

Feature Engineering Techniques to Improve Identification Accuracy for Offline Signature Case-Bases

Shisna Sanyal, Jadavpur University, India

Anindta Desarkar, Jadavpur University, India

Uttam Kumar Das, Tata Consultancy Services, India

Chitrita Chaudhuri, Jadavpur University, India

ABSTRACT

Handwritten signatures have been widely acclaimed for personal identification viability in educated human society. But, the astronomical growth of population in recent years warrant developing mechanized systems to remove the tedium and bias associated with manual checking. Here the proposed system performs identification with nearest neighbor matching between offline signature images which are collected temporally. The raw images and their extracted features are preserved using case-based reasoning and feature engineering principles. Image patterns are captured through standard global and local features, along with some profitable indigenously developed features. Outlier feature values, on detection, are automatically replaced by their nearest statistically determined limit values. Search space reduction possibilities within the case base are probed on a few selected key features, applying hierarchical clustering and dendrogram representation. Signature identification accuracy is found promising when compared with other machine learning techniques and a few existing well-known approaches.

KEYWORDS

Case-Based Reasoning, Dendrogram, Feature Engineering, Hierarchical Clustering, Identification Accuracy, Image Case-Base, Nearest Neighbor Classification, Outlier Detection

1. INTRODUCTION

From ancient times, handwritten signature is the most well-known biometric characteristic for appropriate identification of a person or to authenticate a document. Biometrics is broadly categorized into two sections: behavioral and physiological. Handwritten signatures belong to the first category. The mode of collection is also the easiest and cheapest. For these reasons, it has been one of the most popular techniques favored so far. From financial and business transactions to invigilation in examination hall, this mode of identification has been used everywhere most profusely. Besides authenticating one's identity, other application areas include transaction confirmation, civil law contracts, acts of volition, personal cards, administrative forms, formal agreements, acknowledgement of received services etc. The wide and huge usage area obviously demand automatic identification

DOI: 10.4018/IJRSDA.20210101.oa1

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

technique, as in the present era of data avalanche, it would require abhorrent amounts of time and effort, if performed manually.

Acquiring training data is a continuous process – the signatures collected would usually come from different form-filling sessions. Signature of a person tends to changes with time – over and above the fact that no two signatures of a person are exactly the same. Hence the proposed approach has been made adequately robust and efficient to support the need of a heterogeneous and dynamic environment.

According to the data acquisition mechanism, signature identification system is classified into two types: online and offline. Online method needs special set of devices and instruments to capture the pen movements and pressure over the digital medium at the time of writing, thus involving sophisticated and costly tools. On the other hand, the offline technique at most needs a scanner or a digital camera at the input end to receive a digital representation of the signature in the pixel form corresponding to the grey level intensity at each point within the signature image. In this research work, external media such as a piece of paper have been used to capture the signature, which has thereafter been scanned to receive its digitized copy prior to storing it within the system. However, recent mobile devices allow easy capturing of online signatures, which can well be utilized to accommodate dynamic update of training set and test set data.

The objective of our research is to build a classifier which helps to detect the identity of a person, where training begins by comparing presented signature with each case or person preserved in the base. In an ideal situation, the training process is supposed to be improved if a prior clustering process partitions the case base into further segments, depending on some key feature values, leading to reduced searching and comparison times. For this purpose, a hierarchical clustering technique, using dendrograms, have been examined here to survey the prospects. The feature values extracted from the images have been discretized and the prospect of capping them within normal limits to exclude abnormal values, have been studied separately. Avoiding outlier values have been experimentally found to improve accuracy of identification, which has been one of the primary motivation behind the present work.

In our proposed work, Case Based Reasoning (CBR) technique is deployed to utilize incremental learning procedure at the beginning. The machine is trained, on sample signatures of each person preserved within a case base, which are utilized to recognize a particular person whose authentic signature is posed as a problem to be solved by the system. As a further motivating factor, it may also be mentioned here that, as the literature survey in the following section reflects, no comparative work in the domain utilizes CBR methodologies. This novel approach, of utilizing CBR techniques with straight-forward feature engineering in the form of manipulative outlier handling, has been truly justified by the accuracy results obtained with an indigenous dataset accumulated by the present researchers. Here signatures are stored as the combination of various attributes fetched from the captured images. In the domain of machine learning, features play a very important role, as recognition primarily depends on them. The terms attribute and feature shall be used interchangeably in the present document. These collected features may be considered to be the part of a set problem, the solution to which is the identity of the signatory attached as the class value. Once the newly posed part of a problem is matched sufficiently with an already preserved part of a problem in a case, the solution part of the case is the required output in the form of personal identity. The significance of proper Feature Engineering, and Outlier Handling therein, is thus self-evident under the circumstance.

Following Section 2 highlights a few researches on signature identification. Section 3 describes some basic concepts associated with the proposed system. The proposed techniques for identification are presented in Section 4. It also includes the detailed description of datasets and system tool configurations. Experimental results are depicted in Section 5. The final section concludes with inferences drawn from the results, as well as some future directions of research which may further augment the accuracy and performance of the system.

17 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/feature-engineering-techniques-to-improve-identification-accuracy-for-offline-signature-case-bases/273727

Related Content

Enhancement of TOPSIS for Evaluating the Web-Sources to Select as External Source for Web-Warehousing

Hariom Sharan Sinha (2018). *International Journal of Rough Sets and Data Analysis* (pp. 117-130).

www.irma-international.org/article/enhancement-of-topsis-for-evaluating-the-web-sources-to-select-as-external-source-for-web-warehousing/190894

Concept and Practices of Cyber Supply Chain in Manufacturing Context

Anisha Banu Dawood Ganiand Yudi Fernando (2018). *Encyclopedia of Information Science and Technology, Fourth Edition* (pp. 5306-5316).

www.irma-international.org/chapter/concept-and-practices-of-cyber-supply-chain-in-manufacturing-context/184234

The Analysis of Instrument Automatic Monitoring and Control Systems Under Artificial Intelligence

Qinmei Wang (2024). *International Journal of Information Technologies and Systems Approach* (pp. 1-13).

www.irma-international.org/article/the-analysis-of-instrument-automatic-monitoring-and-control-systems-under-artificial-intelligence/336844

In-Service Teachers' Use of ICT for the Promotion of Collaborative Professional Learning

Ana García-Valcárcel and Juanjo Mena (2018). *Global Implications of Emerging Technology Trends* (pp. 130-144).

www.irma-international.org/chapter/in-service-teachers-use-of-ict-for-the-promotion-of-collaborative-professional-learning/195826

Fog Caching and a Trace-Based Analysis of its Offload Effect

Marat Zhanikeev (2017). *International Journal of Information Technologies and Systems Approach* (pp. 50-68).

www.irma-international.org/article/fog-caching-and-a-trace-based-analysis-of-its-offload-effect/178223