An Empirical Study of Dynamic Incomplete-case Nearest Neighbor Imputation in Software Quality Data

Mr. Jianglin Huang
PhD Candidate
Department of Systems Engineering and Engineering Management
City University of Hong Kong

<table>
<thead>
<tr>
<th>Date</th>
<th>17 Aug 2015 (Monday)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time</td>
<td>10:30am</td>
</tr>
<tr>
<td>Venue</td>
<td>P6921, 6/F, AC1</td>
</tr>
</tbody>
</table>

Abstract

Software quality prediction is an important yet difficult problem in software project development and management. Historical datasets can be used to build models for software quality prediction. However, the missing data significantly affects the prediction ability of models in knowledge discovery. Instead of ignoring missing observations, we investigate and improve incomplete-case k-nearest neighbor based imputation. K-nearest neighbor imputation is widely applied but has rarely been improved to have the most appropriate parameter settings for each imputation. This work conducts imputation on four well-known software quality datasets to discover the impact of the new imputation method we proposed. We compare it with mean imputation and other commonly used versions of k-nearest neighbor imputation. The empirical results show that the proposed dynamic incomplete-case nearest neighbor imputation performs better when the missingness is completely at random or non-ignorable, regardless of the percentage of missing values.

About the Speaker

Jianglin Huang received his B.S. degree in 2012 from University of Electronic
Science and Technology of China. He is currently a PhD candidate at the Department of Systems Engineering & Engineering Management, City University of Hong Kong. His research interests include empirical software quality/effort estimation.

Enquiry: 3442 8408

All are Welcome!

SEEM Seminar 2015-2016/006