

METODOLOGÍA PARA LA APLICACIÓN DE PROCESOS DE EXPLOTACIÓN DE INFORMACIÓN EN ESTABLECIMIENTOS DE SALUD

Lujan, F., Pytel, P., Pollo-Cattaneo, M.F.

Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS)

Universidad Tecnológica Nacional. Facultad Regional Buenos Aires.

Medrano 951 (C1179AAQ) Ciudad Autónoma de Argentina. Buenos Aires Tel +54 11 4867-7511

lujan.facundo@hotmail.com , fpollo@posgrado.frba.utn.edu.ar

RESUMEN

La Explotación de Información es la sub-disciplina de la Informática que provee las herramientas para el análisis y síntesis con el fin de extraer conocimiento no trivial que se encuentra implícito en los datos disponibles en distintas fuentes de información. Este proyecto busca aprovechar las características especiales de dicha disciplina a la hora de procesar el conjunto de datos e historiales clínicos disponibles en Establecimientos de Salud. Para ello, se plantea una metodología que, basada en los procesos de Explotación de Información, provea información útil y oportuna para el diagnóstico y tratamiento de pacientes.

Palabras Clave: *Explotación de Información, Metodología, Toma de Decisiones, Establecimientos de Salud.*

Contexto

Este proyecto articula líneas de trabajo del Grupo de Estudio en Metodologías de Ingeniería de Software (GEMIS) de la Facultad Regional Buenos Aires de la Universidad Tecnológica Nacional (UTN-FRBA) junto con el programa de intercambio UTN-DAAD.

INTRODUCCIÓN

El término Explotación de Información es utilizado para describir una etapa dentro de un proceso mayor llamado *Extracción de Conocimiento en Bases de Datos (Knowledge Discovery in Databases o KDD)*. Lo que en verdad hace la Explotación de Información es reunir las ventajas de varias áreas (estadística, inteligencia artificial, computación gráfica, bases de datos y procesamiento masivo), utilizando principalmente como materia prima las bases de datos. Esta sub-disciplina de la Informática, provee las herramientas para el análisis y síntesis con el fin de extraer conocimiento no trivial, que se encuentra implícito en los datos disponibles en distintas fuentes de información.

Se ha observado en [García-Martínez et al., 2013], el uso indiscriminado de los términos “Minería de Datos” (“Data Mining”, en inglés) y “Explotación de Información” para referirse al mismo cuerpo de conocimientos. Una confusión similar se da al utilizar como sinónimos “Ciencias de la Computación” y “Sistemas de Información” [Pollo-Cattaneo et al., 2012]. La Minería de Datos se relaciona con la tecnología (algoritmos) que se necesita para la obtención de conocimiento y, la Explotación de Información se encuentra más relacionada con los procesos y las metodologías que se

necesitan para realizar el proyecto con éxito. De esta manera, la Minería de Datos está más cerca de las tareas de programación mientras que, la Explotación de Información se encuentra más cercana a Ingeniería de Software.

Una fuente de información en constante expansión son los Establecimientos de Salud. Según la Dirección de Estadísticas e Información en Salud [2000], este término incluye a aquellas organizaciones destinadas al cuidado y a la atención de la salud a través de la prevención y/o el diagnóstico y/o el tratamiento, en una sola ubicación física. Por lo cual están bien equipados con monitores y otros dispositivos de recolección de datos que proveen medios relativamente baratos para recoger y almacenar datos.

Los registros de pacientes obtenidos para el diagnóstico y el pronóstico generalmente abarcan valores de la anamnesis, parámetros clínicos y de laboratorio, así como resultados de los informes particulares, específicos de una tarea dada. Estos datos se caracterizan usualmente por no estar completos (faltan valores de parámetro), estar incorrectos (ruido sistemático o aleatorio en los datos), estar dispersos (los historiales del paciente disponibles son escasos y/o no representables), y ser inexactos (selección inadecuada de parámetros para una tarea dada) [Maimon & Rokach, 2005].

A partir del procesamiento del conjunto de historias clínicas, es posible verificar y obtener: nuevos patrones de desarrollo y propagación de enfermedades, niveles de efectividad de los tratamientos aplicados y relaciones entre factores de riesgo, entre otras relaciones.

En el ámbito de la salud se ha definido como tema crítico la necesidad de obtener información para la toma de decisiones [Bellazzi & Zupan, 2008], y la creciente importancia de un componente

especializado en un sistema de Historia Clínica Electrónica [Luna et al., 2007].

En este contexto, se han aplicado exitosamente algoritmos de Minería de Datos para la predicción de eventos clínicos en pacientes con enfermedades crónicas [González Bernaldo de Quirós et al., 2012], para la evaluación de la eficacia de tratamientos específicos en determinados tipos de cáncer [Reparaz et al., 2008], y para incrementar el nivel de confianza de los diagnósticos médicos mediante la producción de diagnósticos diferenciales [Kuo et al., 2001], entre otras aplicaciones.

Se considera necesario definir una metodología que establezca una serie de pasos ordenados para aplicar, convenientemente, tecnologías de Minería de Datos basada en procesos de EdI [Britos & García Martínez, 2009].

OBJETIVOS E HIPÓTESIS DE INVESTIGACION

Este proyecto pretende valerse de las características especiales de la Explotación de Información a la hora de aprovechar el conjunto de datos e historiales clínicos recolectados en Establecimientos de Salud.

Entre las hipótesis que impulsaron este proyecto se encuentran los siguientes:

- *Hipótesis I:* La información clínica y la eficiencia de los tratamientos médicos puede ser enriquecida por el resultado de aplicar procesos de Explotación de Información
- *Hipótesis II:* Los Establecimientos de Salud poseen datos disponibles, que procesados adecuadamente permitirían realizar exitosamente el diagnóstico y tratamiento de pacientes
- *Hipótesis III:* Es posible generar una metodología que puede ser aplicada por profesionales de la salud con un mínimo conocimiento en Explotación de Información y sus herramientas.

Teniendo en cuenta dichas hipótesis, se tiene como *Objetivo General* plantear una metodología que, aplicando procesos de Explotación de Información, permita obtener patrones de conocimiento para facilitar el diagnóstico y tratamiento de pacientes.

RESULTADOS ESPERADOS

La finalidad de este proyecto es generar un conjunto de pasos que guíe a profesionales de la salud para aplicar procesos de Explotación de Información en la información disponible en Establecimientos de Salud. Por otro lado, se pretende obtener un mejor entendimiento de las causas de enfermedades, conduciendo a un mejor diagnóstico, reconocimiento de patrones de desarrollo de enfermedades endémicas y detección temprana de enfermedades epidémicas. De este modo, se busca maximizar la eficiencia de los tratamientos médicos brindados, y al mismo tiempo, sin postergar la eficacia, minimizar los recursos sanitarios invertidos en las prestaciones médicas.

No obstante, este proyecto contribuye a la prevención sanitaria en todos los niveles, ya que con el nuevo conocimiento adquirido, los Establecimientos de Salud podrán promover el cuidado de la salud con nuevas campañas de prevención, mejores diagnósticos e implementar tratamientos más eficientes con el fin de evitar intervenciones que sean innecesarias dentro del sistema de salud.

Una vez definida la metodología, se considera valiosa su aplicación en ámbitos públicos y privados teniendo en cuenta una posible adecuación de la misma.

FORMACIÓN DE RECURSOS HUMANOS

Este proyecto pretende tanto la obtención de nuevos conocimientos, como la mejora en la formación del recurso humano. El grupo de trabajo se encuentra formado por dos investigadores tesistas de doctorado y un investigador en formación.

Por otra parte, el trabajo propuesto forma parte de un proyecto de investigación que ha sido preseleccionado en el marco del programa de intercambio UTN-DAAD (Deutscher Akademischer Austausch Dienst o Servicio Alemán de Intercambio Académico en español). Dicho programa, tiene como objetivo impulsar a la modernización de la Ingeniería en la República Argentina a través de un acercamiento científico, tecnológico y cultural entre ambos países [Oficina de Relaciones Internacionales, 2010].

REFERENCIAS

- Bellazzi, R., & Zupan, B. (2008). Predictive data mining in clinical medicine: current issues and guidelines. *International journal of medical informatics*, 77(2), 81-97.
- Britos, P. V., & García Martínez, R. (2009). Propuesta de Procesos de Explotación de Información. En XV Congreso Argentino de Ciencias de la Computación.
- Dirección de Estadísticas e Información en Salud (2000). Guía de Establecimientos de la salud. Ministerio de Salud de la Nación Argentina. Disponible en: <http://www.deis.gov.ar/guia.htm>
- García-Martínez, R., Britos, P., & Rodríguez, D. (2013). Information Mining Processes Based on Intelligent Systems. In *Recent Trends in Applied Artificial*

- Intelligence (pp. 402-410). Springer Berlin Heidelberg.
- González Bernaldo de Quirós, F., Luna, D., Baum, A., Plazzotta, F., Otero, C., Benítez, S. (2012). Incorporación de tecnologías de la información y de las comunicaciones en el Hospital Italiano de Buenos Aires.
- Kuo, W. J., Chang, R. F., Chen, D. R., & Lee, C. C. (2001). Data mining with decision trees for diagnosis of breast tumor in medical ultrasonic images. *Breast cancer research and treatment*, 66(1), 51-57.
- Luna, D., Soriano, E., González Bernaldo de Quirós, F. (2007). Historia Clínica Electrónica. *Revista del Hospital Italiano de Buenos Aires, Vol. 27 (Nº 2)*. Disponible en: http://www.hospitalitaliano.org.ar/archivos/noticias_attachs/47/documentos/10347_27-2.Actualizaci%C3%B3n.pdf
- Maimon, O. Z., & Rokach, L. (Eds.). (2005). *Data mining and knowledge discovery handbook* (Vol. 1). New York: Springer.
- Oficina de Relaciones Internacionales (2010). Programa de Intercambio UTN-DAAD. Universidad Tecnológica Nacional. Disponible en: http://www.utn.edu.ar/programasalemania/05_internacionales.utn
- Pollo-Cattaneo, M., García-Martínez, R., Britos, P., Pesado, P., Bertone, R., Rodríguez, D. & Vanrell, J. (2012). Elementos para una Ingeniería de Explotación de Información. *Proyecciones*, 10(1), 67-84.
- Reparaz, D., Merlino, H., Rancan, C., Rodríguez, D., Britos, P. V., & García Martínez, R. (2008). Determinación de la eficacia de la braquiterapia en tratamiento de cáncer basada en minería de datos.
- In X Workshop de Investigadores en Ciencias de la Computación.