Wong Lim Soon

From: Loo Line Fong [loolf@comp.nus.edu.sg]
Sent: Monday, March 31, 2008 8:59 AM
To: acad-cs-l@comp.nus.edu.sg; rs-l@comp.nus.edu.sg
Cc: wongls@comp.nus.edu.sg; Wynne Hsu; paelaips@nus.edu.sg; leongty@comp.nus.edu.sg
Subject: Postgraduate Seminar by Ms Lin Li

SCHOOL OF COMPUTING,NUS

POSTGRADUATE SEMINAR BY

MISS LIN LI

Efficient Mining of Haplotype Patterns for Disease Prediction

Meeting Room 1 (COM1 03-43)

9 April 2008, 9.00am

Abstract:

It was quoted by J. Han (2006) that we are at the stage of being data rich but information poor; the profusion in data collection does not correspond with an exponential development in efficient techniques to extract valuable and useful knowledge from data. The filling of such knowledge gap is a challenge faced by all data miners.

This work focuses on knowledge extraction from domain specific data known as haplotypes. A major issue in pattern extraction from haplotypes is the ability to identify valuable and useful information for disease pattern prediction which can be applied in prognosis and carrier detection.

Firstly, we propose a new method known as LinkageTracker for disease gene location inference (or linkage disequilibrium mapping) from haplotypes. This method was compared with some leading methods in linkage disequilibrium mapping such as Haplotype Pattern Mining (HPM; Toivonen et al, 2000), HapMiner (Li & Jiang, 2005), Blade (Lu et al 2003), and GeneRecon (Mailund et al, 2006). LinkageTracker provides good predictive accuracies while taking up reasonably short processing times. Furthermore, LinkageTracker does not require any population ancestry information about the disease and the genealogy of the haplotypes. It is a useful tool for linkage disequilibrium mapping when the users do not have much information about their datasets.

Secondly, we introduce a novel algorithm called ECTracker for extracting useful haplotype patterns for genetic analysis and carrier detection. Experimental studies show that ECTracker is capable of deriving useful patterns when the dataset is very small. In classification, ECTracker is capable of producing good predictive accuracies that are comparable to some leading machine learning methods. Using biological datasets from wet experiments, ECTracker could efficiently extract patterns that allow for predictive disease classification. Furthermore, it is able to classify samples as unknown if they are almost indistinguishable from the defined classes. This work, in most cases, outperforms the existing methods in classification accuracies for datasets like haplotype patterns for disease class prediction.

* * * * * * * * *

LOO Line Fong (Ms) :: Management Assistant Officer :: Graduate Studies :: Dean's Office :: School of Computing, National University of Singapore :: Computing 1, 03-68 Law Link, Singapore 117590 :: 65-6516 4707 (DID) :: 65- 6779 4580(Fax) :: loolf@comp.nus.edu.sg (Email) :: www.comp.nus.edu.sg (Website) :: Company Registration No: 200604346E

4/3/2008