Chapter 1

Introduction

1.1 Learning Paradigms

Based on the methodology used in the training and testing phases, the learning paradigms can be broadly classified into eager learning and lazy learning, whereas ensemble learning can be considered as midway between these two extremes. In this study, we draw attention to the shortcomings of the existing learning paradigms and propose a novel paradigm EaZy Learning, which can be considered as an adaptive midway between eager and lazy learning.

1.1.1 Eager Learning

Eager learning aims to build a single functional model that captures the entire set of operating conditions underlying the dataset. As eager learning maintains only one hypothesis or a hypotheses space, it fails to provide reasonable local approximations in its target function.

The working mechanism of eager learning is illustrated in Figure 1.1. The training phase in this paradigm takes training data as input and generates a hypothesis or a hypothesis space that acts as a trained model. This model is supplied to the testing phase, where the correct discrete class/continuous value for the query instance x_q is determined. Since eager learning tries to build a general, explicit description of the target function in the training phase, it fails to provide reasonable local approximations in its target function [9]. Eager learning aims to build a single functional model that captures the

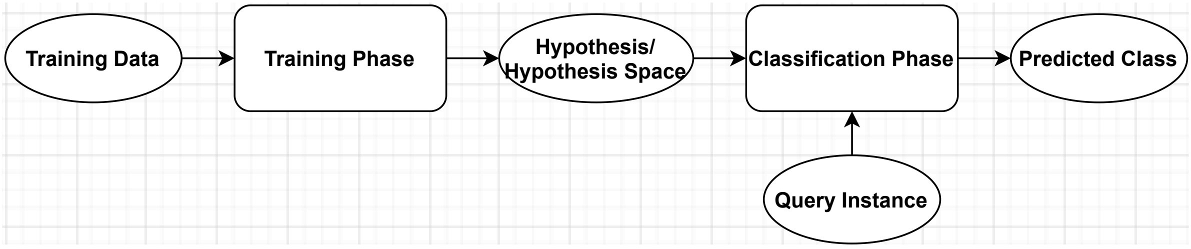

Figure 1.1: Schematic Representation of Eager Learning.

entire set of operating conditions underlying the dataset. As eager learning maintains only one hypothesis or a hypotheses space, it fails to provide reasonable local approximations in its target function. Eager learning approaches are suitable for datasets where most of the records are similar to each other, i.e. their similarity quotient should be high. These approaches perform inefficiently when encountering a dataset where all records are dissimilar to each other [10].

1.1.2 Lazy Learning

Lazy learning is a memory-based learning approach that delays all the computations until it encounters a new query instance. It stores all the training examples in the memory and locates relevant data in response to a query instance. Lazy learning approaches suffer from high storage requirements and inefficiency in the classification phase.

Lazy learning or instance-based learning delays the generalization beyond the training data until a query is made to the system [11]. Therefore, it does not perform any operations in the training phase, apart from storing the training examples. As illustrated in Figure 1.2, this learning paradigm has substantial storage requirements. Few difficulties of instance-based learning are as follows:

- These approaches are sensitive to the choice of the algorithm's similarity function.
- Lazy learning provides little insight into what has already been learned.
- Since the paradigm focuses on storing all the training examples, it is slow for large datasets.

Lazy learning approaches are suitable for datasets where records are highly dissimilar, i.e. their similarity quotient is very low.

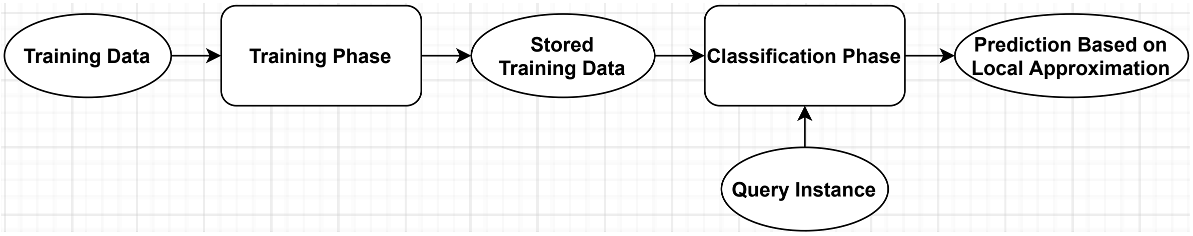

Figure 1.2: Schematic Representation of Lazy Learning.

1.1.3 EaZy Learning

Our proposed work EaZy Learning uses the same foundation as ensemble learning, i.e., instead of taking one expert's opinion, let several experts discuss and come up with a decision. In Multiple Classifier Systems (MCSs), our goal is to generate a set of k hypotheses that are accurate and diverse [12]. The generated hypotheses should be consistent with the subset of data on which they are trained and disjoint from each other. EaZy Learning does not restrict itself to a particular type of data; it is flexible enough to adapt to the problem behaviour.

As represented in Figure 1.3, EaZy learning examines the properties intrinsic to the dataset while generating a pool of hypotheses. Unlike eager learning, EaZy learning yields better local approximations in its target function, and unlike lazy learning, it requires lesser space to store the hypotheses.

EaZy learning is similar to ensemble learning as it generates an ensemble of base classifiers and integrates them to make a prediction. Still, it differs in the way it generates the base classifiers. EaZy learning generates an ensemble of entirely disjoint base classifiers, which positively influences the underlying ensemble's diversity. Also, it integrates the predictions made by these base classifiers based on their performance on the validation data.

1.1.4 Incremental Learning

Incremental learning enables the learner to accommodate new knowledge without retraining the existing model. It is a challenging task that requires learning from new data and preserving the knowledge extracted from the previously accessed data. This challenge is known as the stability-plasticity dilemma. We propose AILearn, a generic model for incremental learning that overcomes the stability-plasticity dilemma by carefully inte-

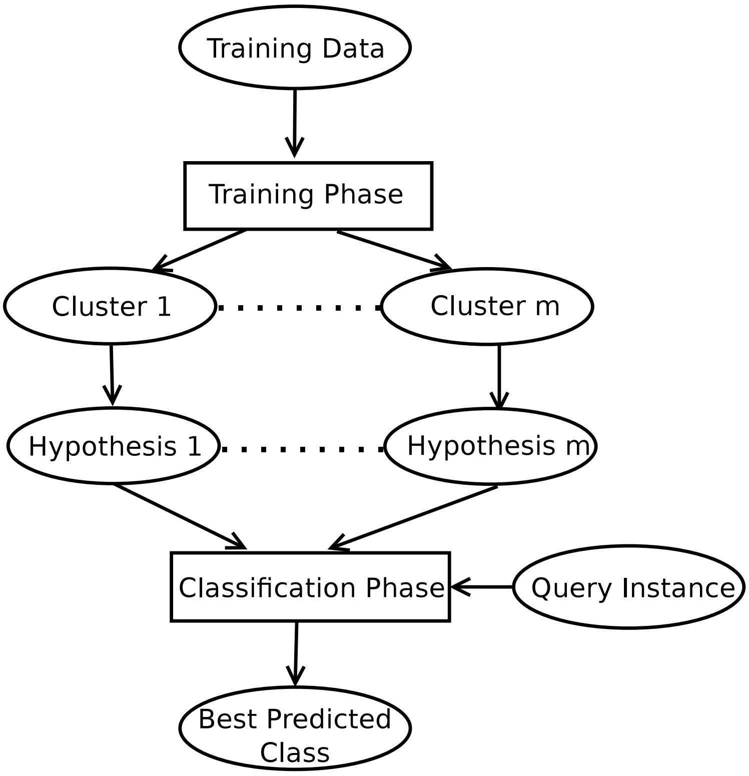

Figure 1.3: Conceptual model of EaZy learning.

grating the ensemble of base classifiers on new data with the current ensemble without retraining the model from scratch using entire data. We demonstrate the efficacy of the proposed AILearn model on spoof fingerprint detection application. One of the significant challenges associated with spoof fingerprint detection is the performance drop on spoofs generated using new fabrication materials.

AILearn is an adaptive incremental learning model that adapts to the features of the "live" and "spoof" fingerprint images and efficiently recognizes the new spoof fingerprints and the known spoof fingerprints when the new data is available. To the best of our knowledge, AILearn is the first attempt in incremental learning algorithms that adapts to the properties of data for generating a diverse ensemble of base classifiers. From the experiments conducted on standard high-dimensional datasets LivDet 2011, LivDet 2013 and LivDet 2015, we show that the performance gain on new fake materials is significantly high. On average, we achieve 49.57% improvement in accuracy between the consecutive learning phases.

1.1.5 Ensemble Based Models

Stacking and bagging are widely used ensemble learning approaches that make use of multiple classifier systems. Stacking focuses on building an ensemble of heterogeneous classifiers, while bagging constructs an ensemble of homogeneous classifiers. There exist some applications where it is essential for learning algorithms to be adaptive towards the training data.

We propose A-Stacking and A-Bagging, adaptive versions of stacking and bagging, respectively, that take into consideration the similarity inherently present in the dataset. One of the main motives of ensemble learning is to generate an ensemble of multiple "experts" that are weakly correlated. We achieve this by producing a set of disjoint experts where each expert is trained on a different subset of the dataset.

We show the working mechanism of the proposed algorithms on spoof fingerprint detection and automatics hate speech detection. The proposed versions of these algorithms are adaptive as they conform to the features extracted from the fingerprint images or the features extracted from text. We establish that A-Stacking and A-Bagging give competitive results on both balanced and imbalanced datasets from our experimental results.

1.2 Spoof Fingerprint Detection

A fingerprint liveness detector is a pattern classifier that is used to distinguish a live finger from a fake (spoof) one in the context of an automated fingerprint recognition system. As liveness detectors or presentation attack detectors are vulnerable to presentation attacks, fingerprint recognition's security and reliability are compromised. Therefore, it is essential to perform liveness detection of a fingerprint before authenticating it.

To enhance the fingerprint spoof detectors' efficiency, the learning models need to be adaptive towards the data. We propose a generic model, EaZy learning, that can be considered an adaptive midway between eager learning and lazy learning. We show the usefulness of this adaptivity for spoof fingerprint detection when fingerprints generated using unknown fabrication materials are introduced to the spoof detector.

Fingerprint liveness detection mechanisms perform well under the within-dataset environment but fail miserably under cross-sensor (when tested on a fingerprint acquired by a new sensor) and cross-dataset (when trained on one dataset and tested on another) settings. To enhance the generalization abilities, robustness and interoperability of the fingerprint spoof detectors, the learning models need to be adaptive towards the data. We show the usefulness of this adaptivity under cross-sensor and cross-dataset environments. Experiments conducted on the standard high dimensional datasets LivDet 2011, LivDet 2013 and LivDet 2015 prove the model's efficacy under cross-dataset and cross-sensor environments.

Presentation attack detection mechanisms rely on handcrafted or deep features to classify an image as live or spoof. In addition, to strengthen the security, fingerprint liveness detectors should be robust to presentation attacks fabricated using unknown fabrication materials or fingerprint sensors. In this study, we conduct a comprehensive study on the impact of handcrafted and deep features from fingerprint images on the classification error rate of the fingerprint liveness detection task. We use LBP, LPQ and BSIF as handcrafted features and VGG-19 and Residual CNN as deep feature extractors for this study. As the problem is targeted as an open-set problem, the emphasis is on achieving better robustness and generalization capability. In our observation, handcrafted features outperformed their deep counterparts in two of the three cases under the within-dataset environment. In the cross-sensor environment, deep features obtained a better accuracy, and in the cross-dataset environment, handcrafted features obtained a lower classification error rate.

1.3 Automatic Hate Speech Detection on SMPs

Social media platforms generate an enormous amount of data every day. Millions of users engage themselves with the posts circulated on these platforms. Despite these platforms' social regulations and protocols, it is difficult to restrict some objectionable posts carrying graphic content. Automatic hate speech detection on social media platforms is an essential task that has not been solved efficiently despite various researchers' multiple attempts. It is a challenging task that involves identifying hateful content from social media posts. These posts may reveal hate outrageously, or they may be subjective to the user or a community. Relying on manual inspection delays the process, and the hateful content may remain available online for a long time.

The current state-of-the-art methods for tackling hate speech perform well when tested on the same dataset but fail miserably on cross-datasets. Therefore, we propose an ensemble learning-based adaptive model for automatic hate speech detection, improving the cross-dataset generalization. The proposed expert model for hate speech detection works towards overcoming the strong user-bias present in the available annotated datasets. We conduct our experiments under various experimental setups and demonstrate the proposed model's efficacy on the latest issues such as COVID-19 and US presidential elections. In particular, the loss in performance observed under cross-dataset evaluation is the least among all the models. Also, while restricting the maximum number of tweets per user, we incur no drop in performance.

To deal with large-scale data efficiently and accurately, we need a simple, scalable and robust framework. Therefore, we propose parallelization to the standard ensemble-based algorithms so that they can be used for speeding up the automatic hate speech detection on SMPs. This study parallelizes bagging, A-stacking, and random sub-space algorithms and tests both serial and 'parallel versions on the standard high-dimensional datasets for hate speech detection. We observe a significant speedup with high efficiency that claims that the proposed models are suitable for the considered application. We observed that the accuracy is not affected while parallelizing the algorithms compared with serial algorithms executing on a single machine.

The central motivation of this dissertation is to highlight the need for considering the inherent characteristics of data for generating the base learners of the ensemble. We emphasize that the learning models must conform to the underlying dataset to achieve better local approximation for the pattern mining applications. In these applications, where the task is to find a specific pattern in a massive amount of data, models' adaptiveness towards data properties plays an important role.

We evaluate our proposed models' performance on two different pattern mining applications: spoof fingerprint detection and automatic hate speech detection on social media platforms. As these applications are different in nature, they show the generalization abilities of our proposed models.

In addition, our focus is on cross-dataset evaluation of the models. We claim that for the chosen applications, it is always required to test the models' performance under various test scenarios. Therefore, to test the robustness and reliability of the proposed models in real-world situations, we test them under a cross-dataset environment.

1.4 Structure of the Thesis

This thesis is arranged into eight different chapters. A brief description of the chapters is as follows:

Chapter 2 presents a survey on various learning models and the current state-of-the-art for the two pattern mining applications.

Chapter 3 gives an introduction to our proposed model EaZy learning. Here, we discuss the architectural details of the model along with experimental results and analysis.

Chapter 4 discusses the importance of incremental learning and introduces the proposed model AILearn, an adaptive incremental learning model.

Chapter 5 introduces A-Stacking and A-Bagging, the adaptive versions of the existing ensemble based models stacking and bagging respectively.

Chapter 6 presents a comprehensive study on the impact of handcrafted and deep features on spoof fingerprint detection. We conduct a rigorous analysis of various types of features when used under different test scenarios.

Chapter 7 discusses automatic hate speech detection on social media platforms and its importance during the global pandemic COVID-19. It also presents the parallelized versions of various ensemble based models.

Chapter 8 brings the concluding remarks and suggestions for future work.