Fine-Tuned the Pre-Trained Chinese RoBERTa Model for Predicting Penalty Types in Court Decisions of Public Insult Cases

Decheng Hsieh

Department of Computer Science and Information Engineering, College of Science and Technology, National Chi Nan University, Nantou County 545301, Taiwan

Lieuhen Chen

Department of Computer Science and Information Engineering, College of Science and Technology, National Chi Nan University, Nantou County 545301, Taiwan

Taiping Sun

Department of Electrical Engineering, College of Science and Technology, National Chi Nan University, Nantou County 545301, Taiwan

Abstract

In this study, we present a systematic investigation into predicting penalty types—fine or detention—in public insult cases by fine-tuning a pre-trained Chinese RoBERTa-wwm model. Addressing critical gaps in prior research, which often focused on high-level judicial decisions and suffered from inadvertent inclusion of outcome-related cues, we propose a novel approach that systematically evaluates the contribution of different judicial text components. By collecting 1,115 judgments from Taiwan's Judicial Yuan, we extract and combine crime facts, objective facts influencing sentencing, and judges' names to form four distinct datasets. To address the significant class imbalance inherent in the data, cost-sensitive learning techniques are incorporated, and hyperparameter optimization is conducted via the tree-structured Parzen estimator. Experimental results indicate that the best-performing model, trained on a dataset combining crime facts, objective facts influencing sentencing, and judges' names, achieved an accuracy of 79.16%, a macro F1 score of 0.6494, and an ROC AUC of 0.7512. These findings underscore the importance of integrating general case information with individual judicial tendencies for robust penalty prediction, thereby contributing to the development of an interpretable and cost-effective AI-driven legal decision support system. Our work lays a solid foundation for further enhancements in domain-specific pre-training and classifier design.

Keywords

Penalty Prediction, Legal Judgment Prediction, Public Insult Cases, Chinese RoBERTa-wwm.