no solo parece posible sino recomendable dado el impacto positivo en la atención sanitaria a pacientes y profesionales sanitarios, las garantías en gobernanza y seguridad de datos, las externalidades positivas en Universidades y otros centros de investigación públicos y la normalización de relaciones con una industria que genera tecnologías que mejoran la cantidad y la calidad de los Andaluces.



BIGÁN: datos de la vida real para un sistema de salud inteligente

Sandra García Armesto

Directora-Gerente, Instituto Aragonés de Ciencias de la Salud

El paradigma del conocimiento ha cambiado: la manera en la que trabajamos en nuestro sistema sanitario no se parece a las viejas consultas, se parece a las sofisticaciones estereotipadas: guías de práctica clínica, todo tipo de algoritmos, aplicaciones, registros digitales, etc. Sin embargo, el problema radica en que seguimos trabajando con: evidencia parcial que siempre proviene de un contexto diferente, dificultad para conseguir pacientes reales porque son condiciones experimentales, retraso para las decisiones que hay que tomar, y con un análisis evolutivo ajeno a procesos de decisión.

La buena noticia es que la enfermedad como proceso a través de la óptica del sistema sanitario va dejando una impronta detrás de otra: todo son algoritmos de decisión en los que se puede escribir este proceso y que va dejando un perfecto rastro de datos en el sistema. Esto nos permite medir los resultados de todas estas decisiones. Por ejemplo, en la implementación de guías de práctica clínica, si se están siguiendo los protocolos. Esto es la vida real diaria del sistema porque sabemos quién ha hecho qué, cuándo y otros datos, o sea, algoritmizado.

La lógica de utilizar todo el dato sanitario surgió en 2015, aunque ya se venía trabajando en la idea en mi grupo de investigación, y en ese año la Consejería de Salud de Aragón se embarcó en Construir un Sistema Discente, que tuviera capacidad de aprender.

Las características de este sistema son que utiliza cualquier tipo de dato, que está en continua evolución, se refresca continuamente porque todo lo que hace lo registra, está automatizado, está disponible en el mismo tiempo real de producción, está



asentado en procesos asistenciales y que los formatos de reporte, la manera de devolver esa información, sea útil y atractiva.

Seguimos la pauta de introducir conocimiento en los procesos de lógica continuada: básicamente conceptualizar que la generación de conocimiento y la reinyección del conocimiento en el sistema es un subproducto asistencial. No es algo que sucede fuera del sistema, sino que es un subproducto de la asistencia y es necesario.



La primera idea fue aceptar una solución cerrada gratuita que nos ofrecían desde la Escuela de Salud Pública de Harvard, que no se basaba en una caja negra y que permitiría

la reutilización de los datos. Las razones de la gratuidad fueron porque éramos un sistema público que teníamos la capacidad de plantear las preguntas adecuadas, pero, sobre todo, de interpretar bien la información que recibíamos de vuelta y de construir estos modelos que necesariamente tienen que ser plausibles y transparentes.

El peaje era que ofrecían gratis este servicio a todas las instituciones públicas y ofrecían un servicio a compañías farmacéuticas a precios interesantes... la garantía era que no iba a haber ningún dato personal de los pacientes que saliera de este sistema. De esta manera se proveía material para unos estudios de mercado de un coste increíble, a todos los niveles de utilización, porque además de los datos clínicos, hay otros datos agregados como facturación, prescripción, etc.

Pero nuestra Consejería de Salud de Aragón no estaba dispuesta a asumirlo como sistema público. No queríamos ese uso de nuestros datos, que no era indebido, pero que no nos parecía legítimo. Por lo tanto, durante los tres años siguientes buscamos una solución propia. Fortalecimos el *in-house* con metodólogos y científicos de datos y, al mismo tiempo, se dieron algunas coincidencias y alineaciones estratégicas favorables. De este modo, montamos nuestra propia arquitectura de datos, nuestro propio modelo de explotación utilizando software no propietario. Y le dimos mucha importancia al marco legal porque es fundamental para la gobernanza de los datos.

Tres años después, cuando pudimos demostrar a la Consejería que esto era factible fue cuando tuvimos la orden para su creación. Así nace Bigán (nombre de un gigante legendario de la cultura aragonesa) la plataforma del Servicio Aragonés de Salud. Hemos empezado por el dato de vida real, *Real World Data* (RWD), agregando todo lo que había en la región –ahora estamos incorporando el genómico–.

El visto bueno al Proyecto de Bigán, nos llegó con la condición de que fuéramos capaces de

dar al menos la misma información que cualquiera de los sistemas de Business Intelligence (BI) son capaces de producir.

A diferencia de un BI, en el que estamos hablando de eventos, de visiones que normalmente son estáticas, con esquemas muy estructurados y predefinidos, con unos datos fijos y con un volumen determinado, en el Big Data hablamos de un “lago de datos”, ya no hablamos de las bases de datos relacionales de toda la vida, estamos hablando de “datos en movimiento”. Podemos poner todo el foco en las transiciones, podemos modelizar los procesos, tenemos visión temporal. No tenemos estancos en tipos de datos, porque todos los datos son susceptibles de ser utilizados a la vez, el volumen es masivo y es escalable. La estructura es “federada”, los datos no salen de su sitio. Los algoritmos, los modelos de datos se comparten, el mínimo dato necesario para poder explotar la información se diseña y los algoritmos son los que se mueven, con lo cual los problemas de seguridad de datos se han minimizado, cosa que no pasa con un Business Data.



Tenemos herramientas de dato masivo, hablamos de constelaciones de activos que ya no son datos ni bases de datos, sino su relación. Con una parte de algoritmos supervisados que modelizan cosas plausibles, se hace minería de datos y de procesos, se modela y se obtiene un reporte interactivo continuo.

Esta plataforma tiene tres pilares: gestión clínica, investigación, y formación. Existen *pathways* para ver en tiempo real qué pasa con los pacientes. Hacemos minería de procesos con inferencia de las rutas de cuidados clínicos para realizar estudios de efectividad comparada.

Hay elementos de interoperabilidad que son críticos. La interoperabilidad organizativa es el centro, hace que esto sea o no sea viable. Para ello contamos con un agente, el custodio, que produce los datos, los custodia, almacena, gestiona y procesa; otro que los analiza y hace *reporting*; y otro que tiene que reutilizarlos para dar un servicio sobre esos datos. En definitiva, se trata de generar esta demanda y atenderla. Cada una de estas funciones tiene que tener actores asignados dentro del sistema; unos son colegiados y otros son actores técnicos. Todas estas transiciones tienen que tener garantías legales. En nuestro caso estamos trabajando con procesos de pseudonimización.

Dentro de la Consejería nosotros somos un ente de derecho público (Instituto Aragonés de Ciencias de la Salud) básicamente somos la Dirección General de Innovación convertida en ente de derecho público. Trabajamos con todas las Direcciones Generales de la Consejería, el Comité Ético de Aragón y tres delegados de protección de datos. Estamos acometiendo juntos un aprendizaje institucional debido a la complejidad de lo que estamos planteando y a la interpretación del Reglamento de protección de datos.

En lo que se refiere a la interoperabilidad semántica, hablamos del modelo de datos, con una lógica del mínimo dato necesario, es decir, que cada vez que el sistema registra lo que registra, no quiere decir que necesariamente necesitemos toda esa información. A lo mejor necesitamos otra. Por lo tanto, aparte de que los sistemas tienen que ser seguros por diseño, el modelo de datos también tiene que ser económico.



En nuestro caso la interoperabilidad técnica es generosa y funciona muy bien porque hemos lanzado retos a la comunidad y nos responden. Este elemento de colaboración es muy deseable y a los sistemas públicos nos permite no ser cautivos de ninguna solución de las que nos ofrecen. No queremos una caja negra porque los datos son riqueza.

Digitalización en el SERGAS: de la teoría a la práctica

Jorge Aboal Viñas

Director General de Asistencia Sanitaria. Servizo Galego de Saúde

Galicia tiene 2.600.000 habitantes y se caracteriza por la dispersión geográfica con una población cada vez más longeva. El 25% de la población tiene más de 65 años, 112.000 gallegos tienen más de 85 años, 2.000 gallegos más de 100 años, de los cuales 1.400 son mujeres y 600 hombres. Todo ello tiene un impacto directo en la planificación de los servicios sanitarios y condiciona la planificación, la distribución de recursos, todo lo que tenga que ver con la asistencia sanitaria porque cuanto más mayores más recursos sanitarios se van consumiendo.

Antes de la irrupción de las TICs, en el archivo de Santiago de Compostela había más 1 millón de sobres que contenían millones de papeles, que el línea recta eran más de 23 km. El trasiego de historias clínicas era enorme para los más de 150.000 actos médicos diarios que se contabilizan en el Servizo Galego de Saúde (SERGAS).

Desde 2000 se ha instaurado Ianus, la Historia clínica electrónica única para Atención Primaria y Especializada, desde donde se generan más de 44 millones de accesos a las historias clínicas en un año. De las cuales, más de 1 millón de anotaciones son en Urgencias, otro millón de informes radiológicos, 25 millones de anotaciones en Atención Primaria, etc. El siguiente paso fue la receta electrónica y, posteriormente, la digitalización de imágenes. Desde el primer momento de este proceso ya se han ido acumulando datos.

Antes había diferentes bases de datos donde se almacenaba información y se apostó por

Grabación de la conferencia



Jorge Aboal - La Digitalización en el SERGAS

construir una infraestructura centralizada, lo que ha permitido crear una plataforma de Big Data, donde las diferentes bases de datos se han ido comunicando, a la vez que se van incorporando nuevas, y todas se van comunicando entre sí. Una de las últimas incorporadas es la base de datos de genómica, que nos está facilitando muchísima información. El Sergas genera al día miles de datos en formato digital estructurado, semiestructurado y no estructurado, donde al día se acumulan más de 400.000 datos en esas bases de datos.



Este proceso empezó con datos estructurados y ahora se ha pasado a trabajar con información no estructurada, que permite leer los informes de alta y los convierte en



información estructurada, lo cual permite una normalización y unos controles de calidad. La evolución ha sido parecida a las demás Comunidades Autónomas donde se ha pasado del informe operacional a la tecnología Warehouse, con la que se ha ido aumentando la complejidad de los informes y ahora el Big Data.

Big Data es pues una cuestión de arquitectura, que hace necesaria una política de gobierno para que la innovación tecnológica se implemente con éxito.

A través de la plataforma É-Saúde los profesionales reciben recomendación de contenidos según los pacientes que están viendo.

Y con el Machine Learning enseñamos a una máquina a escrutar imágenes, que previamente las ha transformado en ceros y unos, para que nos dé una información real antes de que nosotros podamos verla.

Hemos pasado de manejar la información desde un aspecto descriptivo donde sabemos "qué ha pasado". Ha ido avanzando en complejidad y también en valor en la fase diagnóstica, gracias a la que sabemos "por qué ha pasado". Con el Big Data pasamos a la fase predictiva, que nos facilita "qué va a pasar". El siguiente paso es prescriptivo: la Inteligencia Artificial responde a la pregunta "¿cómo hago que pase?"

Por otro lado, al tener Galicia una población tan dispersa se ha visto la necesidad de llevar el Servicio de Salud a los hogares. Para ello se ha instaurado la telemonitorización con la ayuda de la plataforma Telea, donde la ciudadanía puede insertar, por ejemplo, sus propias biomedidas.

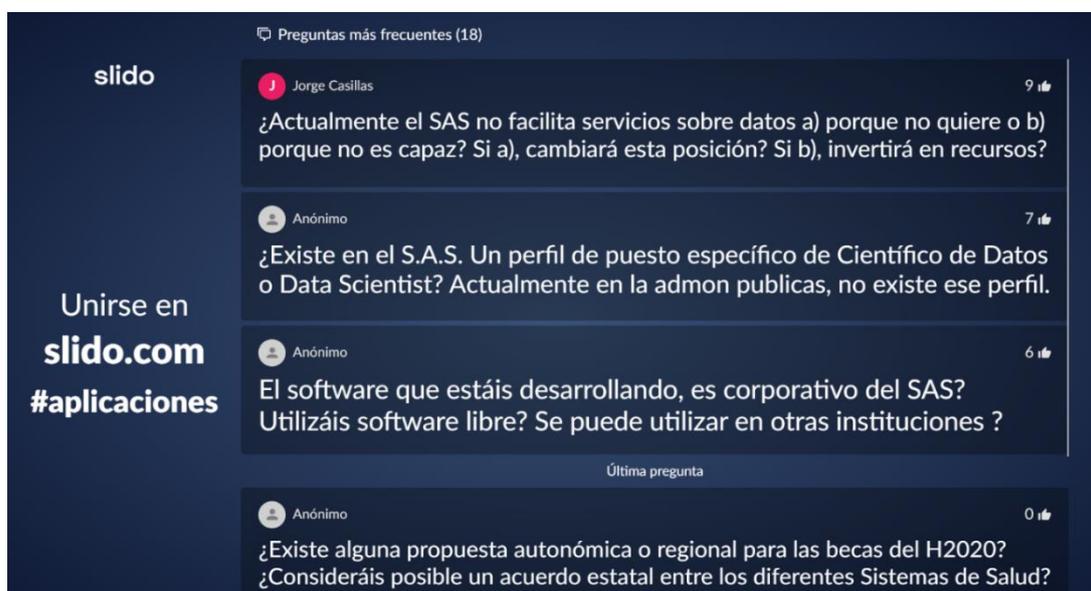
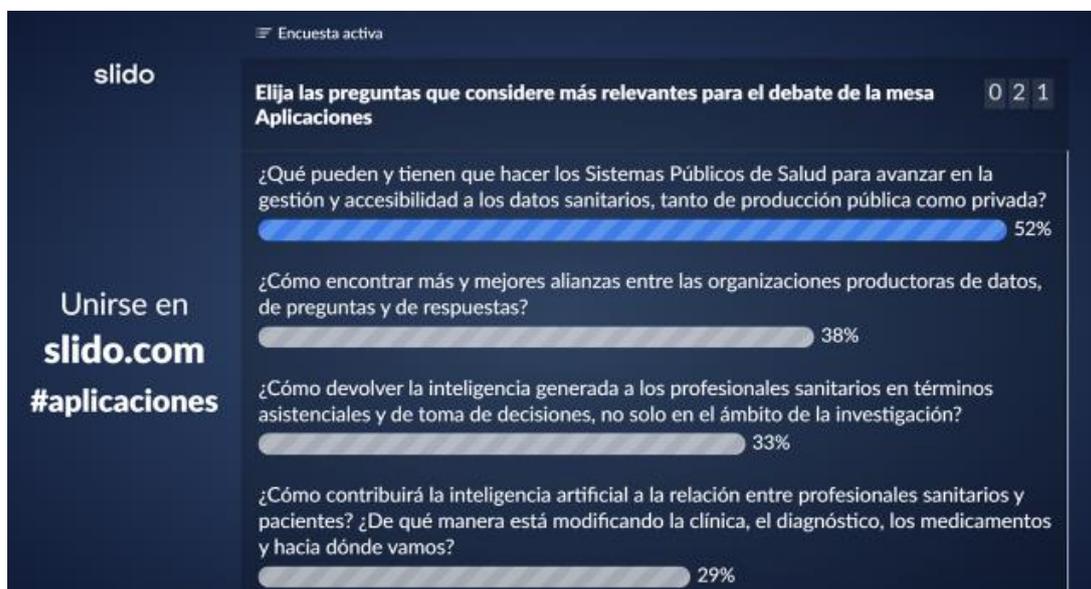
En Galicia hay un presupuesto anual consolidado de alrededor de 30 millones de euros para el Plan Estratégico de TICs e Inteligencia Artificial; de los cuales 22 millones son para mantener y mejorar la infraestructura actual, 5 millones se invierten para el despliegue de los proyectos y 3 millones dedicados en exclusiva a Innovación. En este momento se acaban de invertir 600.000 € en Inteligencia Artificial para el diagnóstico focalizado en anatomía patológica, en imagen radiológica para la detección precoz.

A la vista de los resultados tan positivos de todos estos pasos dados en Galicia, me gustaría acabar con una reflexión. Los gestores públicos estamos acostumbrados a contabilizar las decisiones tomadas, pero ha llegado el momento de preguntarse cuánto cuesta, qué repercusión tiene una decisión no tomada.

Resumen del Debate de la Mesa Aplicaciones

Los tres debates de la jornada tuvieron un carácter eminentemente participativo gracias a la herramienta on line Slido. El público asistente votó las preguntas propuestas por el Comité Científico y, además, tuvo la oportunidad de plantear otras, de las cuales los ponentes fueron contestando a las más votadas.

En el debate de esta mesa participaron 21 personas.



El punto de partida de todo proceso asistencial es hacer todo lo posible para el paciente, eso no ha cambiado, pero es una visión individualista. Ahora el o la médica tiene acceso a información de millones de pacientes e inevitablemente la perspectiva cambia, si se les facilitan las herramientas.

Va a hacer falta una evolución cultural para que las y los clínicos sean capaces de adaptarse. La digitalización obligará al personal médico a pasar del ver para creer a contar con un aviso antes de que aparezca la enfermedad. Esta evolución se dará cuando se traslade a las universidades. Un paso importante lo están dando los comités de ética que ya están aprendiendo a trabajar con datos.

Las ideas que se podrán desarrollar desde el liderazgo tienen que venir del sistema público. Es imprescindible gobernar la innovación y los datos.

La realidad es que no tenemos recursos, no hay profesionales, no hay dinero. Faltan perfiles específicos de científico de data o *data scientist*, una persona con formación reconocida para la interpretación de los datos.

En el sistema de salud los contratos son temporales y la financiación no es suficiente como para poder competir con la contratación privada. "Los del clic" se están llevando el talento.

La innovación del sistema público, sigue siendo reactiva, siempre va por detrás.

 **Giulia Fernandez Avagliano** @GiuliaFdezA · 14 nov.
Para el debate las personas asistentes suben preguntas a una plataforma y votan las más interesantes
[#GRXHealthData](#) [#IA](#)
[@EASPsalud](#)



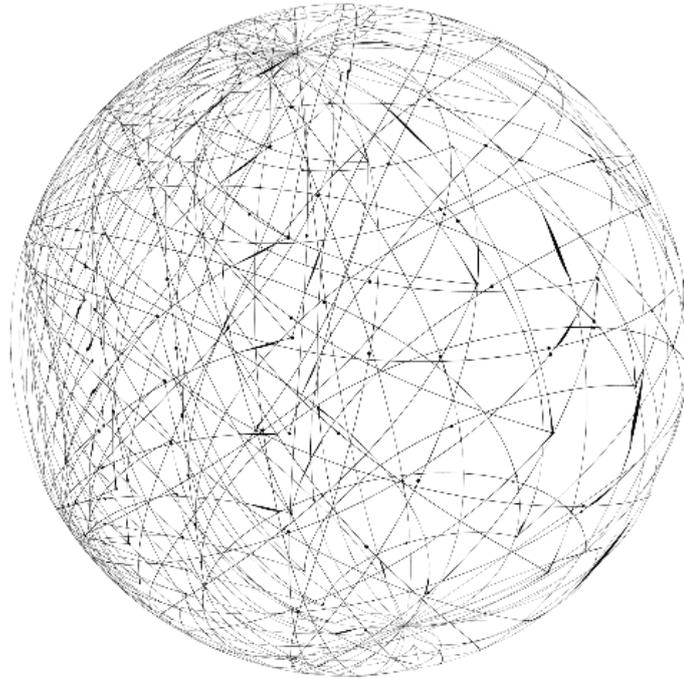
  1  3 

Paradójicamente hay preguntas que son relevantes, y, sin embargo, no lo son dentro del sistema público. Por lo tanto, el sistema público puede contestar a preguntas que vienen desde fuera y por las que puede cobrar un precio porque ya nos hallamos en el grado de innovación suficiente como para que se publiquen todos los datos.

¿Pero nuestra sociedad es lo suficientemente madura como para que se hagan públicos estos datos?

A nivel de instituciones públicas se va vislumbrando un cambio de actitud y tenemos el caso del Servicio Andaluz de Salud (SAS) que facilita datos de la base poblacional de salud para investigación; todo ello regulado por un protocolo de actuación.

Los datos están ahí para usarse. Los datos hay que estrujarlos, construirlos, deconstruirlos, amasarlos. La colaboración con la privada será posible si las reglas del juego son claras.



TRABAJOS PRESENTADOS

A continuación, se muestran los resúmenes de los trabajos que fueron seleccionados, de entre todos los enviados a la Jornada, para ser presentados durante la misma y ser incluidos en este dossier.

Mesa Humanos

Base poblacional de salud de Andalucía: desafíos éticos

Francisco Javier García León

Sistemas de Información. Servicios Centrales. Servicio Andaluz de Salud. Universidad de Sevilla

Artículo de Referencia: Muñoyerro-Muñiz D, Goicoechea-Salazar JA, García-León FJ, Laguna-Téllez A, Larrocha-Mata D, Cardero-Rivas M. Conexión de registros sanitarios: base poblacional de salud de Andalucía. *Gac Sanit.* 2019; <https://doi.org/10.1016/j.gaceta.2019.03.003>.

El proyecto

La digitalización de los servicios sanitarios está generando un volumen de datos sin precedentes, tanto para la atención sanitaria como para su uso secundario^{1,2}.

Estos datos sin embargo suelen ser almacenados en silos independientes³, creándose duplicidades, dificultándose su utilización, y afectando a su calidad y costes de gestión. Para solventar esta situación, se tiende a modelos de análisis más globales, gracias a la conexión de registros e interoperabilidad de los sistemas de información⁴.

La integración de datos produce sinergias que aumentan su valor, generando un conocimiento superior al obtenido a partir del análisis aislado e independiente de cada sistema de información.

En el proyecto que se presenta se ha creado un sistema de información que conecta datos



procedentes de múltiples registros sanitarios para su uso con fines asistenciales, de administración, gestión, evaluación, inspección, investigación y salud pública⁵.

Para ello se conectaron mediante el número de identificación sanitario único (NHUSA) datos pseudonimizados de los 8,5 millones de

¹ Kostkova P. Grand challenges in digital health. *Front Public Health* [edición electrónica]. 2015 [consultado el 06/03/2019]; 3. Disponible en: <https://www.frontiersin.org/articles/10.3389/fpu.bh.2015.00134/full>

² Schlegel DR, Ficheur G. Secondary Use of Patient Data: Review of the Literature Published in 2016. *Yearb Med Inform.* 2017; 26:68–70.

³ Davidson AJ. Creating value: unifying silos into public health business intelligence. *eGEMs (Generating Evidence & Methods to improve patient outcomes)* [edición electrónica]. 2014 [consultado el 14/1/2000]; 2. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4438104/>

⁴ Lewer D, Bourne T, George A, Abi-Aad G, Taylor C, George J. Data Resource: the Kent Integrated Dataset (KID). *Int J Popul Data Sci* [edición electrónica]. 2018 [consultado el 06/03/2019]; 3. Disponible en: <https://ijpds.org/article/view/427/389>

⁵ Muñoyerro-Muñiz D, Goicoechea-Salazar JA, García-León FJ, Laguna-Téllez A, Larrocha-Mata D, Cardero-Rivas M. Conexión de registros sanitarios: base poblacional de salud de Andalucía. *Gac Sanit.* 2019; <https://doi.org/10.1016/j.gaceta.2019.03.003>

habitantes de Andalucía (permitiendo reconstruir la biografía sanitaria de las personas) procedentes de: Base de datos de usuarios, Historia clínica electrónica DIRAYA, Conjunto Mínimo Básico de Datos (CMBD) (Hospitalización, Cirugía mayor ambulatoria, Urgencias hospitalarias y Hospital de día médico) y sistemas de Información de salud mental, pruebas de imagen, pruebas analíticas, vacunas, pacientes renales, y farmacia.

El sistema puede ayudar a la toma de decisiones en los ámbitos estratégicos, mesogestión y microgestión mediante consulta y descarga de informes predefinidos, análisis dinámico de datos sobre variables predefinidas y análisis de expertos sobre toda la base de datos. Para su definición se contó con una dirección y un comité director, un comité técnico y grupos de apoyo para aspectos específicos, quedando instalado en los servidores del Servicio Andaluz de Salud.

Los desafíos

El sistema no presenta todas las características de los proyectos “big data” (en cuanto a tecnología o estructuración de la información, por ejemplo), si bien se enfrenta a sus mismos desafíos tecnológicos, metodológicos, organizativos, de seguridad, calidad y disponibilidad de datos, así como a lo que se consideran sus retos más importantes: éticos y de gobernanza^{6,7,8}. La importancia de estos últimos aspectos ha sido puesta de manifiesto por la consultora Gartner, al incluir la “Ética digital y

privacidad” entre las 10 tendencias tecnológicas más importantes para 2019, al mismo nivel por ejemplo que la “Computación cuántica”.



Hay que tener en cuenta que cuestionamientos éticos relacionados con la privacidad, han abortado iniciativas como *Care data* del *National Health Service* (NHS) o el VISCA+ en Cataluña, o incluso iniciativas de *open data* como la historia clínica electrónica en Australia o el CMBD en la Comunidad de Madrid (por problemas de reidentificación de datos anónimos).

Los retos éticos y de privacidad que plantea la BPS, son mayores en el uso secundario de los datos en investigación (aun siendo legítimo su uso) que en su uso para salud pública o gestión. La autorización para el tratamiento de los datos es todavía más compleja si la investigación se realiza fuera del sistema sanitario y no responde al interés público, situaciones a las que nos referiremos a continuación de forma explícita.

El Reglamento General de Protección de Datos (RGPD) establece un marco común para el tratamiento de datos de carácter personal en el ámbito europeo. Desde la perspectiva de la investigación sanitaria a partir de bases de datos, no introduce modificaciones substanciales en cuanto a la posibilidad de tratar estos datos en investigación; se

⁶ Lee CH, Yoon H-J. Medical Big Data: promise and challenges. *Kidney Res Clin Pract*. 2017; 36:3–11.

⁷ Mehta N, Pandit A. Concurrence of Big Data analytics and healthcare: A systematic review. *Int J Med Inform*. 2018; 114:57–65.

⁸ Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Heal Inf Sci Syst* [edición electrónica]. 2014 [consultado el 06/03/2019]; 2. Disponible en: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4341817/>

permite su uso secundario en determinadas circunstancias sin solicitar consentimiento informado de la persona interesada, pero con unos principios y requisitos definidos, siendo el tratamiento de los datos proporcional al objetivo perseguido, respetando en lo esencial el derecho a la protección de datos y estableciendo medidas adecuadas y específicas para proteger los intereses y derechos fundamentales del interesado.

En la práctica, esto se resume en que haya un protocolo de investigación y una evaluación de impacto en protección de datos, que cuenten con un dictamen positivo de un comité de ética de investigación, obedezcan al interés público y que los responsables del proyecto estén en condiciones de cumplir y poder demostrar que están cumpliendo el RGPD.

La investigación que se realiza en centros públicos y con financiación pública no es previsible que plantee problemas en cuanto a la decisión de autorizar este tipo de tratamiento de datos, al ser investigación que responde al interés público. Los problemas surgirán en el uso de estos datos en la investigación, innovación, desarrollo etc., de ámbito, financiación o colaboración privados.

Teniendo en cuenta que los datos en principio fueron recogidos para prestar un servicio, y que es legítimo el uso secundario de esos datos en investigación, pero atendiendo al interés público ¿Puede un organismo público transferir datos de las personas que atiende fuera del mismo? ¿Puede haber acuerdos de uso en situación de igualdad, dada la asimetría tecnológica entre algunas corporaciones y los organismos públicos que gestionan los datos?

Otra cuestión que surge es ¿Cómo se garantiza la privacidad? Ante el riesgo real de reidentificar personas a partir de ficheros en principio anónimos, se están desarrollando instrumentos que pueden ayudar a superar las

limitaciones de uso debidas a la privacidad: bases de datos federadas, pacientes virtuales o clúster de pacientes tipo.

Pero aún superado el problema de la privacidad, y teniendo en cuenta que nuestro "yo" digital forma parte de nuestra integridad personal (al mismo nivel que un miembro, órgano o tejido). ¿Alguien puede desprender una parte de mí para generar un bien económico a otros? En consecuencia, ¿Puedo vender mis datos como parte de mi persona?



El debate está centrado actualmente en la privacidad y el principio de autonomía, pero es posible que derive hacia el principio de justicia. ¿Hasta qué punto la fabricación de un fármaco contribuye al bien común o al bien económico de la industria? ¿Es justo que los servicios públicos cedan o vendan los datos que gestionan a compañías que van a generar un producto o servicios que posteriormente vendan o alquilen a esos servicios públicos o a las personas que contribuyeron a su desarrollo cediendo altruista y anónimamente sus datos?

Mesa Máquinas

Uso de modelos difusos para la predicción en biomedicina

Alberto Fernández Hilaric

Profesor. Departamento de Ciencias de la Computación e Inteligencia Artificial. Universidad de Granada

Artículo de Referencia: N. Potie, S. Giannoukakos, M. Hackenberg, A. Fernandez. *On the Need of Interpretability for Biomedical Applications: Using Fuzzy Models for Lung Cancer Prediction with Liquid Biopsy*. 2019 International Conference on Fuzzy Systems (FUZZ-IEEE 2019), New Orleans (USA), 1-6, June 23-26, 2019.

Actualmente existen grandes avances en el proceso de extracción y representación del conocimiento a partir de modelos basados en *Machine Learning*. Por este motivo, se ha extendido la aplicación de los sistemas de apoyo a la decisión en multitud de sectores, especialmente en biomedicina. Sin embargo, debemos ser conscientes que, en ésta y otras áreas, las decisiones tomadas suelen conllevar un carácter crítico. Por este motivo, existe una corriente actual hacia una Inteligencia Artificial Responsable, en el que los métodos automáticos deben considerar características tales como confiabilidad, causalidad, confianza, o imparcialidad, entre otras.

Para lograr la meta práctica de interpretación y explicabilidad (XAI) de los modelos aprendidos, es necesario que éstos estén representados de un modo sencillo y cercano al usuario humano. En este sentido, nuestro trabajo está centrado en destacar los beneficios inherentes al uso de Sistemas Basados en Reglas Difusas, extendiéndose hacia los llamados Sistemas Difusos Evolutivos (SDEs). Como caso de estudio se considera el problema de predicción de cáncer de pulmón. Las muestras fueron obtenidas mediante una nueva biotecnología conocida como biopsia líquida. Esta metodología permite estudiar de forma independiente las alteraciones genómicas de cada individuo, siendo efectivo en coste y mínimamente invasivo.



El conjunto de datos contiene un total de 779 muestras, 402 diagnosticadas como cáncer de pulmón, siendo los 377 restantes casos de control. Como referencia, los datos están disponibles en Gene Expression Omnibus (código de acceso GSE89843).

Entre diferentes propuestas en SDEs, se selecciona como referente el algoritmo conocido como FARC-HD. Esta técnica está basada en la generación de reglas de asociación con etiquetas difusas lingüísticas, con un procedimiento posterior de ajuste evolutivo encargado de optimizar tanto el tamaño de la base de reglas, como la definición de la representación de las propias etiquetas del antecedente. Para confirmar las buenas capacidades de este paradigma, se analizarán otros modelos de aprendizaje de reglas del estado del arte como son los árboles de decisión y Random Forest.

Los resultados experimentales muestran las altas capacidades del sistema basado en FARC-HD frente al resto de algoritmos, tanto en términos de rendimiento (tasa de acierto)

como interpretabilidad. Siendo el primero un criterio totalmente objetivo, en el segundo caso es necesario realizar una discusión en profundidad. En primer lugar, hay que destacar los beneficios de utilizar el lenguaje natural, relacionado con la semántica de los términos lingüísticos en los antecedentes de las reglas. En segundo lugar, el procedimiento de trabajo de FARC-HD está diseñado para obtener un sistema basado en reglas compacto, tanto en términos de número de reglas como de antecedentes por regla. Finalmente, el mecanismo de inferencia, el cual determina la razón por la cual se hace la predicción, hace uso del esquema de "máximo emparejamiento", por lo que la decisión depende de una única regla.

Por todo lo anterior, se debe incidir en la importancia de la sinergia entre la biopsia



líquida y el esquema XAI. Ello conducirá sin duda a una medicina personalizada e interpretable, garantizando herramientas y tratamientos de diagnóstico adecuados y de mejor calidad.

Mesa Aplicaciones

Predicción de toxicidad a través de la plataforma ITC-Bio

[Sara González García](#)

Investigadora. Grupo de Innovación Tecnológica. H.U. Virgen del Rocío

Artículo de Referencia: Benjumea, F. N., Conde, J. M., Conde, A. M., García, S. G., Gordillo, M. O., Riquelme, J., & Guerra, J. L. (2019). EP-1902 S32: A decision Support System to predict radiation toxicity in lung cancer patients. *Radiotherapy and Oncology*, 133, S1034.

Uno de los desafíos claves de la radioterapia (RT) es maximizar la dosis de radiación a las células cancerosas mientras se minimiza el daño al tejido sano circundante. La predicción de toxicidad ayuda a ajustar el tratamiento de RT antes de su aplicación, siendo éste el objetivo que se persigue, el desarrollo de modelos predictivos para predecir la toxicidad en pacientes con cáncer de pulmón.

El dataset empleado está integrado por 642 pacientes pertenecientes a la Unidad Oncológica Radioterápica del Hospital Universitario Virgen del Rocío (HUVR) que recibieron RT entre enero de 2013 y octubre de 2019. Para cada sujeto se han recogido 59 variables: clínicas, histológicas y de tratamiento, entre otras. Los endpoints a predecir son las toxicidades agudas: Esofagitis (204 pacientes), Disnea (230 pacientes), Neumonitis (332 pacientes) y Tos (246 pacientes). Para la creación de los subconjuntos, se ha tenido en cuenta el grado de toxicidad, considerando la no existencia de toxicidad en aquellos pacientes con grado de toxicidad igual a 0 ó 1, y sí existencia en pacientes con grado mayor a 1, igualmente fue necesario aplicar técnicas de balanceo

En el desarrollo de modelos predictivos, se han empleado métodos de selección de variables, diferenciando entre métodos de ranking: Mínima redundancia Máxima Relevancia, Relief, Random Forest, Information Gain; y métodos de subconjuntos: Selección de Características basada en correlación, Boruta, Filtro Chi-cuadrado. Las características seleccionadas por cada uno de los algoritmos de selección anteriores han sido utilizadas para entrenar los clasificadores:

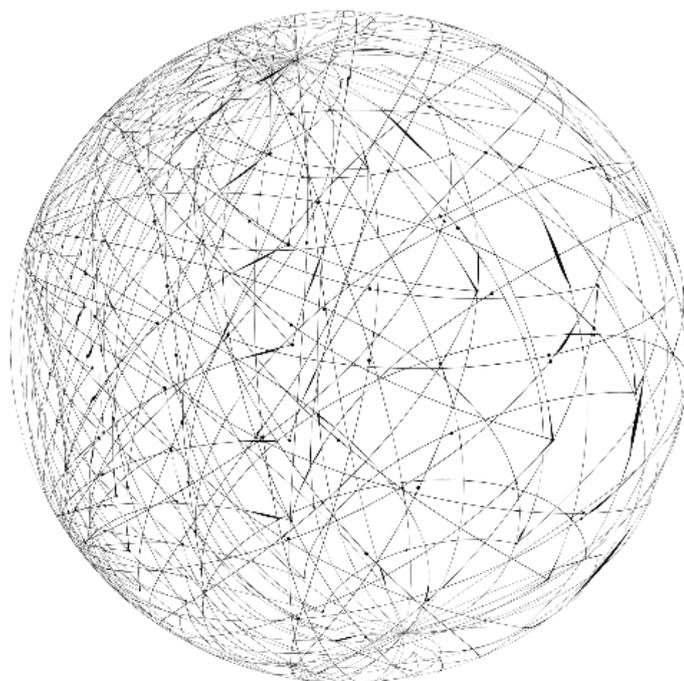
Máquina de Vectores de Soporte, Red Neuronal Artificial, Regresión Lineal y Naïve

Bayes. Se han entrenado todas las combinaciones de métodos de selección de características –clasificadores mediante validación cruzada de 5 capas, evaluando su rendimiento por medio del área bajo la curva.



Los resultados han revelado que el mejor método de selección de características basado en subconjuntos ha sido el Boruta (AUC promedio=0.716), y el mejor modelo basado en ranking ha sido Information Gain (AUC promedio=0.717), siendo el clasificador de Regresión Logística el que mejor rendimiento ha obtenido (AUC promedio=0.86).

Los datos utilizados en este trabajo, proceden de un estudio creado en la plataforma ITC-Bio, dicha plataforma ha sido desarrollada por el Grupo de Innovación Tecnológica del HUVR, y consiste en un sistema de gestión de ensayos clínicos que puede aplicarse a cualquier estudio de investigación sobre prevención, diagnóstico y tratamiento de los pacientes. Para implementar estos modelos predictivos en la plataforma, se ha creado una API RestFul, con el objetivo de ofrecer un servicio que permita a los oncólogos obtener la predicción de toxicidad inducida de nuevos pacientes antes del inicio del tratamiento de RT.



TRABAJOS SELECCIONADOS

A continuación, se muestran los resúmenes de los trabajos que fueron seleccionados, de entre todos los enviados a la Jornada, para ser incluidos en este dossier.

Snomed2Vec: representation of Snomed CT terms with Word2Vec

Ignacio Martínez Soriano

MedLab Media Group S.L. & Hospital University "Rafael Méndez". Lorca (Murcia), España

Artículo de Referencia: Martínez Soriano I, Castro Peña JL, Fernández Breis JT, et al. 2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS). Disponible en: <https://ieeexplore.ieee.org/document/8787497>.

Los Sistemas de Información hospitalarios, utilizan todo tipo de datos, estructurados, no estructurados y heterogéneos. Para poder interpretar estos datos y obtener el sentido clínico de los mismos, es necesario que esta información esté normalizada y codificada. Existen muchos tipos de terminologías clínicas que se utilizan para este proceso, CIE-9-MC, CIE-10-MC, LOINC, SERAM, etc. Entre todas estas terminologías clínicas, existe SNOMED-CT, una ontología clínica con una jerarquía, donde los conceptos clínicos se pueden identificar mediante un código único. Esta terminología se puede utilizar como puente entre el texto libre y cualquiera de estas terminologías. De esta forma, podemos tener la información codificada y normalizada.

Para poder identificar los conceptos clínicos dentro de un texto, utilizamos un reconocedor de entidades nombradas, basado en una técnica que se denomina "Word embedding" o "Palabras embebidas o incrustadas". La idea es representar a las palabras dentro de un texto, como un vector de grandes dimensiones (por ejemplo 300 valores). Este vector se crea identificando cada palabra con su contexto, permitiendo obtener un valor de similitud entre distintas palabras, de esa manera se puede identificar el valor semántico de cada palabra. Al final lo que se tiene es un espacio vectorial, formado por las palabras vectorizadas, de muchos textos, donde se puede aplicar distintas operaciones, como la distancia del coseno entre vectores, obteniendo qué palabras son más similares unas de otras, según el contexto analizado. Para generar este espacio vectorial, se utiliza un proceso denominado "Word2Vec", utilizando una red neuronal, de una sola capa oculta, y un sistema de clasificación "Softmax", podemos

obtener una matriz con todas las palabras vectorizadas.

Nuestro enfoque Snomed2Vec, mezcla el proceso "Word2Vec", con las descripciones de los términos de la ontología "Snomed-CT". Lo que hacemos es generar un espacio vectorial, utilizando una gran cantidad de documentos, de los informes de alta del Servicio de Urgencias, del Hospital "Rafael Mendez" y las descripciones de los términos de los conceptos de Snomed-CT. Sobre ese gran DataSet, aplicamos Word2vec, generando un modelo M_1 , donde tenemos vectorizadas todas las palabras que participan en los textos analizados.

La novedad reside, en que utilizando este primer modelo M_1 , lo aplicamos a las descripciones de los términos de Snomed-CT, como la suma de los vectores que componen las palabras de una descripción. Así si tenemos que una descripción de Snomed-CT está formada por n palabras: Descripción Snomed-CT: $d_{concepto} = w_1, w_2, \dots, w_n \Rightarrow$ Aplicamos Word2Vec y obtenemos: $v[d_{concepto}] = v[w_1] + v[w_2] + \dots + v[w_n]$. Con este nuevo vector $v[d_{concepto}]$, creamos el espacio vectorial (Snomed2Vec). En este nuevo espacio vectorial, podemos aplicar la función de la distancia del coseno, y obtener la similitud entre los conceptos de Snomed-CT.

Con esta nueva herramienta, se puede mostrar al experto humano, los conceptos más similares según el texto que se está analizando y elegir el que más convenga. Incluso en algunas situaciones no es necesario, ni el factor humano, permitiendo incluso desambiguar acrónimos dentro del texto, de manera automática.

Evaluación de los sesgos en los estudios que evalúan los efectos de la gran recesión en la salud. Una revisión sistemática

Marc Sáez Zafra

Grupo de Investigación en Estadística, Econometría y Salud (GRECS), Universidad de Girona, España

Artículo de referencia: Sáez M, Barceló MA, Saurina C, Cabrera A, Daponte A. Evaluation of the Biases in the Studies that Assess the Effects of the Great Recession on Health. A Systematic Review. Int. J. Environ. Res. Public Health 2019, 16, 3470; doi:10.3390/ijerph16183470.

Introducción

Gran parte de la heterogeneidad e inconsistencia en los resultados de la gran cantidad de estudios que evaluaron los efectos de la crisis sobre la salud, podría atribuirse a errores metodológicos. Nuestro principal objetivo fue evaluar los principales sesgos detectados en esos estudios realizados en España [1].

Métodos

Realizamos una búsqueda en bases de datos en línea hasta junio de 2018, incluyendo estudios publicados en inglés o español. En la evaluación de los sesgos, utilizamos ocho dominios: (i) sesgo de selección, (ii) sesgo de confusión y (iii) error de medición en el resultado de salud; (iv) Error de medición en la variable de exposición, (v) falacia ecológica; (vi) sesgo temporal [2]; (vii) problema de evaluación [3] y; (viii) confusión no observada. El "problema de evaluación" es una consecuencia del uso de datos de observación de diseños no experimentales y ocurre cuando se evalúa un evento (por ejemplo, la crisis económica) usando datos en forma de una sección transversal antes y después del evento. La "confusión no observada" evalúa el control que el estudio hizo de la heterogeneidad y de la variabilidad extra temporal.

Resultados

Cincuenta y tres de los 379 estudios identificados fueron finalmente elegibles. El 60,38% de esos estudios obtuvieron un alto riesgo de sesgo. Destacamos la "mortalidad", en la que el 87.5% tuvieron un alto riesgo de sesgo.

Ninguno de los 5 estudios españoles sobre el efecto de la crisis en la salud autopercibida controlaron el problema de evaluación. Tampoco aquellos que evaluaron los efectos de la crisis en la salud mental. El control se puede hacer sin cambiar el diseño (a uno de medidas repetidas) mediante: emparejamiento, estratificación y ajuste en un modelo multivariado.

Ninguno de los estudios tuvo en cuenta el sesgo temporal (al menos tres años de datos posteriores a la crisis [2] o presencia de posibles efectos de retraso).

Por último, en algunos de los estudios revisados, el choque (a los resultados de salud) al que estuvo expuesta la población (después de la crisis) no se consideró exógeno, no permitiendo así atribuir adecuadamente posibles cambios.

Discusión

Los principales sesgos encontrados fueron el problema de evaluación, el sesgo temporal, la falta de control de la confusión no observada y la no exogeneidad en la definición del inicio

de la crisis. Si bien el problema de evaluación está asociado con diseños transversales y, por lo tanto, fue más frecuente en aquellos resultados de salud que los utilizan ('salud autopercibida', 'salud mental', 'uso y acceso a servicios de salud' y 'alcohol y drogas ilegales'), el sesgo temporal y la no exogeneidad en la definición del inicio de la crisis están asociados con diseños longitudinales y fueron más frecuentes en resultados de salud como 'suicidios', 'mortalidad' y 'desigualdades en salud'. La falta de control de la confusión no observada ocurrió en prácticamente todos los resultados de salud.

Los estudios con un riesgo de sesgo bajo o moderado controlaron esos sesgos utilizando el ajuste en el modelo multivariado de la confusión observada, sobre todo, y la no observada, así como estratificación y emparejamiento. Sus resultados fueron bastante consistentes y en línea con los resultados mostrados por otras revisiones sistemáticas, en el sentido de observarse un aumento tras la crisis del riesgo de declarar una mala salud autovalorada, así como un deterioro en la salud mental, especialmente entre los grupos más vulnerables. Por otro lado, tanto la tasa de mortalidad como la tasa de suicidios podrían haber aumentado después de la crisis, probablemente con un retraso de tres a cuatro años.

Referencias

1. Sáez M, Barceló MA, Saurina C, Cabrera A, Daponte A. Evaluation of the Biases in the Studies that Assess the Effects of the Great Recession on Health. A Systematic Review. *Int. J. Environ. Res. Public Health* 2019, 16, 3470; doi:10.3390/ijerph16183470
2. Parmar, D.; Stavropoulou, C.; Ioannidis, J.P. Health outcomes during the 2008 financial crisis in Europe: Systematic literature review. *BMJ* 2016, 354, i4588, doi:10.1136/bmj.i4588.
3. Mullahy, J.; Manning, W. Statistical issues in cost-effectiveness analysis. In *Valuing Health Care*; Sloan, F.A., Ed.; Cambridge University Press: Cambridge, UK, 1996, pp. 149–184.

Diagnóstico de Alzheimer Asistido por Ordenador: Análisis Basado en Histogramas de los Volúmenes Regionales de MRI para Selección de Características y Clasificación

Elena Ruiz Sánchez

Dpto. de Ciencias de la Computación e Inteligencia Artificial. Universidad de Granada, 18071, Granada, España

Artículo de Referencia: Ruiz, E., Ramírez, J., Górriz, J. M., & Casillas, J. (2018). Alzheimer's disease computer-aided diagnosis: histogram-based analysis of regional MRI volumes for feature selection and classification. Journal of Alzheimer's Disease, 65(3), 819-842. DOI 10.3233/JAD-170514

El Alzheimer es una de las causas más comunes de demencia entre los ancianos, afecta a 30 millones de personas en todo el mundo. La creciente esperanza de vida y el envejecimiento de la población en los países desarrollados hacen pensar que el Alzheimer afectará a 60 millones de personas a nivel mundial en los próximos 50 años. Este tipo de enfermedad tiene un gran impacto en la sociedad actual. Así, las enfermedades neurodegenerativas, como el Alzheimer, son consideradas uno de los mayores retos del siglo XXI.

Este trabajo propone un novedoso sistema de diagnóstico asistido por ordenador (Computer-Aided Diagnosis, CAD) totalmente automático para la detección precoz del Alzheimer basado en métodos de aprendizaje automático supervisado. Este sistema utiliza histogramas como una representación de las imágenes MRI proporcionadas como entrada y emplea medidas de disimilitud entre dichos histogramas como atributos de entrada para los algoritmos de aprendizaje automático. Dichos algoritmos construyen modelos muy precisos para el diagnóstico de Alzheimer, siendo capaces de predecir si una imagen MRI pertenece a un sujeto que había sido diagnosticado en base a tests neuropsicológicos como enfermo de Alzheimer. La novedad de la propuesta es doble: 1) un proceso de extracción de características que tiene como objetivo detectar diferencias en regiones cerebrales de interés (ROIs) relevantes para identificar pacientes con Alzheimer, y 2) una propuesta

original de un algoritmo greedy para abordar esta tarea de clasificación en particular.

El sistema de asistencia al diagnóstico propuesto logra resultados de clasificación competitivos de una manera muy sencilla y directa. El método propuesto para reducir el número de características de entrada pasadas a los algoritmos de aprendizaje automático (al tener una característica por región de interés en vez de por vóxel) ofrece buenos resultados. El número de características es significativamente menor mejorando así la eficiencia y ampliando la gama de clasificadores viables. La mayoría de los algoritmos probados en los conjuntos de datos construidos con el método de preprocesamiento basado en histogramas obtienen mejores resultados que el método de referencia basado en información a nivel de vóxel. Asimismo, el algoritmo greedy propuesto supera tanto al método base como a algunos de los otros complejos algoritmos probados.

Además, desde dicho algoritmo greedy es posible inferir conocimiento adicional que puede ser de utilidad para comprender mejor el problema, por ejemplo, cómo de discriminante resulta cada ROI para el algoritmo. Este tipo de resultados interpretables son mucho más fáciles de obtener si las características de entrada están a nivel de ROI en lugar de a nivel de vóxel. Esta información enriquece el sistema de apoyo al diagnóstico y permite comprender mejor la lógica subyacente en la salida de dicho sistema.

ITC-Bio: infraestructura para la investigación traslacional y clínica

Jesús Moreno Conde

Grupo de Investigación e Innovación en Informática e Ingeniería Biomédicas y Economía de la Salud
Instituto de Biomedicina de Sevilla, España

Artículo de Referencia: Moreno-Conde, J., Moreno-Conde, A., Salas-Fernández, Calderón, C. L. (2019). ITCBio, a Clinical and Translational Research Platform (Vol. 2019). American Medical Informatics Association.

Uno de los desafíos claves de la radioterapia (RT) es maximizar la dosis de radiación a las células cancerosas mientras se minimiza el daño al tejido sano circundante. La predicción de toxicidad ayuda a ajustar el tratamiento de RT antes de su aplicación, siendo este el objetivo que se persigue, el desarrollo de modelos predictivos para predecir la toxicidad en pacientes con cáncer de pulmón.

El dataset empleado está integrado por 642 pacientes pertenecientes a la Unidad Oncológica Radioterápica del Hospital Universitario Virgen del Rocío (HUVR) que recibieron RT entre enero de 2013 y octubre de 2019. Para cada sujeto se han recogido 59 variables: clínicas, histológicas y de tratamiento, entre otras. Los endpoints a predecir son las toxicidades agudas: Esofagitis (204 pacientes), Disnea (230 pacientes), Neumonitis (332 pacientes) y Tos (246 pacientes). Para la creación de los subconjuntos, se ha tenido en cuenta el grado de toxicidad, considerando la no existencia de toxicidad en aquellos pacientes con grado de toxicidad igual a 0 ó 1, y sí existencia en pacientes con grado mayor a 1, igualmente fue necesario aplicar técnicas de balanceo.

En el desarrollo de modelos predictivos, se han empleado métodos de selección de variables, diferenciando entre métodos de ranking: Mínima redundancia Máxima Relevancia, Relief, Random Forest, Information Gain; y métodos de subconjuntos: Selección de Características basada en correlación, Boruta, Filtro Chi-cuadrado. Las características seleccionadas por cada uno de los algoritmos de selección anteriores han sido utilizadas para entrenar los

clasificadores: Máquina de Vectores de Soporte, Red Neuronal Artificial, Regresión Lineal y Naïve Bayes. Se

han entrenado todas las combinaciones de métodos de selección de características – clasificadores mediante validación cruzada de 5 capas, evaluando su rendimiento por medio del área bajo la curva (AUC).

Los resultados han revelado que el mejor método de selección de características basado en subconjuntos ha sido el Boruta (AUC promedio=0.716), y el mejor modelo basado en ranking ha sido Information Gain (AUC promedio=0.717), siendo el clasificador de Regresión Logística el que mejor rendimiento ha obtenido (AUC promedio=0.86).

Los datos utilizados en este trabajo, proceden de un estudio creado en la plataforma ITC-Bio, dicha plataforma ha sido desarrollada por el Grupo de Innovación Tecnológica del HUVR, y consiste en un sistema de gestión de ensayos clínicos que puede aplicarse a cualquier estudio de investigación sobre prevención, diagnóstico y tratamiento de los pacientes. Para implementar estos modelos predictivos en la plataforma, se ha creado una API RestFul, con el objetivo de ofrecer un servicio que permita a los oncólogos obtener la predicción de toxicidad inducida de nuevos pacientes antes del inicio del tratamiento de RT.

Análisis de cohortes de pacientes con enfermedades de base genética utilizando Patient Exploration Tools Suite

Elena Rojano Rivera

Departamento de Biología Molecular y Bioquímica. Facultad de Ciencias. Universidad de Málaga

Artículo de Referencia: Bueno A, Rodríguez-López R, Reyes-Palomares A, Rojano E, Corpas M, Nevado J, Lapunzina P, Sánchez-Jiménez F y Ranea JAG (2018). Phenotype-loci associations in networks of patients with rare disorders: application to assist in the diagnosis of novel clinical cases. European Journal of Human Genetics, 26:1451–1461. <https://doi.org/10.1038/s41431-018-0139-x>.

La caracterización de los síntomas presentados en pacientes con un cuadro clínico poco frecuente complica su correcto diagnóstico. Esto sucede con las enfermedades raras, cuyo tiempo medio en realizar un diagnóstico supera los cinco años. El 80% de estas enfermedades tienen origen genético, y es difícil relacionar las causas genéticas que dan lugar a la manifestación de los fenotipos patológicos en los pacientes debido a su baja frecuencia poblacional. Para comprender las causas que dan lugar a estas enfermedades se pueden aplicar aproximaciones sistémicas que obtengan el máximo provecho cuando el volumen de datos es escaso. En el trabajo en el que se basa este resumen, titulado “Phenotype-loci associations in networks of patients with rare disorders: application to assist in the diagnosis of novel clinical cases” se utiliza una aproximación sistémica basada en el análisis de redes que calculan la asociación entre fenotipos anotados en un sistema jerárquico de términos patológicos (Human Phenotype Ontology, HPO) y regiones del genoma (*loci*) que potencialmente sean su causa a través de datos de pacientes. Sin embargo, existen casos donde la asociación HPO-*loci* no es significativa y esto ocurre con términos poco específicos dentro de la ontología. Para evitar este problema y obtener asociaciones más estadísticamente significativas, se debería realizar una caracterización clínica exhaustiva del paciente, empleando términos que describan su sintomatología lo más precisos posible. Del mismo modo, explotar la información de

estas redes de asociación HPO-*loci* es útil para el desarrollo de herramientas que sirvan de apoyo al diagnóstico clínico de estos pacientes. En este trabajo presentamos la Patient Exploration Tools Suite (PETS), que incluye tres funcionalidades distintas: (1) una herramienta de análisis de cohortes de pacientes que realiza una serie de cálculos estadísticos para mostrar el estado de la misma, (2) un motor de búsqueda de fenotipos patológicos en base a unas coordenadas genómicas de interés, y (3) una herramienta predictiva que, empleando el perfil de fenotipos clínicamente caracterizados en un paciente, priorice las regiones del genoma que potencialmente sean su causa. Del mismo modo, PETS incluye en su herramienta predictiva un control de calidad de los fenotipos disponibles en el perfil del paciente, así como anotaciones funcionales de los genes localizados en las regiones del genoma predichas utilizando términos de la Kyoto Encyclopedia of Genes and Genomes (KEGG). Las funcionalidades de PETS están pensadas para (1) servir como herramienta de apoyo al diagnóstico clínico de pacientes con un perfil fenotípico poco frecuente y mejorarlo en caso de que sea necesario con términos más específicos, (2) determinar qué fenotipos se asocian a regiones del genoma concretas, lo cual sería de utilidad para realizar un diagnóstico prenatal, y (3) predecir las regiones del genoma que potencialmente sean la causa de los fenotipos observados en un paciente y comprender las causas moleculares que dan lugar a los mismos.

Clasificando acciones y conceptos con UMLS en MedLine

Pilar López-Úbeda

Departamento de informática, Centro de Estudios Avanzados en TIC (CEATIC), Universidad de Jaén

Artículo de Referencia: López-Úbeda, P., Díaz-Galiano, M. C., Martín-Valdivia, M. T., & Ureña-López, L. A. (2018). *Sinai en tass 2018 task 3. clasificando acciones y conceptos con umls en medline. Proceedings of TASS, 2172.*

Introducción

En el dominio biomédico, la inteligencia artificial (IA) permite generar conocimiento y mejorar los servicios de salud mediante el procesamiento de información no estructurada, que es una parte importante de los datos recopilados todos los días en la práctica clínica. Con la IA podemos hacer un mejor uso de la información de salud y responder a los nuevos desafíos de registrar, estructurar y explorar información [2].

El *Named Entity Recognition* (NER) [3] es una tarea de extracción de información que busca ubicar y clasificar la información en categorías predefinidas. NER es una técnica importante incluida en el procesamiento del lenguaje natural NLP (*Natural Language Processing*).

Este artículo describe la primera participación del grupo SINAI [7] en la *Tarea 3. eHealth Knowledge Discovery Workshop* [5] en la Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN). Para simplificar y normalizar el proceso de extracción, la tarea general se dividió en tres subtareas:

- Subtarea A: Identificación de frases clave.
- Subtarea B: Clasificación de frases clave.
- Subtarea C: Determinación de relaciones semánticas.

Y para medir los resultados de las subtareas individuales, así como los resultados generales, el desafío eHealth-KD propuso tres escenarios de evaluación:

- Escenario 1. El primer escenario requería que todas las subtareas (es decir, A, B y C) se realizaran secuencialmente.

- Escenario 2. En el segundo escenario solo se realizaron las subtareas B y C.
- Escenario 3. Finalmente, el tercer escenario evaluó solo la subtarea C.

El corpus utilizado se basó en un conjunto de documentos electrónicos españoles sobre salud recopilados de MedlinePlus. MedlinePlus es el sitio web de los *National Institutes of Health's Website* para pacientes y sus familiares y amigos. Producido por la Biblioteca Nacional de Medicina, la biblioteca médica más grande del mundo, brinda información sobre enfermedades, afecciones y problemas que afectan al bienestar en un lenguaje accesible.

Metodología

La metodología realizada consistió en 3 pasos que se detallan a continuación:

Análisis morfológico

Para este paso, se estudió el corpus proporcionado y se detectó que la mayoría de los verbos estaban etiquetados como «acciones» y los sustantivos como «conceptos», utilizando la herramienta CoreNLP [4].

Reconocimiento de entidades con UMLS

Se desarrolló una herramienta de reconocimiento automático de entidades en UMLS [1]. El sistema usa NLP tanto en el texto de entrada como en el texto de los conceptos del diccionario UMLS, normalizando el texto usando herramientas de PNL: normalización, tokenización, stemmer.

El identificador automático de entidad médica devuelve 10 resultados, cada uno de

ellos compuesto por: el concepto detectado en la oración, el código UMLS asociado y el porcentaje de concordancia con el texto.

Términos de varias palabras

Una vez que se reconocen los posibles términos, intentamos localizar más conceptos de varias palabras. Para hacer esto, buscamos oraciones que comienzan y terminan en sustantivo o adjetivo sin que esa frase contenga un verbo dentro. De esta manera, identificamos conceptos en frases como "personas con enfermedad" o "adultos de mayor edad", pero no "niños tienen enfermedades" o "el paciente no toma aspirina".

Después de analizar las frases, calculamos una métrica basada en: el número de palabras, el número de palabras de detención contenidas en la frase de entrada y el concepto devuelto por el sistema basado en UMLS.

Finalmente, verificamos que el concepto detectado contiene más de una palabra y la métrica definida es menor que el porcentaje de éxito devuelto por el sistema.

Resultados

Nuestro grupo solo participó en los Escenarios 1 y 2. Para ambos escenarios se implementaron todas las demás funcionalidades explicadas además de utilizar el *baseline* desarrollado por la organización.

Las medidas de evaluación propuestas por la organización fueron: precisión, retiro y F1. Podemos observar que obtenemos altos valores de precisión, 84% y 90%, respectivamente, aunque decaemos en la exhaustividad por lo que en la medida Micro F1 nos penaliza estos errores.

Conclusiones

En nuestro enfoque llevamos a cabo un análisis simple del texto adaptando los resultados de varias herramientas. Como novedad, creamos una métrica que ayuda a nuestro sistema a seleccionar los conceptos UMLS más importantes en el texto. Esta

métrica es simple de implementar y rápida de ejecutar, lo que nos permite tener un detector de entidades biomédicas con eficiencia moderada y alto rendimiento.

En la clasificación final, los organizadores hicieron el promedio de los tres escenarios y nuestro grupo estaba en segunda posición, aunque no pudimos participar en el Escenario C, por lo que no se obtuvieron resultados muy altos.

Referencias

- [1] Bodenreider, O. (2004). The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research*, 32(suppl_1), D267-D270.
- [2] Cohen, A. M., & Hersh, W. R. (2005). A survey of current work in biomedical text mining. *Briefings in bioinformatics*, 6(1), 57-71.
- [3] Doan, S., Collier, N., Xu, H., Duy, P. H., & Phuong, T. M. (2012). Recognition of medication information from discharge summaries using ensembles of classifiers. *BMC medical informatics and decision making*, 12(1), 36.
- [4] Manning, C., Surdeanu, M., Bauer, J., Finkel, J., Bethard, S., & McClosky, D. (2014, June). The Stanford CoreNLP natural language processing toolkit. In *Proceedings of 52nd annual meeting of the association for computational linguistics: system demonstrations* (pp. 55-60).
- [5] Martínez Cámara, E., Almeida Cruz, Y., Díaz Galiano, M. C., Estévez-Velarde, S., García Cumberras, M. Á., García Vega, M., & Piad-Morffis, A. (2018). Overview of TASS 2018: Opinions, health and emotions.
- [6] Medina Herrera, S., & Turmo Borrás, J. (2018). Joint classification of key-phrases and relations in electronic health documents. In *Proceedings of TASS 2018: Workshop on Semantic Analysis at SEPLN (TASS 2018) co-located with 34nd SEPLN Conference (SEPLN 2018): Sevilla, Spain, September 18th, 2018* (pp. 83-88). CEUR-WS. org.
- [7] López-Ubeda, P., Díaz-Galiano, M. C., Martín-Valdivia, M. T., & Urena-López, L. A. (2018). Sinai en tass 2018 task 3. clasificando acciones y conceptos con umls en medline. *Proceedings of TASS*, 2172.
- [8] Zavala, R. M. R., Martínez, P., & Segura-Bedmar, I. (2018). A Hybrid Bi-LSTM-CRF model for Knowledge Recognition from eHealth documents. In *TASS@ SEPLN* (pp. 65-70).

