Innovative Outlier Removal Techniques to Enhance Signature Authentication Accuracy for Smart Society

Anindita Desarkar, Jadavpur University, Kolkata, India
Shisna Sanyal, Jadavpur University, Kolkata, India
Anupam Baidya, Jadavpur University, Kolkata, India
Ajanta Das, University of Engineering & Management, Kolkata, India
Chitrita Chaudhuri, Jadavpur University, Kolkata, India

ABSTRACT

A smart society is an empowered society, which can improve the lives of its citizens by using the latest innovations and technologies. This improvement can happen in several dimensions out of which security is a major one. Inconsistency and forgery are very common phenomenon where handwritten signatures are often preserved for training a classifier to authenticate a person. The removal of outliers, at the outset, obviously improves the quality of training and the classifier. The present article deals with the mechanized segregation of the poor-quality authentic signatures from reliable ones. Machine learning algorithms for outlier handling utilizing clustering, classification and statistical techniques have been implemented in this context. Subsequent performance evaluation after outlier removal reflects improvement of both true positive and true negative recognition rate accuracy. The performance evaluation presents the significant differences between authentication accuracy and forgery accuracy in the context of building a safe, secure and smart society.

KEYWORDS:
Authentic Accuracy, Bio-metric Authentication, Classification Techniques, Clustering Techniques, Forgery Accuracy, Outlier Detection, Overall Accuracy, Smart Society, Statistical Techniques

1. INTRODUCTION

The major objective of a smart society is to use technology for the benefit of its people. People need smarter and secured solutions in various sectors like Banking, Public Examination Systems and vigilance through Government Identity Cards to name only a few amongst the vast area of applications to fight with the fraudulent activities. Automatic biometric authentication is especially beneficial in these areas and checking offline signatures is one of the cheapest and simplest techniques employed to verify a person. But manual checking frequently involves both intentional and unintentional errors triggering from factors such as vested interest or tedium. Customer satisfactions are also often jeopardized by human interactions. An impersonal mechanized decision, on the other hand, may sometimes be conveniently utilized to ward off unpleasant situations. Hence, we feel the crying need to introduce machine learnt techniques to perform such tasks in a ubiquitous environment.

DOI: 10.4018/IJDST.2019040104
Incidentally, although handwritten signatures are still the most widely used and popular mode of identifying a person throughout the literate world, the process involves major issues resulting from inconsistency and forgery. Inconsistency refers to the differences present between two signatures of the same person due to various reasons like ageing, sudden problem in writing hand, mood fluctuations etc. This is the biggest motivation for detecting outliers. This inspires use of various machine learning algorithms in this context. These machine learning algorithms majorly include classification and clustering techniques which are used to improve authentication accuracy by detecting poor quality signatures. Automatically, decisions become more appropriate and efficient which helps to build a smarter society.

An outlier refers to a data point that is considerably dissimilar from other data points belonging to a dataset. It is sometimes generally referred to as error or abnormality or anomaly, although there are occasions where such exceptional data may itself be of paramount importance. So, detecting outliers often plays a major role in any decision-making process associated with the dataset. For classification tasks, the presence of incongruous data will enhance the risk of getting inappropriate class values. Circumstances leading to generation of outlier data are usually multifarious in nature. These may range from natural occurrences such as changes in system behavior, to faulty instruments, intentional misrepresentation, or simply due to natural deviations in populations, occurring by chance in any distribution, heavy-tailed or otherwise.

There already exist, across the world, numerous outlier detection systems for finding network related issues, credit card frauds, defect in medical diagnosis and image analysis, to name just a few – but surprisingly the role of poor-quality training samples in the area of bio-metric identification, utilizing offline handwritten signatures in particular, have not been explored to the fullest extent. Yet, as we all know, no two signatures of the same person can ever be exactly similar. One can add to this the fact that the handwritten signatures of even the most meticulous person can deviate grossly from the normal, even within a short period of time, because of causes both physical as well as psychological. The justification behind such aberrations occurring may be quite trivial, such as cramped and uncomfortable writing space. On the other hand, they can also be rooted in major factors contributed by debilitating diseases and traumatic conditions.

As established earlier, one of the pitfalls involved in handwritten signatures being used to prove identity of a person, is the chance of malpractices such as presentation of forged signatures. In a smart environment, the way to avoid the circumstance can only be achieved by utilizing classification techniques to segregate the sound and authentic signatures from the fraudulent ones, for each and every person individually.

All classification algorithms employ supervised learning by partitioning data with known class values into training and test sets. The outcome of the learning procedure is a classifier which should be able to discriminate between two specific categories of signature images – the good or authentic ones of a person, and the bad or forged ones. The accuracy of the classifier is determined by its ability to decipher the true positive ones or the good ones, as well as the true negative ones or the bad ones. Hence the test set comprises of both types of signatures of a person. But the training set, which is the major portion of data, cannot contain anything but authentic data, because philosophically a classifier cannot be taught to recognize all types of forgery. Not only are the possibilities endless, moreover it goes against the grain of all ethics to train up a classifier with fraud values at its nascent state.

So, the classifier needs to discriminate the bad from the good by the standards set by the good alone. Hence the strategy is to identify the good ones accurately. Anything outside the range of true good would automatically be classified to be bad. But then the standard of the true goods must be flawless. Hence it is absolutely mandatory to set up the training set immaculately – no outliers can be tolerated, as they would invariably lead to poor decisions during the test phase. This is the principal motivation behind the present research work.

Here our objective is to identify the truly inconsistent signatures as outliers within the authentic training-set for every persons’ signature images. One common and easy method to detect such
Cloud Computing Security: Opportunities and Pitfalls
[www.igi-global.com/article/cloud-computing-security/62997?camid=4v1a](www.igi-global.com/article/cloud-computing-security/62997?camid=4v1a)

Rough Sets and Boolean Reasoning
[www.igi-global.com/chapter/rough-sets-boolean-reasoning/28467?camid=4v1a](www.igi-global.com/chapter/rough-sets-boolean-reasoning/28467?camid=4v1a)