

Datos del proyecto:

Título: EXTRacción de Asociaciones entre Enfermedades y otros conceptos médicos (EXTRAE)

Personal:

UNED: Lourdes Araujo (IP), Juan Martinez Romo, Andrés Duque Fernandez, Fernando López Ostenero

ISCIII: Ricardo Sánchez de Madariaga (IP), Montserrat Carmona Rodríguez, Mario Pascual Carrasco, Adolfo Muñoz Carrero

url: <http://nlp.uned.es/extrae/>

Los profesionales de la salud disponen en la actualidad de acceso a la Historia Clínica Electrónica (HCE) de los pacientes. La disponibilidad de información precisa, completa y estructurada permite mejorar muy considerablemente la toma de decisiones. Sin embargo, cada vez es más difícil tomar estas decisiones dado el gran volumen de datos que ha de considerarse. Este volumen dificulta encontrar manualmente relaciones que pueden ser utilizadas en la extracción de conocimiento.

En este proyecto estamos trabajando en el diseño de algoritmos que ayuden a la identificación de relaciones relevantes entre distintas enfermedades. Esta información es muy útil para realizar nuevos diagnósticos, probar nuevos tratamientos o fármacos, o para prever la posible evolución de la enfermedad, etc.

En la actualidad los médicos tienen que basarse en su experiencia para encontrar estas relaciones. El problema se hace prácticamente intratable cuando el especialista quiere abordar no sólo su área de especialización, sino también otras. Por esta razón sería muy útil disponer de un sistema que realice una preselección de las relaciones entre enfermedades y se las proponga a los especialistas en salud, para su consideración.

Muchas enfermedades comparten uno, o varios aspectos, como síntomas, evolución, tratamiento, etc., pero esto no siempre significa que exista una relación entre ellas. Por ello, es necesario detectar relaciones entre enfermedades que se pueden considerar significativas. La significatividad vendrá dada por la coincidencia de aspectos más allá de la casualidad que se capturará definiendo un modelo estadístico apropiado. Las relaciones entre distintas enfermedades se pueden establecer en base a distintos patrones, separada o conjuntamente:

aparición conjunta, síntomas comunes, similitudes de tratamientos, etc. Estas relaciones entre enfermedades se pueden codificar como Reglas de Asociación (RA), que se pueden considerar formas de representar el conocimiento médico subyacente en el conjunto de HCE almacenadas en el repositorio de información clínica. Sin embargo, la extracción de RA no resulta un proceso evidente o inmediato a partir del repositorio de HCE, por dos razones fundamentales.

Primero, como se ha apuntado más arriba, la producción de relaciones entre enfermedades o RA por parte de los algoritmos de Aprendizaje Máquina (AM) no significa que estas relaciones sean significativas, pues pueden haber sido producidas por el ruido o la casualidad.

Por otra parte, el conocimiento médico codificado en forma de RA debe ser ratificado por la experiencia empírica de los profesionales médicos. Este es un proceso extremadamente largo, complejo y costoso, si se tiene en cuenta que estos profesionales disponen de recursos de tiempo y esfuerzo limitados.

Por lo tanto, es imprescindible identificar reglas que tengan la máxima probabilidad a priori de ser válidas y de interés para ser probadas añadidas al corpus de conocimiento médico establecido.

En este proyecto hemos diseñado un nuevo algoritmo semisupervisado (con la mínima intervención médica inicial posible, ya que este es un recurso muy costoso) para producir listas de RA ordenadas por la máxima probabilidad a priori de ser ciertas, de forma que a los profesionales médicos se les presentan únicamente las nuevas RA mejor colocadas en esta ordenación óptima, y así también se optimice el uso de un recurso tan preciado como el tiempo y el esfuerzo de estos profesionales. Este algoritmo que combina técnicas estadísticas no supervisadas y técnicas supervisadas, ha sido capaz de mejorar los resultados de un sistema supervisado entrenado con una cantidad mucho mayor de datos clasificados manualmente.

