

Extracción Automática de Conocimiento en Bases de Datos e Ingeniería del Software

<http://www.dsic.upv.es/~jorallo/docent/doctorat/index.html>

José Hernández Orallo

jorallo@dsic.upv.es

M^a. José Ramírez Quintana

mramirez@dsic.upv.es

Cèsar Ferri Ramírez

cferri@dsic.upv.es

Programas:

- Máster en Ingeniería del Software, Métodos Formales y Sistemas de Información.
- Máster en Inteligencia Artificial, Reconocimiento de Formas e Imagen Digital.
- Programa de Doctorado: Programación Declarativa e Ingeniería de la Programación

Objetivos (1/2)

- Presentar el problema del análisis inteligente y automático de la información para el descubrimiento de conocimiento útil.
- Presentar las técnicas de aprendizaje automático más habituales y conocer la idoneidad de cada una para diferentes problemas.
- Reconocer la existencia de técnicas inductivas de alto nivel, especialmente las declarativas, que permiten obtener modelos complejos (relacionales y/o recursivos) pero comprensibles, a partir de los datos y de conocimiento previo.

Objetivos (2/2)

- Particularizar las técnicas vistas para las necesidades específicas de la extracción de conocimiento en bases de datos (KDD), y, en concreto, para la minería de datos (*data-mining*).
- Adaptar la extracción automática de conocimiento a otras fuentes de información no estructurada (textos y *web*) y semiestructurada (XML).
- Conocer las posibilidades de desarrollo semi-automático, combinando técnicas inductivas y deductivas, de sistemas *software* para el tratamiento de la información.

Temario (1/4)

1. Introducción.

1.1. El Problema de la Extracción Automática de Conocimiento.

1.2. Técnicas de Aprendizaje Automático.

1.3. Evaluación de Hipótesis

1.4. Inducción Declarativa y Árboles de Decisión

Temario (2/4)

2. Minería de Datos y Extracción de Conocimiento de Bases de Datos

2.1. Nuevas Necesidades de Análisis de Grandes Sistemas de Datos. Los Data-Warehouse.

2.2. El Proceso de Extracción de Conocimiento de Bases de Datos.

2.3. Métodos Específicos de Minería de Datos.

2.4. Técnicas Declarativas para Prospección de Datos.

2.5. Bases de Datos y Lenguajes de Consulta Inductivos.

Temario (3/4)

3. Extracción de Conocimiento a Partir de Información No Estructurada

- 3.1. Los Problemas de la Información No Estructurada.
- 3.2. Extracción de Conocimiento a partir de Documentos HTML y texto.
- 3.3. Extracción de Información semi-estructurada (XML).
- 3.4. Asistentes Web Inteligentes.
- 3.5. Personalización

Temario (4/4)

4. Programación Inductiva e Ingeniería del Software

- 4.1. La Inducción en la Generación de Software.
- 4.2. Aplicaciones del Aprendizaje Automático para el Desarrollo de Software.
- 4.3. Inducción Automática de Requerimientos.
- 4.4. Modelado (Semi-)Automático.
- 4.5. Aplicaciones a Otras Fases del Ciclo de Vida.

Bibliografía (1/9)

APRENDIZAJE AUTOMÁTICO (Generales):

- Bishop, Christopher M. “Neural networks and machine learning” Springer, 1998. [4-63/834B, 4-63/835B, D-SIC/2873D, D-SIC/2895D].
- Laird, P.D. “Learning from good and bad data” Kluwer Academic, 1988. [4-63/87B][4-63/88B][4-63/27F][4-63/28F]
- Langley, P. “Elements of Machine Learning”. San Francisco: Morgan Kaufmann Publishers Inc, 1996.
- Michie, D.; Spiegelhalter, D.; Taylor, C. (eds.) “Machine Learning, Neural and Statistical Classification” Ellis Horwood, 1994.
- **Mitchell, T.M. “Machine Learning” McGraw-Hill 1997. [4-63/102B, 4-63/34F, 4-63/35F] [*]**
- Osherson, D.N.; Stob, M.; Weinstein, S. “ Systems that learn : An introduction to learning theory for cognitive and computer scientists”, Cambridge, Bradford, 1990. [4-63/95B] [4-63/96B] [4-63/9F]
- **Thornton, Chris “Truth from Trash. How Learning Makes Sense” The MIT Press (A Bradford Book), 2000. [4-63/897B] [*]**

Bibliografía (2/9)

APRENDIZAJE AUTOMÁTICO (Específicos):

- Fogel, D.B. “Evolutionary Computation. Toward a New Philosophy of Machine Intelligence”, IEEE Press, 1995. [*]
- Glenberg, Arthur M.; Haines, B. “Learning from data : An introduction to statistical reasoning”, Mahwah : Lawrence Erlbaum, 1996. - XV, [D-SIC/2373D] 2nd ed. [D-SIC/2419].
- Koza, J.R. “Genetic Programming (I and II)” Cambridge, MA, MIT Press, 1992 and 1994.
- Mitchell, M. “An Introduction to Genetic Algorithms” Cambridge, MA, MIT Press, 1996.
- Quinlan, J.R. “C4.5, Programs for Machine Learning” Morgan Kaufmann, San Mateo, 1993. [4-63/793B] [*]
- Sutton, Richard S.; Barto, A.G. “Reinforcement learning : an introduction” Cambridge ; London : Bradford, 1998. [D-ISA/589D, D-SIC/3014D]

Bibliografía (3/9)

APRENDIZAJE DECLARATIVO (ILP, IFLP, ...):

- Bratko, I.; Muggleton, S. “Applications of Inductive Logic Programming” Comm. ACM, Vol. 35, No. 11., 1995. [*]
- **Dzeroski, S.; Lavrac, N. “Relational Data Mining” Springer 2001. [D-SIC/3347] [*]**
- Muggleton, S. (ed) “Inductive Logic Programming” Academic Press, 1992. [*]
- **Muggleton, S.; De Raedt, L. “Inductive Logic Programming: Theory and Methods” J. Logic Prog., 629-679, 1994. [*]**
- **Niehuys-Cheng, S.H.; de Wolf, R. “Foundations of Inductive Logic Programming” Springer, LNAI 1228, 1997. [*]**
- ILP Workshops (1989 – 2004) [*]
- Journal of Logic Programming, Special Issue: ILP, Vol. 40, n°2/3, 1999.
- 1st ICML Workshop on Inductive Programming, 2005.

Bibliografía (4/9)

DATA-MINING Y KDD:

- Adamo, J.M. “Data Mining for Association Rules and Sequential Patterns” Springer 2000. [D-SIC/3243] [*]
- **Berry M.J.A.; Linoff, G.S. “Mastering Data Mining” Wiley 2000. [1-68/1891B] [*]**
- **Berthold, M.; Hand, D.J. (ed) “Intelligent Data Analysis. An Introduction” Springer 1999. (2nd Edition 2002). [D-SIC/3346] [*]**
- De Raedt, L.; Dehaspe, L. “Clausal Discovery” Machine Learning, 26, 99-146, 1997. [*]
- **Dunham, M.H. “Data Mining. Introductory and Advanced Topics” Prentice Hall, 2003. [DSIC/3475D][*]**
- Fayyad, U.M.; Piatetskiy-Shapiro, G.; Smith, P.; Ramasasmy, U. “Advances in Knowledge Discovery and Data Mining”, AAAI Press / MIT Press, 1996. [D-SIC/2667] [*]

Bibliografía (5/9)

DATA-MINING Y KDD:

- **Fayyad, U.M.; Grinstein, G.; Wierse, A. “Information Visualization in Data Mining and Knowledge Discovery” Morgan Kaufmann, Harcourt Intl., 2001. [D-SIC/3313] [*]**
- **Freitas, Alex A. “Data Mining and Knowledge Discovery with Evolutionary Algorithms”, Springer, Natural Computing Series, 2002.**
- **Han, J.; Kamber, M. “Data Mining: Concepts and Techniques” Morgan Kaufmann, 2001. [D-SIC/3274] [*]**
- **Hand, D.J.; Mannila, H. and Smyth, P. “Principles of Data Mining”, The MIT Press, 2000. [D-SIC/3349] [*]**
- **Hernández, J.; Ramírez, M.J.; Ferri, C. “Introducción a la Minería de Datos”, Prentice Hall / Addison Wesley, 2004. [*]**
- **Pyle, D. “Data Preparation for Data Mining” Morgan Kaufmann, Harcourt Intl., 1999. [D-SIC/3283] [*]**

Bibliografía (6/9)

DATA-MINING Y KDD:

- Thuraisingham, B. “Data Mining. Technologies, Techniques, Tools, and Trends”, CRC Press, 1999. [D-SIC/3339] [*]
- Witten, I.H.; Frank, E. “Tools for Data Mining”, Morgan Kaufmann, 1999. [D-SIC/3281] [*]
- Wong, P. C. “Visual Data Mining”, Special Issue of *IEEE Computer Graphics and Applications* , Sep/ Oct 1999, pp. 20- 46. [*]
- **Zhang, C.; Zhang, S. “Association Rule Mining” Lecture Notes in Artificial Intelligence vol. 2307, Tutorial, Springer Verlag 2002. [D-SIC/3546] [*]**

Bibliografía (7/9)

XML, Datos Semiestructurados, Web Mining

- Abiteboul, S.; Buneman, P.; Suciu, D. “Data on the Web: from Relations to Semistructured Data and XML” Morgan Kaufmann, 1999. [*]
- Etzioni, O. “The World- Wide Web. Quagmire or Gold Mine” Communications of the ACM, November 1996, Vol. 39, nº 11, 1996. [*]
- Communications of the ACM “Special Issue on Personalization” August 2000, Vol. 43, No. 8. [*]
- **Kosala, R.; Blockeel, H. “Web Mining Research: A Survey” ACM SIGKDD Explorations, Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining, June 2000, Vol. 2, nº1, pp. 1-15. [D-SIC/] [*]**

Bibliografía (8/9)

XML, Datos Semiestructurados, Web Mining

- **Liu, B.; Chang, C.C. “Special Issue on Web Content Mining” ACM SIGKDD Explorations, Newsletter of the ACM Special Interest Group on Knowledge Discovery and Data Mining, December 2004, Vol. 6, nº1, [*]**
- **Mena, Jesus “Data Mining Your Website”, Digital Press, July 1999. [D-SIC/3271] [*]**
- **Chakrabarti, S. “Mining the Web: Discovering Knowledge from Hypertext Data” Morgan-Kaufmann 2003.[D-SIC/3530] [*]**

Bibliografía (9/9)

PROGRAMACIÓN INDUCTIVA E INGENIERÍA DEL SOFTWARE:

- Bergadano, F.; Gunetti, D. “Inductive Logic Programming. From Machine Learning to Software Engineering” The MIT Press 1995. [4-64/5344B, D-SIC/2730D, 4-64/241F] [*]
- Hernández, J.; Ramírez, M.J. “Predictive Software” J. of Automated Software, Kluwer Academic Publishers, Dordrecht, vol. 8, n°2, 2001. [*]
- **Partridge, D. “The Case for Inductive Programming” IEEE Computer, January 1997, 36-41. [*]**
- Lieberman, Henry “Your Wish is My Command: Programming By Example” Morgan Kaufmann, [D-SIC/3248] [*].
- Zhang, Du.; Tsia, Jeffrey J. P. (eds.) “Machine Learning Applications in Software Engineering”, World Scientific, 2005.