

Una propuesta metodológica para Adaptive Business Intelligence

Marcelo Fabio Roldán¹, Germán Montejano², Ana Funes²

¹Departamento de Ciencias Exactas, Físicas y Naturales –
Universidad Nacional de La Rioja
La Rioja – Argentina

marcelofabio01@yahoo.com.ar

²Universidad Nacional de San Luis, Ejército de los Andes 950
5700 San Luis, Argentina
{gmonte, afunes}@unsl.edu.ar

Resumen

Las etapas metodológicas para el desarrollo de aplicaciones basadas en Adaptive Business Intelligence [MIC07] [IEE07] [MOS03] abarcan la comprensión del problema, de los datos, de su preparación, modelado, búsqueda para acercarse a los objetivos e implementación a través de una aplicación de negocios.

La interpretación correcta de estas etapas para el desarrollo de aplicaciones autoadaptativas basadas en técnicas de Business Intelligence, dan lugar al uso óptimo de métodos de predicción y técnicas de optimización, los cuales aplicados a un problema de referencia, dan como resultado parámetros de mayor o menor precisión, error, nivel de confianza, consumo de recursos o tiempo de entrenamiento. [ZBI06]

En la presente línea de investigación nos proponemos la formulación de una nueva metodología a partir del mapeo de las principales etapas para la solución de problemas, tomando como referencia Adaptive Business Intelligence, en donde a través de diversos casos de estudio de la medicina se pretende encontrar patrones metodológicos comunes.

Contexto

El presente trabajo de investigación se encuentra enmarcado en una colaboración entre investigadores del Proyecto de Incentivos código 22/F822 “Ingeniería de Software: Conceptos, Métodos y Herramientas en un contexto de Ingeniería de Software en Evolución”, de la Universidad Nacional de San Luis e investigadores del Departamento de Ciencias Exactas, Físicas y Naturales de la Universidad Nacional de La Rioja.

Esta propuesta de investigación se desarrolla como una tesis de la Maestría en “Ingeniería de Software” de la Universidad Nacional de San Luis.

Palabras clave: Minería de Datos, Adaptive Business Intelligence, metodologías de extracción de conocimiento.

Introducción

El Business Intelligence se nutre de grandes volúmenes de datos para permitir que los usuarios y aplicaciones accedan a ellos y los usen para realizar toma de decisiones. [WAT07] [AYA06] [REI00] Es así que metodologías como CRISP-DM [CHA00] [GUT07] [AZE08], SEMMA [SAS11] [TAA11] y, en general, el KDD Process

[ASE04], nos introducen al proceso utilizado para alcanzar tal objetivo por medio del uso de técnicas de Minería de Datos.

Ante este escenario, la definición de una metodología, que concentre y sintetice las virtudes de estas metodologías, facilita la producción de conocimiento a partir del aprendizaje y del descubrimiento, en especial cuando se trata de aquellas áreas de las ciencias, como las médicas, farmacéuticas, génicas o similares, que conceptualmente se pueden considerar como alejadas del espíritu del Adaptive Business Intelligence.

En este trabajo abarcaremos el dominio de los problemas de predicción, incluyendo la regresión, la clasificación y el análisis de series temporales, y su configuración dentro de una metodología, que facilite la aplicación de la técnica de minería de datos más recomendable en base a la experiencia y conocimientos previos sobre la materia. [PAL05]

Para ilustrar y demostrar el comportamiento de la metodología, como así también describir el trabajo metodológico interdisciplinario, se han elegido dos aplicaciones médicas diferentes. Una relativa a la sepsis intrahospitalaria y otra al hipotiroidismo.

Buscamos de esta forma facilitar el desarrollo informático y, mediante la aplicación de esta metodología en un trabajo interdisciplinario, demostrar su eficacia en un marco orientado a herramientas autoadaptativas, usando problemas reales como eje ilustrativo de su aplicación.

Resultados y Objetivos

Este trabajo de investigación tiene como objetivo principal definir una metodología, basada en las estructuras conceptuales de Adaptive Business Intelligence.

El trabajo es llevado a cabo a partir de casos de estudio, y de bibliografía de

referencia que permitan facilitar el uso de métodos de predicción orientados en particular a problemas médicos y afines disciplinariamente tales como biotecnología, genómica, cuidado de la salud, ciencia y farmacéutica médica [BIO11].

Asimismo, se pretende aplicar la metodología desarrollada a problemas reales como eje demostrativo.

Para su logro, se considera cubrir los aspectos relacionados con los aportes metodológicos existentes, considerando para ello tres metodologías: CRISP-DM [ARA11], KDD Process [CIO07] y SEMMA [SAS11], las cuales servirán como referencia para la determinación de los pasos fundamentales a seguir cuando se trate de aplicaciones médicas, siempre en el entorno de Adaptive Business Intelligence y con énfasis en la búsqueda de conocimiento orientado a problemas de clasificación, regresión o series temporales.

Se espera que como resultado del seguimiento de los pasos metodológicos propuestos, se proporcione, además, un software aplicativo que permita ejemplificar el comportamiento de la metodología con un producto final funcional cuyo aporte a la medicina sea válido.

Para esto, hemos de plantear conceptos sintéticos de las diferentes técnicas de predicción y métodos de optimización, con una breve introducción en los aspectos exploratorios y de pre procesamiento de la información, ya que el conocimiento de estos corresponde mayormente al dominio del especialista médico [VAI07].

Se realizará una caracterización sobre las metodologías existentes orientada a determinar aquellos datos significativos a la investigación, identificando los de mayor relevancia y reduciendo las variables que sumen complejidad innecesaria, buscando permanentemente patrones metodológicos comunes a ellas [AZE08].

En un paso posterior a esto, se buscará construir el método que facilite la

administración del flujo de trabajo para el desarrollador de aplicaciones en el dominio de los problemas de predicción.

Para el desarrollo de la aplicación autoadaptativa de diagnóstico de sepsis, se recabarán datos de historias clínicas de pacientes (antecedentes hereditarios, signos clínicos, valores de los exámenes sanguíneos y otros) a través de un sistema informático y en el segundo caso (diagnóstico de estadio de hipotiroidismo) [MDL11] [HAL09] se utilizarán las muestras recogidas por un centro de salud disponible en un banco de datos.

Líneas de Investigación y Desarrollo

La línea de investigación aquí presentada, la cual está siendo llevada adelante entre investigadores del proyecto de incentivos de la Universidad Nacional de San Luis, código 22/F822: “Ingeniería de Software: Conceptos, Métodos y Herramientas en un Contexto de Ingeniería de Software en Evolución” e investigadores de la Universidad Nacional de La Rioja, se trata de una nueva línea para la ejecución de una tesis de Maestría en Ingeniería de Software de la Universidad Nacional de San Luis.

Presentamos acá los lineamientos generales para esta nueva línea de investigación, en la que nos encontramos abocados a elaborar una propuesta metodológica sobre la base de Adaptive Business Intelligence, enfocado a un dominio de aplicación diferente, el de las ciencias médicas.

Formación de Recursos Humanos

El presente trabajo es la base para el desarrollo de una Tesis de Maestría en Ingeniería de Software de la Universidad Nacional de San Luis. La misma se encuentra en etapa de desarrollo, llevada a delante por un alumno de la Universidad Nacional de la Rioja, y siendo dirigida por

investigadores de la Universidad Nacional de San Luis, bajo un acuerdo entre ambas universidades.

Referencias

[ARA11] J. G. Arancibia. Metodología para el Desarrollo de Proyectos en Minería de Datos CRISP-DM.

URL(<http://yoshibauco.wordpress.com/>) Último acceso: 2011

[ASE04] V.V.Asencios. Data Mining y el Descubrimiento del Conocimiento. Revista de la Facultad de Ingeniería Industrial. Vol(7). 2004

[AYA06] Alejandro Peña Ayala. Inteligencia de Negocios: Una Propuesta para su Desarrollo en las organizaciones. Instituto Politécnico Nacional. Dirección de Publicaciones.. 2006

[AZE08] A. Azevedo. M.F.Santos. KDD, SEMMA y CRISP-DM: A Parallel Overview. IADIS European Conference Data Mining 2008 (part of MCCSIS 2008)

[BIO11] The open access publisher. URL(<http://www.biomedcentral.com/>). Último acceso: 2011

[CHA00] Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer and Rüdiger Wirth. CRISP-DM 1.0 Step-by-step data mining guide, NCR Systems Engineering Copenhagen (USA and Denmark), DaimlerChrysler AG (Germany), SPSS Inc. (USA) and OHRA Verzekeringen en Bank Groep B.V (The Netherlands). Copyright © 1999, 2000

[CIO07] K.J.Cios, W.Pedrycz, R.W.Swiniarski, L.A.Kurgan. Data Mining. A Knowledge Discovery Approach. Springer. 2007

[GUT07] P-Chapman. J. Clinton. R. Kerber. T. Khabaza. T. Reinartz. C. Shearer. R. Wirth. Guía paso a paso de Minería de Datos.

[HAL09] Mark Hall, E. Frank, G. Holmes, B. Pfahringer. The WEKA Data Mining Software: An Update. Pentaho Corporation. 2009

[IEE07] Watson, H.J.; Wixom, B.H. The Current state of the Business Intelligence.

Univ. of Georgia, Athens. 2007. Último acceso 2011 <http://ieeexplore.ieee.org/>

[MIC07] Z. Michalewicz, M. Schmidt, M. Michalewicz, C. Chiriac. Adaptive Business Intelligence. Springer 2007

[MLD11] Machine Learning Data Set Repository. MLdata.

<http://mldata.org/repository/data/viewslug/datasets-uci-hepatitis/> Último acceso 2011

[MOS03] Larissa T. Moss, Shaku Atre. Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications. Addison Wesley. 2003

[PAL05] Pal, N.R., Jain, L.C. Advanced Techniques in Knowledge Discovery and Data Mining, Springer Verlag. (Eds.) 2005.

[REI00] Joerg Reinschmidt, Allison Francoise. Business Intelligence Certification Guide IBM. International Technical Support Organization. 2000

[SAS11] SAS Institute. United Kingdom. <http://www.sas.com/offices/europe/uk/technologies/analytics/datamining/miner/semma.html> Último acceso: 2011

[TAA11] Academic Business Intelligence System Development using SAS Tools. A. Ta'a, Universiti Utara Malaysia, Kedah, Malaysia. SAS Global Forum 2008.

[VAI07] S. Vaidya L.C. Jain H. Yoshida (Eds.) Advanced Computational Intelligence Paradigms in Healthcare-2. Springer 2007

[WAT07] Hugh J. Watson. Barbara H. Wixom. The Current State of Business Intelligence. University of Georgia. University of Virginia. 2007. IEEE Computer Society
www.computer.org

[ZBI06] Adaptive Business Intelligence , Zbigniew Michalewicz, Springer 2006