Dr. Lihua Wang[1*], Prof. Ryo Nojima[1,2], Atsushi Waseda[3]

1. National Institute of Information and Communications Technology (NICT), Koganei, Japan
2. Ritsumeikan University, Ibaraki, Japan
3. Tokyo University of Information Sciences, Chiba, Japan

Differentially Private (Random) Decision Tree with k-Anonymity and no Added Noise

Abstract: Recently, the progress in data analysis technologies and computing power has led to extensive utilization of various types of data held by organizations. However, in most cases, such data contain personal information. Hence, privacy preservation has become a primary concern. Typical privacy preservation approaches employed during model training include anonymity frameworks as well as statistics and encryption methods. Differential privacy is a well-known technique used for preserving data privacy from trained models, and it is typically achieved by introducing Laplace or Gaussian noise into training data. However, this technique can distort the original data, particularly in sensitive domains such as healthcare.

Decision trees, which are popular models used in traditional machine learning, are particularly attractive because of their efficiency and interpretability. However, they are vulnerable to information leakage via attacks such as homogeneity and uniqueness attacks, necessitating robust privacy protection methods for underlying data. To address data security concerns and privacy preservation challenges, this study explores various methods to incorporate differential privacy into decision trees via the frameworks of k-anonymity and ℓ-diversity. We first present a k-anonymization technique customized for decision trees and its integration with ℓ-diversity to mitigate potential attacks. In addition, we demonstrate the way differential privacy can be realized in randomized decision trees without the need for noise addition. Our approach enhances the feasibility of using sensitive financial and medical data while preserving privacy.

Keywords: privacy preservation; (random) decision tree; k-anonymity; ℓ-diversity; differential privacy

12

Biography: Dr. Lihua Wang is a senior researcher at Cybersecurity Research Institute, NICT, Japan, focusing on cryptography and its applications in privacy-preserving machine learning. Her work has resulted in several patents and numerous publications in journals and international conferences. Notably, her co-authored paper published in IEEE Transactions on Information Forensics and Security won the 2023 IEEE SPS Best Paper Award. In addition, as a principal investigator, she has led three R&D projects funded by JSPS KAKENHI. Today's discussion was partially supported by JST CREST Grant Number JPMJCR21M1, Japan.

ISBN: 978-1-0685851-7-3