

MINERIA DE DATOS EN INTELIGENCIA DE NEGOCIOS

García-Martínez, R., Britos, P., Diez, E., Ochoa, M. Merlino, H., Fernández, E., Rancán, C.

Centro de Ingeniería del Software e Ingeniería del Conocimiento
Escuela de Postgrado. Instituto Tecnológico de Buenos Aires
Tel: 54-11-6393-4843 - E-mail: rgm@itba.edu.ar
Web: <http://www.itba.edu.ar/capis>

1. Objetivos del Proyecto

Este proyecto tiene por objetivos: [a] definir un marco teórico que permita determinar de qué manera, los sistemas inteligentes puede utilizarse para descubrir automáticamente descripciones cualitativas de una base de datos que mejoren la calidad de la inteligencia de negocio, [b] determinar la calidad (por contraste con expertos del dominio) del conocimiento cualitativo inferido automáticamente y [c] desarrollar técnicas y fases metodológicas que mejoren el proceso de minería de datos con énfasis en la inteligencia de negocio.

2. Resultados

Los resultados del proyecto pueden agruparse en tres áreas: aspectos metodológicos, fundamentos y aplicaciones.

2.1. Aspectos Metodológicos de la Minería de Datos en Inteligencia de Negocios

Dentro de los aspectos metodológicos se ha trabajado en: la definición de técnicas de modelado centradas en el entendimiento del negocio [Ochoa, 2006; Ochoa *et al*, 2006], la definición de métodos de preprocesamiento de datos orientado al uso de explotación de información basado en sistemas inteligentes [Merlino, 2004; Merlino *et al*, 2005], la especificación, diseño y desarrollo de un gestor metodológico para proyectos de minería de datos [Fernández *et al*, 2005; Fernández, 2006] y el desarrollo de una metodología para la selección de herramientas de minería de datos [Britos *et al*, 2006].

2.2. Fundamentos de la Minería de Datos en Inteligencia de Negocios

En el área de fundamentos se ha trabajado en la integración metodológica de herramientas inteligentes para explotación de información [Ochoa, 2004; Ochoa *et al*, 2006], en el estudio de herramientas de obtención de sub-óptimos basada en algoritmos genéticos [Cottone, 2004], en el análisis de clasificadores bayesianos [Fernández, 2004], en el nivel de significación estadística del aprendizaje en estos [Césari, 2006] y el estudio comparativo de distintos algoritmos [Groppo Parisi, 2006], en la detección automática de reglas de asociación [Cartagenova, 2005]; y en ambientes de integración: de herramientas para exploración de datos centrados en la web [Merlino, 2006] y sistemas basados en conocimiento y de descubrimiento [Rancán *et al*, 2006; 2007].

2.3. Aplicaciones de la Minería de Datos en Inteligencia de Negocios

Dentro de las aplicaciones de la minería de datos en inteligencia de negocios se ha explorado el uso en: predicción de ventas basada en redes neuronales [Nojec *et al*, 2003], análisis de los deudores del sistema financiero [Sal, 2005], identificación de patrones de caras humanas [Britos *et al*, 2005], detección en el patrón de consumo de usuarios de un servicio [Britos *et al*, 2006], en la identificación de variaciones de cubos sísmicos [Cersosimo *et al*, 2006], a análisis de variables agrometeorológicas [Cogliati *et al*, 2006a; 2006b; 2006c], diagnóstico de calidad de datos en sistemas de mantenimiento [Cuello *et al*, 2006], detección de patrones en imágenes médicas [Ferrero 2006; Ferrero *et al*, 2006], identificación de hábitos de uso de sitios web utilizando SOM [Martinelli *et al*, 2006], en la detección de patrones delictivos en Argentina [Perversi, 2007] y en la estratificación de afiliados en una compañía de medicina prepaga con el objeto de determinar grupos de alto impacto económico y patrones de indeseabilidad para la selección de afiliados [Abatti, 2007].

3. Líneas de Trabajo

A la fecha de la presentación de esta comunicación se continúa trabajando en: la integración de sistemas basados en conocimiento y de descubrimiento y su aplicación a la toma de decisiones estratégicas operacionales [Rancán *et al*, 2006] y en la cartografía de textos centrada en métodos iconográficos de observación, exploración y comunicación aplicados a la minería de textos [Césari, 2007].

4. Formación de Recursos Humanos

A la fecha de esta comunicación se han radicado en el Proyecto: dos tesis de doctorado, siete tesis de magíster, siete trabajos finales de especialidad, tres tesis de grado en ingeniería y los planes de investigación de cuatro docentes.

5. Referencias

- Abatti, A. 2007. *Aplicación de Redes Neuronales para la Estratificación de Afiliados en una Compañía de Medicina Prepaga con el Objeto de Determinar Grupos de Alto Impacto Económico y Patrones de Indeseabilidad para la Selección de Afiliados*. Proyecto de Tesis de Grado en Ingeniería Industrial. Instituto Tecnológico de Buenos Aires.
- Britos, P., Abasolo, M., García-Martínez, R. y Perales, F. 2005. *Identification of MPEG-4 Patterns in Human Faces Using Data Mining Techniques*. Proceedings 13 th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision'2005. Pág. 9-10.
- Britos, P., Grosser, H., Sierra, H. y García Martínez, R. 2006. *Unusual Changes in Consumption Detection in Mobile Phone Users..* Research in Computing Science Journal, 21: 195-204.
- Britos, P., Merlino, H., Fernández, E., Ochoa, M., Diez, E. y García Martínez, R. 2006. *Tool Selection Methodology in Data Mining*. Proceedings V Ibero-American Symposium on Software Engineering. Pág. 85-90.

- Cartagenova, S. 2005. *Detección Automática de Reglas de Asociación*. Trabajo Final de Especialidad en Ingeniería de Sistemas Expertos. Escuela de Postgrado. Instituto Tecnológico de Buenos Aires.
- Cersosimo, S., Ravazzoli, C., García-Martínez, R. 2006. *Identification of Velocity Variations in a Seismic Cube Using Neural Networks*. IFIP International Federation for Information Processing, Volume 218, ISSN 1571-5736, Professional Practice in Artificial Intelligence, eds. J. Debenham, (Boston: Springer), pp. 11-19.
- Césari, M. 2006. *Nivel de Significación Estadística para el Aprendizaje de una Red Bayesiana*. Trabajo Final de Especialidad en Tecnologías de Explotación de Información. Escuela de Postgrado. Instituto Tecnológico de Buenos Aires.
- Cesari, M. 2007. *Cartografía de Textos. Métodos Iconográficos de Observación, Exploración y Comunicación Aplicados a la Minería de Textos*. Proyecto de Tesis de Magister en Ingeniería del Software. Escuela de Postgrado. Instituto Tecnológico de Buenos Aires.
- Cogliati, M., Britos, P. y García-Martínez, R. 2006a. *Análisis de Variables Agrometeorológicas en Noches de Heladas u tilizando Mapas Auto Organizados y Algoritmos de Inducción*. XI Reunión Argentina de Agrometeorología. Facultad de Ciencias Agrarias y Forestales. Universidad Nacional de La Plata.
- Cogliati, M., Britos, P., García-Martínez, R. 2006b. *Análisis del Viento en el Valle del Río Negro Mediante Mapas Auto Organizados y Algoritmos de Inducción*. Avances en Energías Renovables y Ambiente Vol. 10. Pág. 11.137-11.144.
- Cogliati, M., Britos, P., García-Martínez, R. 2006c. *Patterns in Temporal Series of Meteorological Variables Using SOM & TDIDT*. IFIP International Federation for Information Processing, Volume 217, ISSN 1571-5736, Artificial Intelligence in Theory and Practice, ed. M. Bramer, (Boston: Springer), pp. 305-314.
- Cottone, A. 2004. *Estudio de Una Herramienta de Obtención de Sub-óptimos Basada en Algoritmos Genéticos*. Trabajo Final de Especialidad en Ingeniería de Sistemas Expertos. Escuela de Postgrado. Instituto Tecnológico de Buenos Aires.
- Cuello, G., Britos, P. y García-Martínez, R. 2006. *Diagnóstico de Calidad de Datos en Sistemas de Mantenimiento. Comparación entre Mapas de Karnaught y Algoritmos de Inducción*. Memorias de las Iras. Jornadas de Calidad en Mantenimiento e Inspección del Instituto Argentino del Petroleo y Gas.
- Fernández, E. 2004. *Análisis de Clasificadores Bayesianos*. Trabajo Final de Especialidad en Ingeniería de Sistemas Expertos. Escuela de Postgrado. Instituto Tecnológico de Buenos Aires.
- Fernández, E. 2006. *Asistente para la Gestión de Documentos de Proyectos de Explotación de Datos*. Tesis de Magister en Ingeniería del Software. Escuela de Postgrado. Instituto Tecnológico de Buenos Aires.
- Fernández, E., Merlino, H., Ochoa, M., Diez, E., Britos, P. y García-Martínez, R. 2005. *Gestión Asistida de Documentos. en una Metodología de Explotación de Información*. Proceedings del II Workshop de Ingeniería de Software y Bases de Datos del XI Congreso Argentino de Ciencias de la Computación. Pág. 504-514.
- Ferrero, G. 2006. *Detección de Patrones en Imágenes Médicas*. Tesis de Magister en Ingeniería del Software. Escuela de Postgrado. Instituto Tecnológico de Buenos Aires.
- Ferrero, G., Britos, P., García-Martínez, R. 2006. *Detection of Breast Lesions in Medical Digital Imaging Using Neural Networks*. IFIP International Federation for Information Processing, Volume 218, ISSN 1571-5736, Professional Practice in Artificial Intelligence, eds. J. Debenham, (Boston: Springer), pp. 1-10.

- Grosso Parisi, T. 2006. *Learning Bayesian Networks Skeleton: A Comparison Between TPDA and PMMS Algorithm*. Tesis de Grado en Ingeniería Industrial. Instituto Tecnológico de Buenos Aires (En cooperación con la Université Claude Bernard Lyon I).
- Martinelli, D., Merlino, H., Britos, P., García-Martínez, R. 2006. *Identificación de Hábitos de Uso de Sitios Web Utilizando SOM*. Reportes Técnicos en Ingeniería del Software, 8(1): 17-24 .
- Merlino, H. 2004. *Un Método de Preprocesamiento de Datos Orientado al Uso de Explotación de Información Basado en Sistemas Inteligentes*. Trabajo Final de Especialidad en Ingeniería de Sistemas Expertos. Escuela de Postgrado. Instituto Tecnológico de Buenos Aires.
- Merlino, H. 2006. *Ambiente de Integración de Herramientas para Exploración de Datos Centrados en la Web*. Tesis de Magister en Ingeniería del Software. Escuela de Postgrado. Instituto Tecnológico de Buenos Aires.
- Merlino, H., Britos, P., Ierache, J., Diez, E. y García-Martínez, R. 2005. *Un Método de Transformación de Datos Orientado al Uso de Explotación de Información*. Proceedings del II Workshop de Ingeniería de Software y Bases de Datos del XI Congreso Argentino de Ciencias de la Computación. Pág. 22-32.
- Nojec, S., Britos, P., Brossi, B., García Martínez, R. 2003. *Pronóstico de Ventas: Comparación de Predicción entre Redes Neuronales y Métodos Estadísticos*. Revista Eletrônica de Ciência Administrativa. Volumen 2 N° 1 ISSN: 1677-7387.
- Ochoa, A. 2004. *Herramientas Inteligentes para Explotación de Información*. Trabajo Final de Especialidad en Ingeniería de Sistemas Expertos. Escuela de Postgrado. Instituto Tecnológico de Buenos Aires.
- Ochoa, A. 2006. *Uso de Técnicas de Educación para el Entendimiento del Negocio*. Tesis de Magister en Ingeniería del Software. Escuela de Postgrado. Instituto Tecnológico de Buenos Aires.
- Ochoa, M., Britos, P. y García-Martínez, R. 2006. *Una Prototipo de Entendimiento del Negocio para Metodologías de Desarrollo de Sistemas*. XII Congreso Argentino de Ciencias de la Computación. San Luis. Argentina.
- Perversi, I. 2007. *Aplicación de Minería de Datos para la Exploración y Detección de Patrones Delictivos en Argentina*. Proyecto de Tesis de Grado en Ingeniería Industrial. Instituto Tecnológico de Buenos Aires.
- Rancán, C., Pesado, P. y García-Martínez, R. 2006. *Hacia una Propuesta Integradora de Sistemas Basados en Conocimiento y de Descubrimiento*. XII Congreso Argentino de Ciencias de la Computación. San Luis. Argentina.
- Rancán, C., Pesado, P. y García-Martínez, R. 2007. *Toward Integration of Knowledge Based Systems and Knowledge Discovery Systems*. Journal of Computer Science & Technology, 7(1): 91-97.
- Sal, E. 2005. *Análisis de los Deudores del Sistema Financiero a través de la Explotación de Datos*. Trabajo Final de Especialidad en Ingeniería de Sistemas Expertos. Escuela de Postgrado. Instituto Tecnológico de Buenos Aires.