香港城市大學
City University of Hong Kong

Concept drift in Empirical Software Engineering

Presenter:

Md Alamgir Kabir

Supervisor:

Dr. Jacky Keung

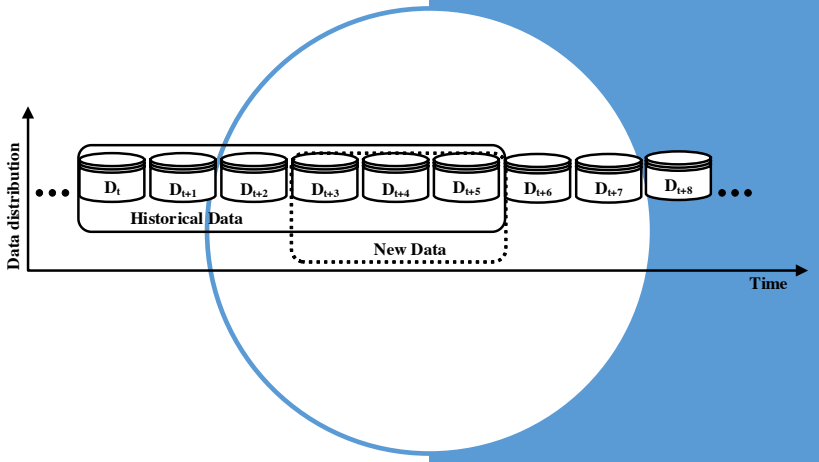
Department of Computer Science, CityU

Outline

1. Investigating Concept Drift on Software Defect Datasets
2. Concept Drift and Class Imbalance
3. Work Plan

Investigating Concept Drift on Software Defect Prediction

- Detection Method:
  - DDM (Drift Detection Method)
- Statistical Test:
  - Chi-square test with Yate's continuity Correction
- Datasets:
  - jm1 and prop (comparatively large)
- Drift Types:
  - Sudden drift (in *jm1* datasets)
  - Gradual drift (in *prop* datasets)



Class Imbalance and Concept Drift

- **Combined problem** where both class imbalance learning and concept drift **coexist**
- **How** to best **overcome** concept drift in online learning with class imbalance
- The problem becomes particularly challenging when **they occur simultaneously**
- It is also unknown **whether and how applying class imbalance techniques** (e.g., resampling methods) affects concept drift detection and online prediction.

Wang, S., Minku, L. L., & Yao, X. (2018). A systematic study of online class imbalance learning with concept drift. *IEEE Transactions on Neural Networks and Learning Systems*.

Work Plan

I. Deal with small datasets
    I. Fisher Exact test (recommended by the literature)
II. Window size is fixed by the user
    I. Adaptive Window
III. Concept drift with Class Imbalance Learning
    I. Resampling methods

CityU
香港城市大學
City University of Hong Kong

Thanks for listening!