



DEPARTMENT OF COMPUTER SCIENCE

PhD Degree Oral Presentation

PhD Candidate:	Mr Xiaofeng ZHANG
Supervisor:	Dr Kwok Wai CHEUNG
External Examiner:	Dr Korris Fu-lai CHUNG Dr Yiu Ming CHEUNG (Proxy for Dr KLUSCH)
Time:	26 October 2007 (Friday) 10:30 am – 12:30 pm (35 mins presentation and 15 mins Q & A)
Venue:	R905, Conference Room, Sir Run Run Shaw Building, HSH Campus

“A Model-based Approach for Distributed Data Mining”

Abstract

Recently, there exist a number of cases where data are distributed and cannot be shared due to local constraints, such as privacy concerns or bandwidth limits, which greatly challenge traditional data mining technologies. In this thesis, we focus on studying how the model-based approach can be applied to distributed data mining with privacy concern. A generic distributed model-based data mining approach called learning from abstraction is proposed which computes local data abstraction using hierarchical clustering algorithms. Gaussian mixture model and generative topographic mapping are the global models for two applications --- distributed data clustering and distributed manifold discovery respectively. An EM-like algorithm is derived for learning both global models solely based on the model parameters of the local abstractions. Then we studied the problem of how to set the right trade-off between the abstraction levels of the local data sources and the global model accuracy. The problem is first formulated as a game and the Nash equilibrium of the game gives the best abstraction strategy. To enforce the privacy requirement, the game is further modified so that the trade-off computation is achieved via active exploration of local data's details. Future research directions include the extension of the proposed approach to peer-to-peer computing environments which can be untrusted and dynamic.

***** ALL INTERESTED ARE WELCOME *****