

The Distributed Data-Mining Workshop

Hillol Kargupta and Philip Chan

The Workshop on Distributed Data Mining (DDM), held at the Fourth International Conference on Knowledge Discovery and Data Mining, brought interested researchers and practitioners together and created an environment for crystallizing the fast-growing field of DDM. The workshop focused on the state-of-the-art DDM algorithms, systems, and application-related issues. Approximately 40 participants attended the workshop. The workshop had 13 presentations, including 3 invited talks.

The workshop began with a visionary invited talk on the possible role of mobile agents in DDM by George Cybenko of Dartmouth University. His talk described some of the existing mobile agent-based systems and pointed out the strengths of distributed data analysis using mobile agents. This talk was followed by the general session. Yike Guo and Janias Sutiwaraphun (both of Imperial College, United Kingdom) proposed distributed knowledge probing that uses meta learning for hierarchical supervised learning. Hillol Kargupta, Erik Johnson, Byung-Hoon Park, Daryl Hershberger (all of Washington State University), Eleonara Riva Sanseverino, and Luisa Di Silvestre (both of University of Palermo, Italy) proposed a new approach called collective data mining from distributed data with vertically partitioned feature space. They pointed out that naive local analysis of data can be misleading when features are distributed, and distributed representation of functions using a set of orthogonal basis functions can effectively be used for accurate DDM. Vincent Ng and David Cheung (both

of Hong Kong Polytechnic University, China) presented new techniques for learning distributed associations under constraints. Philip Chan (Brigham Young University) and Salvatore Stolfo (Columbia University) presented the effects of class distribution in the training set on performance and proposed different techniques for measuring performance. Lawrence Hall, Nitesh Chawla, and Kevin W. Bowyer (all of University of South Florida) presented the results of their approach for combining decision trees in parallel. Andreas Prodromidis (Columbia University) and Stolfo suggested different techniques for

The Workshop on Distributed Data Mining (DDM), held at the Fourth International Conference on Knowledge Discovery and Data Mining, brought interested researchers and practitioners together and created an environment for crystallizing the fast-growing field of DDM. The workshop focused on the state-of-the-art DDM algorithms, systems, and application-related issues.

pruning meta-classifiers to make the resultant system less complex. Abdula Tansel (City University of New York) and Necip Ayan (Bilkent University) discussed techniques for discovering association rules in temporal data.

The afternoon session was initiated by a stimulating invited talk by Foster Provost (Bell Atlantic Science and Technology) on the scalability issue of DDM. The preliminary design of PAPHYRUS, a DDM system under development, was presented by Robert Grossman, Stuart Bailey, Simon Kasif, Don

Mon, Ashok Ramu, and Balinder Mathi (all of University of Illinois at Chicago). Ramesh Subramonian (Intel Corp.) and Srinivasan Parthasarathy (University of Rochester) discussed architectural issues in DDM systems. They proposed a novel architecture that maps a DDM application onto a cluster of symmetric multiprocessor environments and provides distributed programming primitives that reduce interprocessor communication. Victor Lesser (University of Massachusetts at Amherst) gave an invited talk on distributed interpretation and its possible implication in DDM. He shared his wealth of experience in building resource-bounded distributed AI-based systems for the last 20 years. The paper sessions ended with two working paper presentations by Billy Wallace and Juan Botia, Mercedes Garjijo, Juan Velasco (all of Technical University of Madrid), and Antonio Skarmeta (University of Murcia).

The concluding session was the panel discussion. Stolfo, George Cybenko (Dartmouth College), and Provost discussed different critical issues for DDM research and applications. Provost pointed out the significant practical reasons behind DDM. Stolfo stressed research issues motivated by the real world and encouraged participants to explore data available from the federal government. Cybenko identified several research directions for DDM using mobile agents.

Organizers sincerely hope that the workshop created a stimulating environment for further growth of the field of DDM. A book entitled "Advances in Distributed Data Mining," expected to be published in early 2000 by the AAAI Press, is being planned to disseminate the latest developments in DDM. Further details about the developments can be found at www.eecs.wsu.edu/~hillol/ddm.html.

Hillol Kargupta is currently an assistant professor of computer science at Washington State University (WSU). His current research interest is in the area of distributed data mining and gene expression-based evolutionary algorithms.

Philip Chan is currently an assistant professor of computer science at the Florida Institute of Technology.