

Exploring the Power of Links in Data Mining

Jiawei Han*

Department of Computer Science
University of Illinois at Urbana-Champaign
hanj@cs.uiuc.edu

Abstract. Algorithms like PageRank and HITS have been developed in late 1990s to explore links among Web pages to discover authoritative pages and hubs. Links have also been popularly used in citation analysis and social network analysis. We show that the power of links can be explored thoroughly in data mining, such as classification, clustering, information integration, and object distinction. Some recent results of our research that explore the crucial information hidden inside links will be introduced, including (1) multi-relational classification, (2) user-guided clustering, (3) link-based clustering, and (4) object distinction analysis. The power of links in other analysis tasks will also be discussed in the talk.

About the Speaker

Jiawei Han is a Professor in the Department of Computer Science at the University of Illinois at Urbana-Champaign. He has been working on research into data mining, data warehousing, database systems, mining spatiotemporal data, multimedia data, stream and RFID data, Web data, social network data, and biological data, and software bug mining, with over 300 conference and journal publications. He has chaired or served on over 100 program committees of international conferences and workshops, including PC co-chair of 2005 (IEEE) International Conference on Data Mining (ICDM), Americas Coordinator of 2006 International Conference on Very Large Data Bases (VLDB). He is also serving as the founding Editor-In-Chief of ACM Transactions on Knowledge Discovery from Data. He is an ACM Fellow and has received 2004 ACM SIGKDD Innovations Award and 2005 IEEE Computer Society Technical Achievement Award. His book “Data Mining: Concepts and Techniques” (2nd ed., Morgan Kaufmann, 2006) has been popularly used as a textbook worldwide.

* The work was supported in part by the U.S. National Science Foundation NSF ITR/CCR-0325603, IIS-05-13678/06-42771, and NSF BDI-05-15813. Any opinions, findings, and conclusions or recommendations expressed here are those of the authors and do not necessarily reflect the views of the funding agencies.