# Positive and Unlabeled Learning for Anomaly Detection

ZHANG JIAQI
Supervisor: Assoc Prof. Yuan Junsong
Co-Supervisor: Prof. Tan Yap-Peng

February 11, 2018

**Abstract**

Anomaly detection is of great interest to big data applications, and both supervised and unsupervised learning methods have been applied for anomaly detection. However, it still remains a challenging problem for machine learning-based methods because of the difficulty of obtaining anomaly data for training, which is usually rare, diversely distributed, and difficult to collect. Thus it is difficult for supervised learning to acquire training data while for unsupervised learning, the performance may not be satisfactory due to the lack of labeled training data.

To address the challenge, we propose a hybrid solution by applying Positive and Unlabeled (PU) Learning for anomaly detection problem. As a semi-supervised method, only normal (positive) data and unlabeled data (could be positive or negative) are required by the proposed method for anomaly detection. We start by using a linear model to extract the most reliable negative instances followed by an iterative self-learning process to add reliable instances with different speeds. Furthermore, we adopt boosting in the self-learning process to advantageously exploit the instability characteristic of PU learning when feedback is available. The classifiers in self-learning process are weighted combined based on the estimated error rate to build the final classifier. Our proposed methods are verified on several benchmark datasets and outperform existing methods under different conditions.

# 1 Publications

1. **Jiaqi Zhang**, Zhenzhen Wang, Junsong Wang, Tap-Peng Tan "Positive and Unlabeled Learning for Anomaly Detection with Multi-features", acceptted by ACM MM Conference, 2017.

2. **Jiaqi Zhang**, Zhenzhen Wang, Jingjing Meng, Tap-Peng Tan, Junsong Wang "Boosting Positive and Unlabeled Learning for Anomaly Detection with Multi-features", submitted to IEEE Transactions on Multimedia.