

Data Mining utilizando Sistemas Inteligentes

Docentes

Lic. Laura Lanzarini (UNLP)

Dr. José Ángel Olivas Varela (Universidad de Castilla La Mancha - España)

Objetivo

Presentar los principios básicos de Redes Neuronales y Algoritmos Evolutivos aplicados a la Minería de Datos. El énfasis estará en la búsqueda eficaz sobre la WEB y en la resolución de problemas de predicción y de agrupamiento (clustering).

Destinatarios

Licenciados en Informática, licenciados en Sistemas, ingenieros de Sistemas y profesionales que trabajen en aplicaciones que requieren Minería de Datos utilizando sistemas inteligentes.

Motivación

Se denomina proceso de extracción de conocimiento o KDD (Knowledge Discovery from Databases) al análisis y modelización de la información disponible con el fin de adquirir conocimiento nuevo, novedoso y potencialmente útil. Este proceso posee una solución tradicional manual donde el análisis e interpretación de la información disponible es realizada por especialistas. Esta forma de actuar es lenta, cara y altamente subjetiva.

La Minería de Datos es una de las etapas más importantes de este proceso que se caracteriza por no suministrar información por extensión (datos) sino por comprensión (conocimiento) y además, el conocimiento no es, generalmente, una parametrización de ningún modelo preestablecido o intuitivo por el usuario, sino que es un modelo novedoso y original extraído completamente por la herramienta.

Resultados de la minería de datos son, por ejemplo, conjuntos de reglas, árboles de decisión, redes neuronales o grafos probabilísticos, los cuales pueden utilizarse para responder cuestiones tales como ¿cuál es el perfil de los clientes que más benefician a mi compañía y qué productos conviene ofrecerles? ¿qué medicación debería suministrarse para esta patología específica?, ¿esta transacción debería considerarse fraudulenta?

La minería de datos utiliza distintas técnicas para alcanzar sus objetivos entre las que se destacan los árboles, las reglas de asociación y clasificación, las redes neuronales y las técnicas de optimización.

Las Redes Neuronales son sistemas conexionistas que se adaptan a la información disponible emulando la manera en que los seres humanos aprendemos. Según el tipo de arquitectura neuronal que se utilice pueden tener aplicaciones muy diversas. Pueden usarse para el reconocimiento de patrones, la comprensión de la información, la reducción de la dimensionalidad, el agrupamiento, la clasificación y la visualización de la información disponible, entre otros.

Otro enfoque muy utilizado lo constituyen las técnicas de optimización basadas en Metaheurísticas poblacionales. En esta línea las más utilizadas son los Algoritmos Evolutivos y la optimización basada en cúmulo de partículas (PSO – Particle Swarm Optimization).

Los primeros trabajan sobre el concepto de población formada por un conjunto de soluciones que mejoran durante el proceso evolutivo. Dicho proceso emula lo afirmado por la Teoría Darwiniana y por lo tanto utiliza operadores genéticos como estrategia de adaptación.

Por su lado, PSO es una técnica de optimización estocástica que ha sido utilizada exitosamente en la resolución de una amplia gama de problemas de optimización, incluyendo el entrenamiento de redes neuronales y la minimización de funciones. Se espera que a partir de una presentación general de estos temas se analice su participación en la resolución de problemas concretos. En particular se analizará el caso de la búsqueda eficiente en la WEB, de gran importancia en la Informática actual.

Data Mining utilizando Sistemas Inteligentes

Programa

Minería de Datos

- Introducción. Fases del proceso de extracción del conocimiento. Relación con otras disciplinas.
- Recuperación de información vs recuperación de datos.
- Proceso de recuperación de información.
- Preparación de los datos. Recopilación. Limpieza. Exploración y Selección.
- Técnicas de Minería de Datos. Métodos basados en una medida de similitud. Agrupamiento y Clasificación. Árboles de decisión y Sistemas de Reglas Reglas de asociación

Redes Neuronales

- Definición. Conexiones. Elementos de procesamiento. Función discriminante.
- Redes Neuronales feedforward. Resolución de problemas de predicción.
- Red neuronal multi-perceptrón. Algoritmo de entrenamiento backpropagation. Ejemplos.
- Redes neuronales competitivas. Entrenamiento no supervisado.
- Resolución de problemas de clustering utilizando mapas auto-organizativos estáticos y dinámicos.

Algoritmos Evolutivos

- Paradigmas Principales. Programación Evolutiva, Estrategias Evolutivas y Algoritmos Genéticos. Comparaciones entre los distintos paradigmas. Terminología. Introducción. Conceptos Básicos. Representación.
- Técnicas de Selección y Cruce.
- Ejemplo: Reglas de clasificación obtenidas a través de un algoritmo genético.

Búsqueda eficaz en la WEB

- Herramientas de búsqueda en Internet (se presentan diversos ejemplos ilustrativos).
 - ✓ Directorios.
 - ✓ Buscadores.
 - ✓ Meta-Buscadores.
 - ✓ Sistemas de pregunta-respuesta.
- El usuario y los buscadores.
 - ✓ Tipos de consultas.
 - ✓ Qué buscan los usuarios.
 - ✓ ¿Por qué buscan los usuarios?
 - ✓ Comportamiento del usuario como realimentación.
- Examen de Google, Yahoo! y Live search.
 - ✓ Operadores de búsqueda.
 - ✓ Resultados.
- Problemas en la recuperación de información, cómo afrontarlos.

Data Mining utilizando Sistemas Inteligentes

Bibliografía Básica

- Baeza-Yates, R., Ribeiro-Neto, B. Modern Information Retrieval. Addison-Wesley, 1999.
- Grossman, D. A., Frieder, O. Ad Hoc Information Retrieval: Algorithms and Heuristics. Kluwer. Academic Publishers, 1998.
- Kowalski, G., Maybury, M., Information Storage and Retrieval Systems. Theory and Implementation, 2nd. Edition. Kluwer Academic Publishers, 1997.
- Salton, G. Automatic Text Processing. Addison Wesley, 1989.
- Witten, I. H., Moffat, A., Bell, T.C., Managing Gigabytes: Compressing and Indexing Documents and Images, 2nd. Edition. Morgan Kaufmann, 1999. Cambridge MA, 1991.
- Hernández Orallo, Ramírez Quintana, Ferri Ramírez. *Introducción a la Minería de Datos*. Prentice Hall. 2004. ISBN84-205-4091-9.
- Sierra Araujo. *Aprendizaje Automático: conceptos básicos y avanzados*. Aspectos prácticos utilizando el software WEKA. Prentice Hall. 2006. ISBN 84-8322-318-X
- Isasi y Galván. *Redes Neuronales Artificiales. Un enfoque práctico*. Pearson. Prentice Hall. 2004. ISBN 84-205-4025-0.
- Kohonen, T. *Self-Organizing Maps*. 2nd Edition. Springer. ISSN 0720-678X. 1997.
- Alahakoon, D., Halgamuge, S.K. & Srinivasan, B. *Dynamic Self-Organizing Maps with Controlled Growth for Knowledge Discovery*. IEEE Transactions On Neural Networks, Vol. 11 (3), pp. 601-614. 2000