Electrical and Computer Engineering Faculty Publications

Electrical & Computer Engineering

10-11-2018

Application of Machine Learning to Biometric Systems- A Survey

Lina Chate
University of Nevada, Las Vegas

Shahram Latifi
University of Nevada, Las Vegas, shahram.latifi@unlv.edu

Follow this and additional works at: https://digitalscholarship.unlv.edu/ece_fac_articles

Part of the Electrical and Computer Engineering Commons

PAPER • OPEN ACCESS

Application of Machine Learning to Biometric Systems- A Survey

To cite this article: Lina Chato and Shahram Latifi 2018 *J. Phys.: Conf. Ser.* **1098** 012017

View the article online for updates and enhancements.

IOP ebooks™

Bringing you innovative digital publishing with leading voices to create your essential collection of books in STEM research.

Start exploring the collection - download the first chapter of every title for free.

Application of Machine Learning to Biometric Systems- A Survey

Lina Chato and Shahram Latifi

Dept. of Electrical and Computer Engineering, UNLV, 4505 Maryland Parkway, Las Vegas, NV 89154, USA

Shahram.latifi@unlv.edu

Abstract. A summary of existing work on the use of machine learning and deep learning methods in biometrics is presented here. Biometrics traits covered include physiological (image, voice) as well as behavioral (gait, signature) features. This study shows that machine learning has a high potential to improve the performance of biometrics systems due to ML's ability to mine, search and analyze big sets of data, performing matching tasks more quickly and reliably than the conventional methods. At the end some key challenges in use of adopting ML in biometrics systems are pointed out.

1. Introduction

Work on Machine Learning (ML) has been accelerated in recent years. Due to their ability to mine, analyze and interpret large files, ML methods have been applied to many disciplines including image processing and biometrics. The basis for any biometrics method is some kind of matching methods which typically are one to one in case of verification and one to many in case of identification. On the other hand, ML methods have shown to improve the performance of biometrics systems and their full potential in helping biometrics achieve 100% performance has not been explored sufficiently. We present, in this paper, a brief survey of existing literature that applied the ML methods to biometrics.

2. State of the Art Techniques

Machine learning (ML) methods have been increasingly employed in many applications. In particular, ML has played a significant role in improving the performance of biometric systems. With embedded ML in biometric systems, sometimes tedious tasks such as one to one or one to many matching tasks can be done automatically and seamlessly. In particular, Deep Learning (DL), a specific ML approach based on neural nets composed of many layers, has been used in different biometrics applications. DL methods exhibit the ability to create robust and reliable authentication models that at times outperform the state-of-the-arts systems as pointed out by some researchers. This work presents a survey of existing literature on how ML can be used to improve the performance of biometrics systems.

In [1], the authors proposed a one-shot learning facial recognition method (DeepFace) based on deep Siamese network and support vector machine (SVM) classifier. They used three datasets to train and test their model. Each dataset consists of pairs of imagesthat belong either to the same person or different persons. A complex 3D alignment process was applied to each pair of images and then fed to Siamese network for extracting features from each image. The absolute difference between the features of each pair of images is computed. Next a fully connected layer is used to map the result into a single logistic unit based on a binary classifier to classify each pair intothe same-person or different-

person classes. This is followed by application of the SVM instead of logistic regression units to complete the classification process. The Social Face Classification (SFC) dataset (collected from Facebook) was used to train the proposed method, and Labeled Faces in the Wild (LFW) dataset which is the benchmark dataset for face verification and the YouTube Face (YTF) dataset were used to test the proposed method. The results indicated that the proposed method outperforms the state-of-the –art technique [1]achieving performance close to human's was (the accuracy of the proposed method 97.35% in LFW dataset compared to 97.5% human cropped, and 91.4% in YTF dataset).

In [2], the deep CNN based on triplet inputs was proposed for one shot-learning. The authors proposed to use 3 input images (person, false positive person, false negative person) to train the CNN. The features are calculated for each of 3 images and then a difference of (person, positive person) features, and (person, negative person) features were calculated to compute very high accuracy ≈99% compared to 97% reported in [1]. This method tested with slandered face recognition datasets (Labeled Faces in the Wild (LFW) dataset, YouTube Faces DB) and achieved the state-of-the-art in recognition performance using only 128 byte per face (LFW accuracy 99.63%, and YouTube Faces DB accuracy 95.12%). Also, this method used minimal alignment compared to [1] which used a complex 3D alignment to extract the facial features. The authors used the proposed method as another application to cluster a users' personal photos into groups of people with the same identity based on agglomerative clustering (This is a "bottom up" approach: each observation starts in its own cluster, and pairs of clusters are merged as one moves up the hierarchy [3]). This application showed no sensitivity to occlusion, lighting, pose and even age variations. As an application of semanese Neural Net, twin-face verification was proposed in [4].

The authors in [5] proposed a supervised deep learning structure called Deep Irisfor iris verification based on pair filter layer (PFL) and convolutional layers. Two standard datasets (Q-FIRE and e CASIA cross sensor datasets) were used in this work. These datasets consist of heterogeneous iris images (collected from different sensors, different distances (resolutions), and different illumination) labeled in two classification classes (intra-class pair (0), and inter-class pair (1)). An input to the Deep learning network is a pair of images which belong to either the same person or different persons. The similarity map is generated by involving a pairwise filter layer as a first layer to convolve the images and summarize the feature map by using a number of filters. Three different structures based on PFL were proposed and tested to produce best iris verification system, and then the results were compared with the state-of-the –art techniques. The results from the best structure showed increasing the accuracy and decreasing the Equal Error Rate (EER) to 1/10 compare to other methods with a high genuine accept rate (GAR) (≈95%)and very low FAR (≈ 10-4).

There has been some work on application of ML to voice biometrics as well. A Voiceprint Authentication System based on DL was proposed in [6]. Raw audio data is not the common input for speech/speaker recognition applications; so, instead of raw audio data, the features extracted from raw audio data is used for identification. The Mel-Frequency Cepstral Coefficients (MFCCs) are one set of features based on frequency representation that are extracted from audio raw data, and the human feature voice has the range of frequencies 85-255 Hz. DNN was proposed in [6] for voice classification trained by MFCCs features (the input to DNN).

The process of modeling the proposed system is described in the following steps:

1. All the audio samples need to have the same length to create same length feature vector. Therefore, the first step is represented by segmenting the audio samples in a specific length to n-second length segments.
2. The MFCC calculation is applied to each segment to extract the required features. The steps of extracting this type of features are shown in Figure. 1. The details about how to create the feature vector are well explained in [6].
3. After applying the Discrete Cosine Transform (DCT) (as the last step of producing MFCC data), the lower DCT coefficients are kept. These coefficients describe the features of the input audio signal. The feature data for each audio sample is represented by matrixMxN where M

represents the number of frames in each audio segment and N represents the number of coefficients kept from MFCC calculation.

4. The MxN feature-matrix is fed to DNN (its structure is shown in [6])
5. The authors used devclean dataset from LibriSpeech which contains 40 different speakers and each speaker has 8 minutes of speaking time. In the first experiment, they used only data for 10 speakers and feed them to a Neural Network to classify them into corresponding individuals. The authors studied different numbers of coefficients to come up with the best model. A total of 40 speakers were involved in training process. However increasing number of the speakers called for an increase in complexity of NN, and the best accuracy achieved was 97% for the training set and 95% for the test set. Also, the results indicated that increasing the time of audio file from 2 sec to 4 sec increased the accuracy.
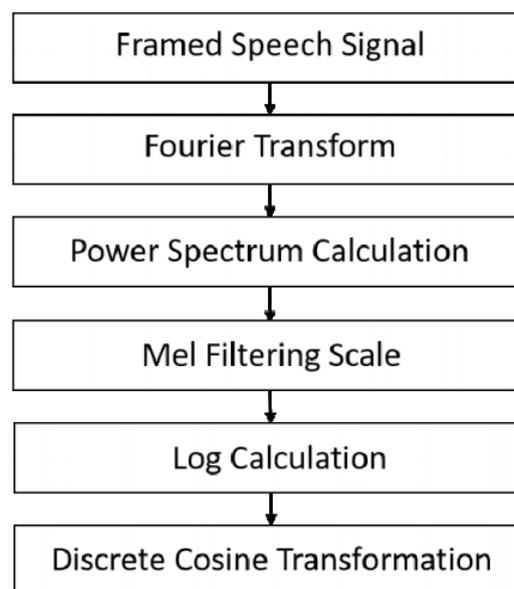

Figure. 1 Mel-Frequency Cepstral Coefficient Calculation Steps [6]

The classical gait identification methods typically use sequences of binary silhouettes to produce gait signatures [7]. However, the authors in [8] used different type of features as the noiseless silhouettes were not easy to obtain. The authors used the optical flow components to describe a gait signature. The optical flow components (spatio-temporal cuboids) were used as inputs to CNN for feature extraction process and then these deep features were used to train an SVM classifier to classify the produced signature gait to a specific individual.

The author tested their proposed method with TUM-GAID dataset which is the standard dataset used in the gait state-of-the-art. The results of this method are very close (same) to the results from state-of-the-art even with using images with 8 times lower resolution.

The authors proposed the following pipeline to develop their proposed methods: (i) compute optical flow (OF) along the whole sequence; (ii) build up a data cuboid from consecutive OF maps; (iii) feed the CNN with OF cuboid to extract the gait signature; and, (iv) apply a classifier to decide the subject identity. However, the following points have been taken into consideration

1. CNN requires a specific input vector size for all input samples. Therefore, the subsequences of L frames were extracted from the full length sequences for both x and y direction optical flow components.
2. The RGB original videos were reduced to 80×60 pixels for each frame, and then the Optical Flow (OF) components were obtained from the reduced videos' size.

3. By removing a part from background, the frame size was reduced 60×60 pixels from the OF as to produce the OF map. Also the overlap 80 % was applied to produce a subsequence of 25 frames for each sample. Therefore, the input to CNN is two sequences of 60x60x25 (one sequence for x component, and the other for y component) or 60x60x50

4. CNN was used to obtain the gait signature (from the last fully connected layer) with softmax classification unit, or feed it to binary linear SVM classifier (instead of softmax) to classify each class one-vs-all. Also, Nearest Neighbor was proposed as alternative classifier which does not require any training process. This will be useful to extend the gait recognition system by adding new samples to the model.

5. All the proposed three classification models were used with three experiments (gait recognition with clothing and carrying condition, elapsed time, and gait-based gender recognition)

The following should be pointed out regarding this work:

a) TUM-GAID dataset consists of people walking indoors under four walking conditions: normal walking, wearing coats, carrying a bag and wearing coating shoes. Also, the same set of subjects are included in different months of the same year. This database contains a total of 305 individuals, and the size of the frame is 640 x 480 pixels with approximate frame rat 30 fps.

b) The whole training process took about 60 hours.

A Multi-view Multi-Class Framework (MVMC) for User Identification on Mobile Devices called "DEEPSERVICE" was proposed in [9] as an efficient Identification method which takes less than 1 ms to perform each identification process. The authors collected mobile data based on user's keystroke information. The sequential tapping information and accelerometer information were collected from 40 users from smartphones for 8 weeks. The collected data set into multiple views (a view of alphabets, a view of numbers and symbols, and a view of tapping acceleration). The view of alphabets was represented in 3 features: Duration, Time since last key, and Number of keystrokes per session. The view of numbers and symbols was represented by 3 features: Medium of frequent key, Maximum of frequent key, and Radio of frequent key. The view of Acceleration features was based on the correlation of different directions of acceleration. The proposed deep structure was based on Gated Recurrent Unit (GRU) bidirectional recurrent Neural network (RNN) which is a special case of RNN with a simple structure. The GRU was used to model the hidden layer for each view, and then Softmax function is used in the output layer for multi-classification task. The accuracy and F1-score were used to measure the performance of the proposed identification system "DEEPSERVICE" with different number of users (2, 5, 10, and 26). The results showed that the proposed method outperformed other approaches(SVM, Decision Tree, Random Forest, Logistic Regression), and achieved accuracy 99% for 2-users identification, 93.51% for 5-users identification, 87% for 10-users identification, and 82% for 26-users identification. It is clear from the results that the accuracy and F1 score decreased with increasing number of users. Therefore, increasing the length of each sample, or/ and increase the complexity of DNN could be a solution to prevent decreasing the accuracy. However, the achieved accuracy of the proposed method still has very promising achievements compare to other approaches. It should be noted that:

a) DEEPSERVICE is not the fastest model but the decision tree model is faster. However, it only takes about 0.657 ms per session which shows its feasibility of real-world usage.

b) The tool can be implemented on the web or the router.

3. Challenges of using ML in biometrics

It is more appropriate to consider the following points as challenges of using ML in biometrics instead of disadvantages due to rapid developments in the ML field which include optimized structures, new deep structures, advanced and cloud-based devices for accelerating computations.

1. To produce a good model needs a general and large dataset.
2. No general ML algorithms and structure can be used for all biometric systems.

3.   Unreliable results due to noisy environments, sensors' problems, and/or acquisition problems.

4.   Optimizing ML parameters is not easy

5.   Time consuming: Need preprocessing method(s) to prepare and improve the quality of datasets. Also, some of these methods are based on deep learning methods. For example, face recognition process needs face detection, segmentation, denoising, and pose and illumination corrections.

6.   Huge memory to store the database.

7.   Need continuous developments of the used method to resist attackers and hackers.

8.   Need to develop antispoofing methods due to increasing availability of synthesized data.

9.   Although some DL models achieved a high accuracy, still additional data is required to develop and retrain identification and recognition models.

References:
[1] Taigman Y, Yang M, Ranzato M, and Wolf L 2014, "DeepFace: Closing the Gap to Human-Level Performance in Face Verification,"*2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp.1701-1708.

[2] Schroff Ϝ , Kalenichenko D , and Philbin J 2015, "FaceNet: A Unified Embedding for Face Recognition and Clustering," *2015 IEEE Conference on Computer Vision and Pattern Recognition* (CVPR)

[3]Wikipedia:Hierarchical,Clustering https://en.wikipedia.org/wiki/Hierarchical_clustering

[4] Sun X, Torfi A, and Nasrabadi N 2018, *"Deep Learning in Biometrics, Chapter 3: Deep Siamese Convolutional Neural Networks for Identical Twins and Look-Alike Identification*", 5 March 2018.

[5] Nianfeng Liu, Man Zhang , Haiqing Li, Zhenan Sun, and Tieniu Tan, "DeepIris: Learning pairwise filter bank for heterogeneous iris verification," Pattern Recognition Letters, Vol. 82, Part 2, pp. 154-161, 15October.

[6] Boles A, and Rad P 2017, "Voice Biometrics: Deep Learning-based Voiceprint Authentication System," *2017 12th System of Systems Engineering Conference (SoSE)*, Waikoloa, HI, 2017, pp. 1-6.

[7] Wang C, Zhang J, Wang L, Pu J, and Yuan X 2012, "Human identification using temporal information preserving gait template," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.34,no.11,pp.2164–2176.

[8] Castro F, Marin-Jimenez M, Guil N, and de la Blanca, N P 2016 "Automatic learning of gait signatures for people identification," *arXiv preprint arXiv:1603.01006.*

[9] Sun L, Wang Y, Cao B, Yu P S, Srisa-an W, and Leow A D 2017, "Sequential keystroke behavioral biometrics for user identification via multi-view deep learning," in *Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML/PKDD).*