IEEE ISI 2008 Invited Talk (II)

Probabilistic Frameworks for Privacy-Aware Data Mining

Dr. Joydeep Ghosh

Schlumberger Centennial Chair in Engineering
University of Texas at Austin

Abstract

Often several cooperating parties would like to have a global view of their joint data for various data mining objectives, but cannot reveal the contents of individual records due to privacy, ownership or competitive considerations. In this talk, we present a probabilistic framework for resolving such seemingly contradictory goals. Rather than sharing parts of the original or perturbed data, the framework shares the parameters of suitable probabilistic models built at each local data site. We mathematically show that the best representative of all the data is a certain "mean" model, and empirically show that this model can be approximated quite well by generating artificial samples from the underlying distributions using Markov Chain Monte Carlo techniques, and then fitting a combined global model with a chosen parametric form to these samples. We also propose a new measure that quantifies privacy in such situations based on information theoretic concepts, and show that decreasing privacy leads to a higher quality of the combined model and vice versa. The method can also be applied to situations where different local datasets may not have identical features by using certain maximum likelihood and maximum entropy principles. We provide empirical results on different data types with continuous vector, categorical and directional attributes to highlight the generality of our framework. The results show that high quality distributed clustering or classification can be achieved with little privacy loss and low communication cost.

Biography: Professor Joydeep Ghosh is currently the Schlumberger Centennial Chair Professor of Electrical and Computer Engineering at the University of Texas, Austin. He joined the UT-Austin faculty in 1988 after being educated at IIT Kanpur, (B. Tech '83) and The University of Southern California (Ph.D?8). He is the founder-director of IDEAL (Intelligent Data Exploration and Analysis Lab) and a Fellow of the IEEE. His research interests lie primarily in intelligent data analysis, data mining and web mining, adaptive multi-learner systems, and their applications to a wide variety of complex engineering and AI problems.

Dr. Ghosh has published more than 200 refereed papers and 30 book chapters, and co-edited 18 books. His research has been supported by the NSF, Yahoo!, Google, ONR, ARO, AFOSR, Intel, IBM, Motorola, TRW, Schlumberger and Dell, among others. He received the 2005 Best Research Paper Award from UT Co-op Society and the 1992 Darlington Award given by the IEEE Circuits and Systems Society for the Best Paper in the areas of CAS/CAD, besides nine other "best paper" awards over the years. He was the Conference Co-Chair of Computational Intelligence and Data Mining (CIDM?7), Program Co-Chair for The SIAM Int'l Conf. on Data Mining (SDM'06), and Conf. Co-Chair for Artificial Neural Networks in Engineering (ANNIE)'93 to '96 and '99 to '03. He is the founding chair of the Data Mining Tech. Committee of the IEEE CI Society. He also serves on the program committee of several top conferences on data mining, neural networks, pattern recognition, and web analytics every year. Dr. Ghosh has been a plenary/keynote speaker on several occasions such as ANNIE?6, MCS 2002 and ANNIE'97 and, and has widely lectured on intelligent analysis of large-scale data. He has co-organized workshops on high dimensional clustering

(ICDM 2003; SDM 2005), Web Analytics (with SIAM Int'l Conf. on Data Mining, SDM2002), Web Mining (with SDM 2001), and on Parallel and Distributed Knowledge Discovery (with KDD-2000).

Dr. Ghosh has served as a consultant or advisor to a variety of companies, from successful startups such as Neonyoyo and Knowledge Discovery One, to large corporations such as IBM, Motorola and Vinson & Elkins. At UT, Dr. Ghosh teaches graduate courses on data mining, artificial neural networks, and web analytics. He was voted the Best Professor by the Software Engineering Executive Education Class of 2004.