Department of Computer Science, The University of Hong Kong

Student Research Seminar

Mining Uncertain Data with Probabilistic Guarantees

Mr. Sun Liwen

Date & Time:
December 4, 2009
Friday
3:30 pm

Venue:
Room 328
Chow Yei Ching Building
The University of Hong Kong

Abstract:

Data uncertainty is inherent in important and emerging applications, such as sensor data management and biological databases. To manage data uncertainty, probabilistic databases have been recently developed, where probability and statistical information is stored together with the data. Given the richness of the information presented in these data, useful knowledge may be discovered from them. In this paper, we investigate how the notions of frequent patterns and association rules, which have been defined for exact databases, can be extended to support the tuple uncertainty model. Due to the exponential number of possible-world instances presented in a probabilistic database, mining these patterns and rules is technically challenging. We propose two algorithms, based on dynamic programming and Fast Fourier transform techniques, in order to generate frequent patterns efficiently. We further present a novel algorithm to convert these patterns to association rules. Extensive experiments, using real and synthetic datasets, were conducted to validate the performance of our methods.

About the Speaker:

Sun Liwen is a full-time MPhil student in the Department of Computer Science, The University of Hong Kong, under the supervision of Prof. David W.L. Cheung and Dr Reynold C.K. Cheng. His research interest is data mining.

All are welcome!
For enquiries, please call 2859-2180 or email enquiry@cs.hku.hk
Department of Computer Science
The University of Hong Kong