



Contribution ID: 51

Type: **not specified**

Training Gradient Boosted Decision Trees on Tabular Data Containing Label Noise for Classification Tasks

Thursday, 16 May 2024 10:00 (20 minutes)

Label noise, the mislabeling of instances in a dataset, is harmful to classifier performance, increases model complexity, and impairs adequate feature selection. It is frequent in large scale datasets and naturally occurs when human experts are involved. While extensive research has focused on mitigating label noise in image and text datasets through deep neural networks, there exists a notable gap in addressing these issues within Gradient Boosted Decision Trees (GBDTs) and tabular datasets.

This study aims to bridge this gap by adapting two noise detection methods, originally developed for deep learning, to enhance the robustness of GBDTs. Through this adaptation, we aim to augment the resilience of GBDTs against label noise, thereby improving their performance and reliability in real-world applications. The algorithms' effectiveness is rigorously tested against several benchmark datasets that have been intentionally polluted with various amounts and types of noise.

One of the devised algorithms achieves with state-of-the-art noise detection performance on the Adult dataset, showcasing its potential to effectively identify and mitigate label noise.

The investigation extends to analyzing the overarching effects of label noise on the performance of GBDTs, the challenges of different types of noise, and the effectiveness of various noise treatment strategies.

The insights derived from this study not only enhance our understanding of the detrimental effects of label noise on the accuracy and reliability of GBDTs but also inform practical guidelines for handling such noise. Through rigorous analysis, the study proposes a direction for future research in enhancing GBDTs' resilience to label noise and ensuring their continued success in tabular data classification tasks.

Type of presentation

Contributed Talk

Primary author: EISENBÜRGER, Anita (Debeka)

Co-authors: Dr OTTEN, Daniel (Debeka); Prof. HOPFGARTNER, Frank (Universität Koblenz)

Presenter: EISENBÜRGER, Anita (Debeka)

Session Classification: Contributed session

Track Classification: Spring Meeting