SYSTEMS ENGINEERING RESEARCH CENTER

Put your logo here on the master slide

Models for Efficient Testing (MET)

Sponsor: OUSD(R&E) | CCDC

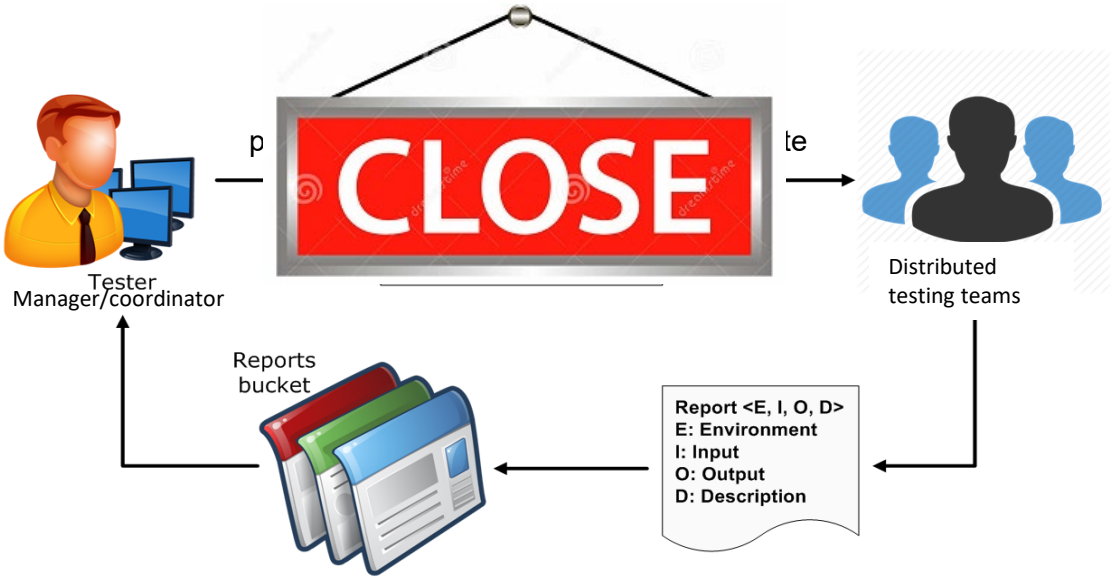
By
Dr. Ye Yang

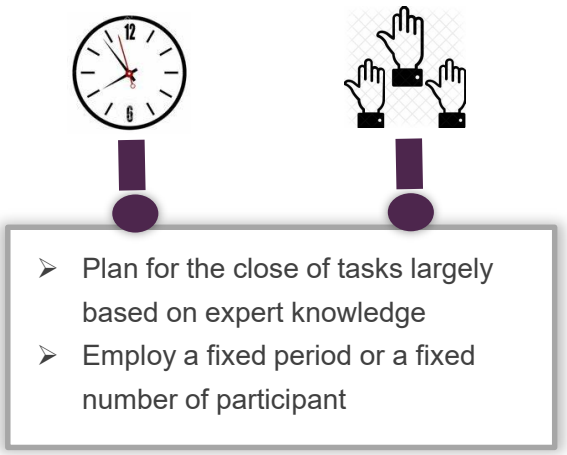
11th Annual SERC Sponsor Research Review
November 19, 2019
FHI 360 CONFERENCE CENTER
1825 Connecticut Avenue NW, 8th Floor
Washington, DC 20009

www.sercuarc.org

Motivation and Objectives

- Trade-off such as “how much is enough” is challenging, esp. in exploratory testing
  - Insufficient testing can lead to unsatisfying product quality, while
  - excessive testing can result in potential schedule delays and low cost-effectiveness.
- Objective
  - To enable actionable, value-driven decision making on resource allocation and utilization faced by testing managers.



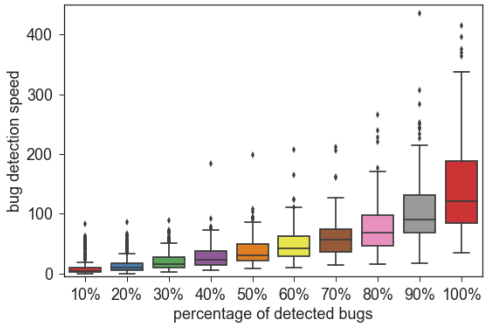


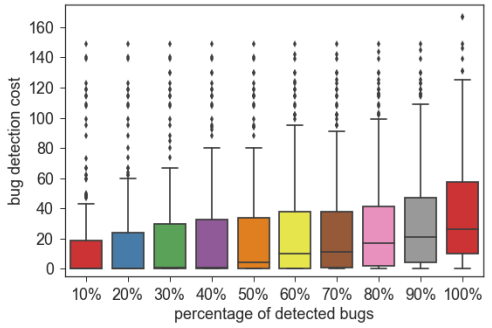
Project Overview

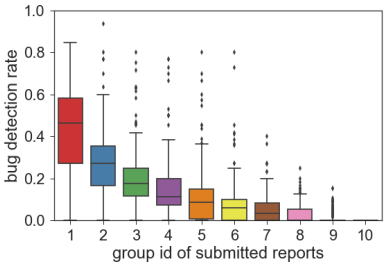
- MET project
  - Investigates, characterizes distributing testing processes
  - Develops a set of Machine Learning-based approaches to support efficient testing management across distributed testing teams
- MET consists of
  - **A testing measurement model**
    - For characterizing the representative contextual factors of a testing process
  - **A in-process team formation model**
    - Matching, learning, ranking, and dynamically tuning the configuration of distributed testing teams to maximize testing adequacy
    - Leveraging natural language processing (NLP) and learning-to-rank algorithms
  - **A early completion detection model**
    - Monitoring, aggregating testing reports, predicting total defects, and automate testing completion detection

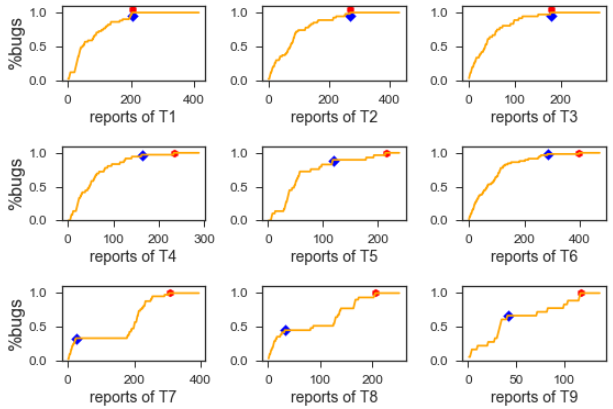
Pilot Study – Analysis Results

- A pilot study: 636 real-world mobile application testing projects.
- Observation: Excessive engagement leading to waste due to duplicated effort.
  - Average wasteful spending over 636 projects: **32%**





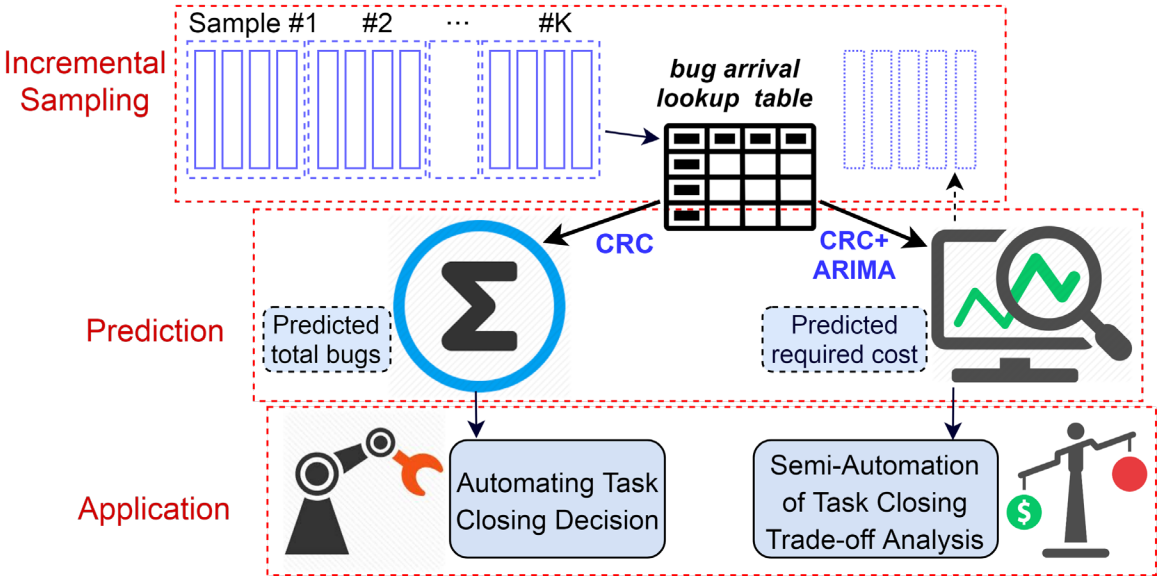


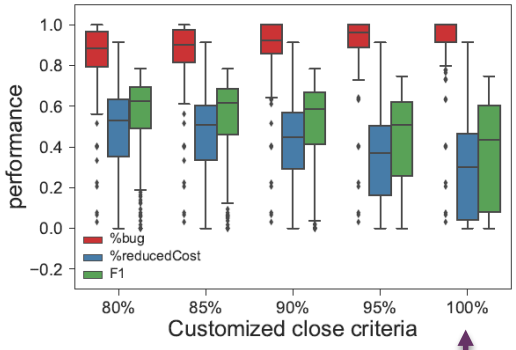


Bug Arrival Patterns:

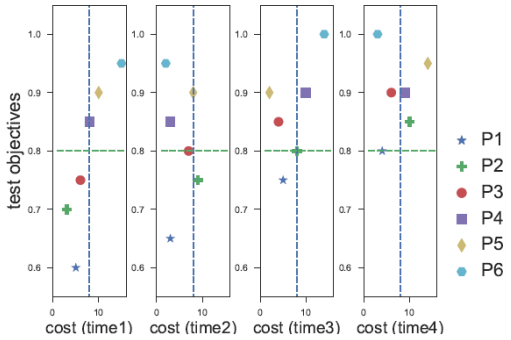
1) Large variation in bug detection speed and cost;
2) Decreasing bug detection rates over time;
3) Plateau effect of bug arrival curves.

iSENSE: Completion-Aware Crowdtesting Management





30% Cost reduction

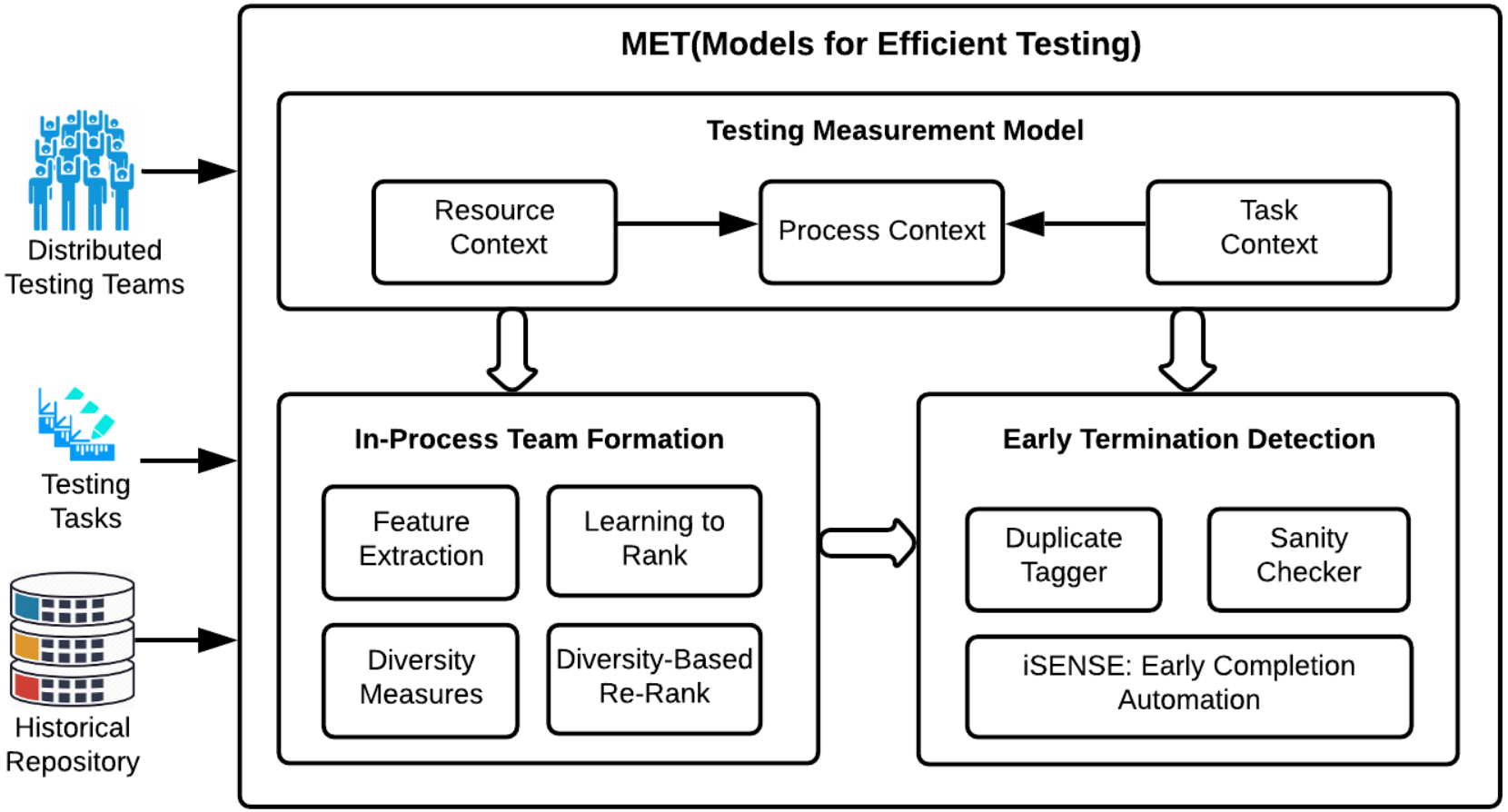


Evaluation: Baidu Crowd Test dataset;

1. Automated close management: 30% cost reduction.
2. Trade-off analysis support.

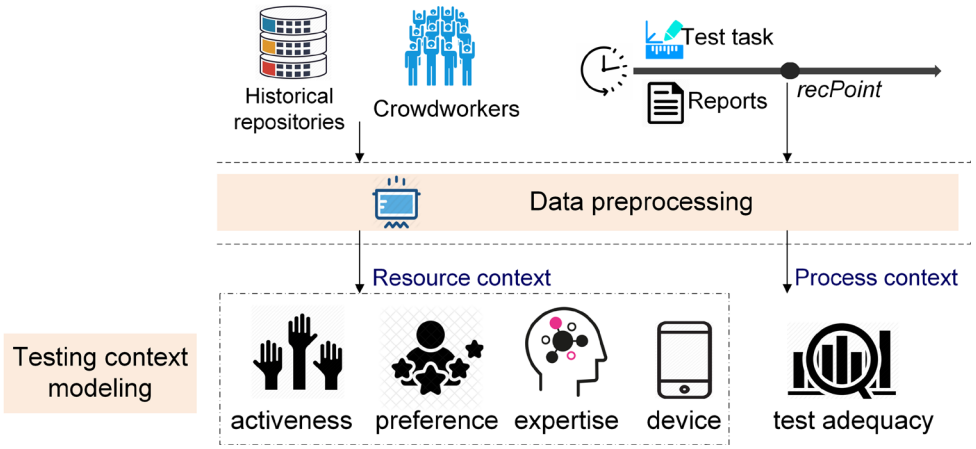
(**ACM SigSoft Distinguished Paper Award**) J. Wang, Y. Yang, R. Krishna, T. Menzies, Q. Wang. "iSENSE: Completion-aware Crowdtesting Management". ICSE 2019.

Research Vision

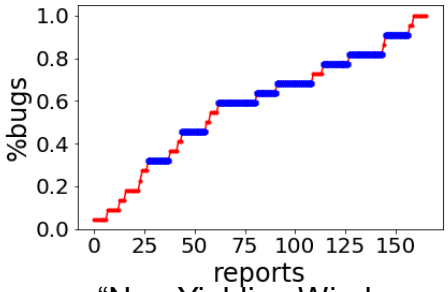


Testing Measurement Model (TMM)

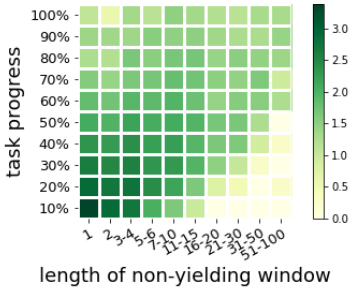
- Two sub-models
  - **Process context sub-model:** in-process progress-oriented information
    - **Testing requirement matrix**: represent the task's requirements in the vector space of descriptive terms from testing requirements.
    - **Test adequacy:** To measure the testing progress regarding to what degree each descriptive term has been tested.
  - **Resource context sub-model:** characteristics of distributed testing teams
    - **Activeness**: a set of metrics to characterize a team's activeness over certain period
    - **Preference:** distribution of term intensity from a testing team's historical reports
    - **Expertise**: distribution of term intensity from a testing team's unique bug reports
    - **Devices**: environmental factors such as testing devices, equipment, etc.



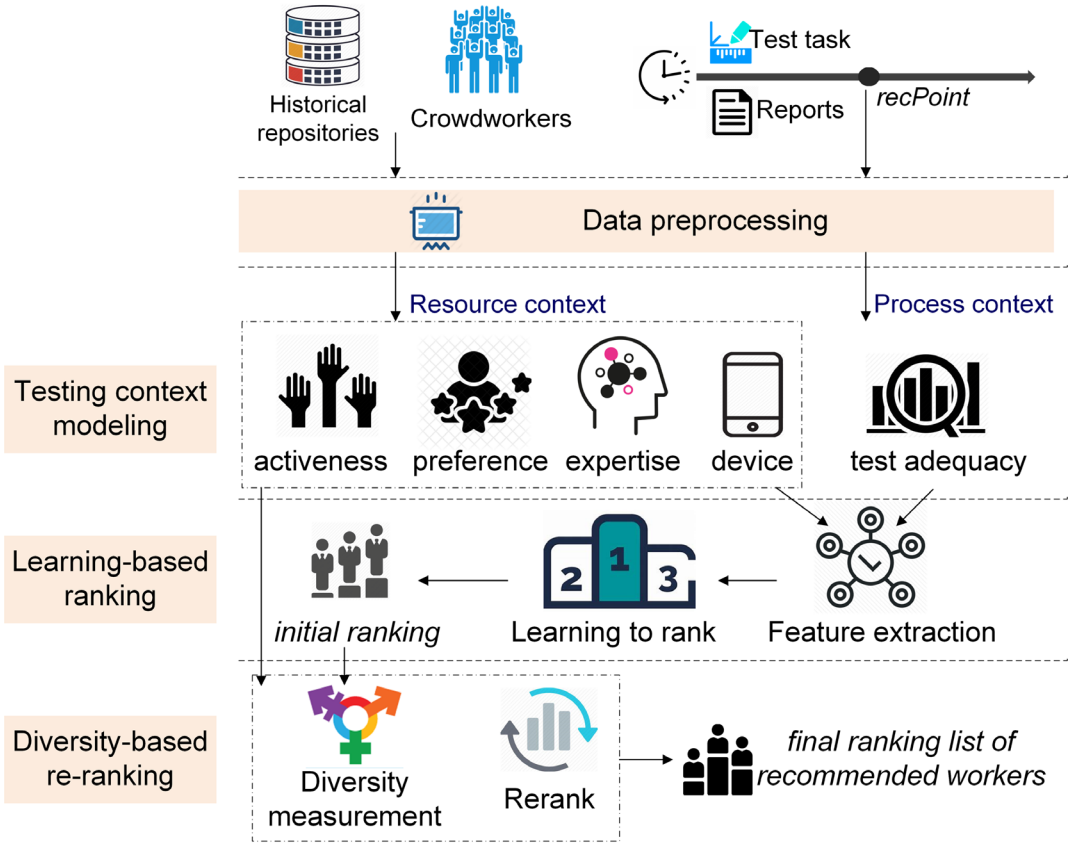
iRec: In-Process Team Recommendation



"Non-Yielding Windows (NYW)"



NYWs distribution across testing cycles

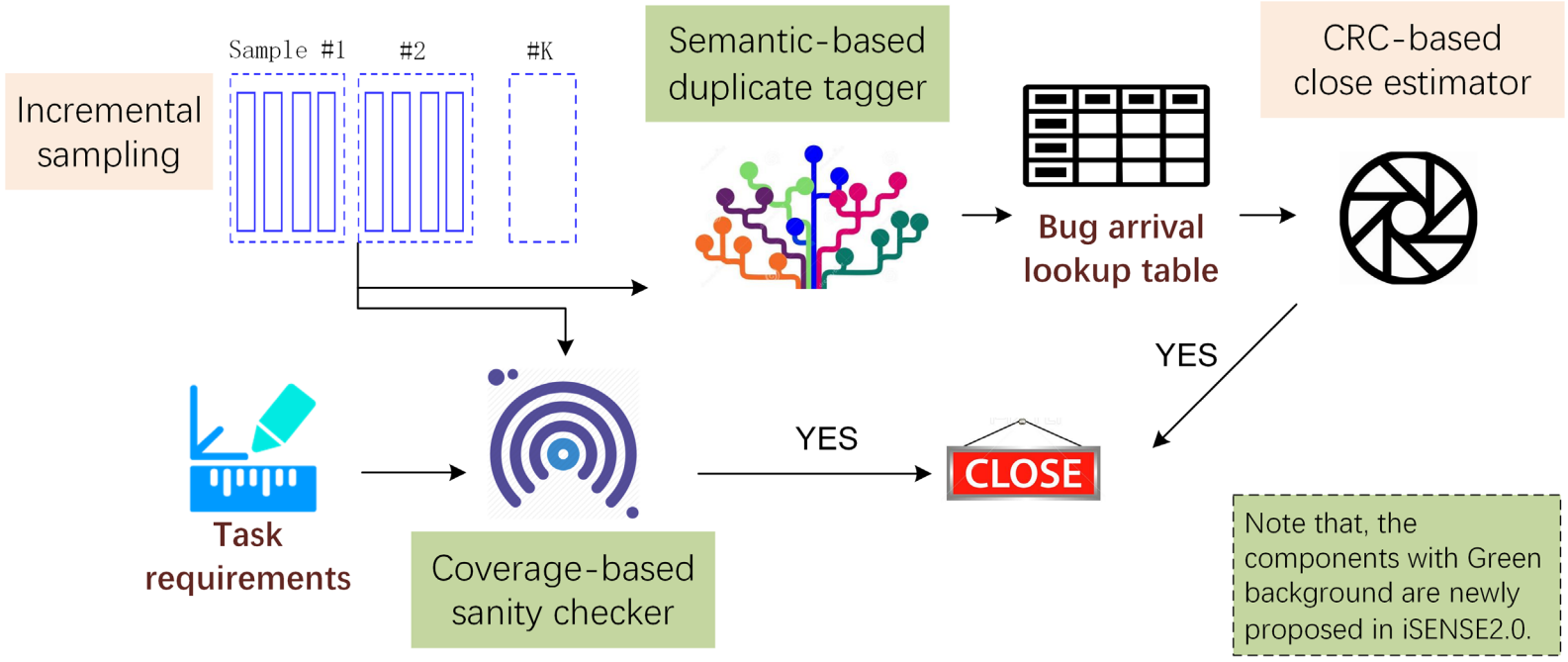


Evaluation: Baidu Crowd Test dataset; Reduction of the NYWs by 50% - 58% ; Reduction of cost by about 10% on median

J. Wang, Y. Yang, S. Wang, Q. Wang. Submitted to ICSE 2020, under review.

Work in progress: iSENSE 2.0

- Extending iSENSE with new components
  - —Automated Duplicate Tagger
    - Analyzing the duplicate status of received crowd reports leveraging on semantic analysis
  - —Coverage-based Sanity Checker
    - Reinforce the stability and performance of close prediction.



Next Steps

- Elaborate the testing measurement model (TMM)
  - To address specific needs for characterizing DoD task/process/resource context
  - Refine core underlying metric: **Testing Process Adequacy**
- Integrate TMM with iRec and iSENSE 2.0
- Empirical validation of existing models on cross-platform datasets
  - In-process team recommendation
  - Early completion detection
- Seeking collaboration in further evaluation in DoD testing projects.

Thank you!

Contact:
ye.yang@stevens.edu