Effective Learning Models on Pattern Mining Applications

Thesis submitted in partial fulfillment
for the award of degree

Doctor of Philosophy

by

Shivang Agarwal

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
Indian Institute of Technology
(Banaras Hindu University)
Varanasi

Roll No: 16071503 2021

Chapter 8

Conclusions and Future Work

In this section, we provide a summary of the work presented in the earlier chapters of this dissertation. We worked with the motivation of using the inherent properties of data while generating the hypotheses to be used for classification. Also, we emphasized on the cross-dataset evaluation of models to be used on pattern mining applications. These requirements were satisfied by the conducted studies.

In Chapter 3, an adaptive midway between eager and lazy learning is proposed. EaZy learning proves to be robust under extreme test environments. We considered fingerprint spoof detection as an open-set problem which requires the learning model to be adaptive. Spoof detectors need to identify presentation attacks made by using unknown presentation attack instruments or biometric sensors. We claim that a spoof detector must adapt to the features of the bonafide presentation (live fingerprint) and attack presentation (spoof) to perform reasonably well under cross-sensor and cross-dataset environments. We raised two research questions and demonstrated the experimental results answering these questions. We performed our experiments on standard high dimensional datasets and explored the working mechanism of the proposed model, along with the traditional ensemble learning models under various experimental settings.

In Chapter 4, we emphasized that an incremental learning algorithm must be able to learn from the newly added data while retaining the already acquired knowledge from the past data. We proposed a novel incremental learning model AILearn and showed its working mechanism on spoof fingerprint detection. The proposed algorithm is able to learn the new spoof fingerprints from the current learning phase while maintaining its performance on the "live", and "spoof" fingerprints learned in the previous learning phase.

AILearn is an adaptive way of learning as it adapts to the similarity inherently present in the data. Also, AILearn is an efficient algorithm as it discards the already seen data and keeps only knowledge extracted from it. By doing so, space can be reused for storing the upcoming data. With this motivation, we conducted our experiments and proved the efficacy of AILearn. We highlighted the stability and plasticity features of AILearn on LivDet2011, LivDet2013 and LivDet2015 dataset and concluded that the proposed framework improves its performance in the new learning phase by 49.57% on an average, without considerable degradation in the existing knowledge. This study provides critical insights into feature level comparison.

In Chapter 5, we explored the behaviour of various ensemble learning approaches to spoof fingerprint detection. We proposed A-Stacking and A-Bagging: the adaptive versions of ensemble learning approaches Stacking and Bagging, respectively. We used logistic regression as the meta-classifier in A-Stacking and showed that our results are always better than the best individual base classifier, which justifies the extra effort of employing a meta-classifier. We use A-Bagging by applying the same base learner to different subsets of data and combine their predictions using weighted majority voting. From our experimental results, we establish that the proposed adaptive models perform reasonably well on spoof detection in terms of accuracy and false positive rate.

In Chapter 6, we conducted an extensive set of experiments and did a comprehensive study on the comparison of handcrafted and deep features for the fingerprint liveness detection task. We targeted the problem as an open-set classification task where the trained model is tested on fingerprint images acquired from novel sensors that were not used in training. We evaluated the performances of various features along with different classifiers and reported overall accuracy and the spoof misclassification rate. In our findings, handcrafted features outperformed their deep features counterparts under the within-dataset category in two of the three cases. Under the cross-sensor evaluation, deep features obtained better accuracy, but handcrafted features obtained a lower spoof misclassification rate. Under the cross-dataset evaluation, no consistency in the performance of hand-crafted and deep features is observed.

In Chapter 7, the importance of automatic hate speech detection on social media is explored. We emphasize the importance of the task during the global outbreak of COVID19, where users are spending more time on social media and information is getting sensitive. With factual information, we have shown that there is a rapid growth in hate speech posts on social media, specifically targeting a community or a country and its citizens. We have also emphasized the importance of this research area during the US presidential elections. We reported the level of sincerity the social media platforms seek to improve their mechanisms for detecting hate speech without manual inspection. To address all these issues, we proposed an adaptive expert model for the task, keeping the motivation that the model must perform in cross-dataset environments. Our proposed model is constituted by deep learning methods for feature extraction and an ensemble-based adaptive classifier for predicting the class labels for the tweets. We performed our experiments on standard datasets along with latest datasets on COVID-19 and US elections and reported the performance under various experimental setups. Also, we proposed parallelized models for ensemble-based algorithms to accelerate the automatic hate speech detection of social media platforms. We used a super-computing facility for achieving this parallelization, but our proposed models can be used on any platform. We achieved reasonable speedup and efficiency as compared with the serial versions.

In future, more image processing feature types can be explored to further improve the accuracy of spoof fingerprint detection task. Also, integration of the hand-crafted and deep features in AILearn can be explored at feature and score level for further performance enhancement. In addition, efficacy of feature level fusion of hand-crafted and deep features can be evaluated for the fingerprint liveness detection. We also reported the fallacies present in the available datasets; therefore, an interesting direction of research will be to build a dataset that is more fine-grained and free from user overfitting effect in the future.