Department of Computer Science, The University of Hong Kong

Student Research Seminar

Clustering Uncertain Data using Voronoi Diagrams and R-Tree Index

Mr. Lee King For

Date & Time:
October 14, 2009
Wednesday
7:00 pm

Venue:
Room 308
Chow Yei Ching Building
The University of Hong Kong

Abstract:

We study the problem of clustering uncertain objects whose locations are described by probability density functions (pdf). We show that the UK-means algorithm, which generalises the k-means algorithm to handle uncertain objects, is very inefficient. The inefficiency comes from the fact that UK-means computes expected distances (ED) between objects and cluster representatives. For arbitrary pdf's, expected distances are computed by numerical integrations, which are costly operations. We propose pruning techniques that are based on Voronoi diagrams to reduce the number of expected distance calculation. These techniques are analytically proven to be more effective than the basic unding-box-based technique previously known in the literature. We then introduce an R-tree index to organise the uncertain objects so as to reduce pruning overheads. We conduct experiments to evaluate the effectiveness of our novel techniques. We show that our techniques are additive and, when used in combination, significantly outperform previously known methods.

About the Speaker:

Lee King For is a MPhil student in the Department of Computer Science, The University of Hong Kong under the supervision of Dr. Benjamin Kao. His research interest is uncertainty and data clustering .

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