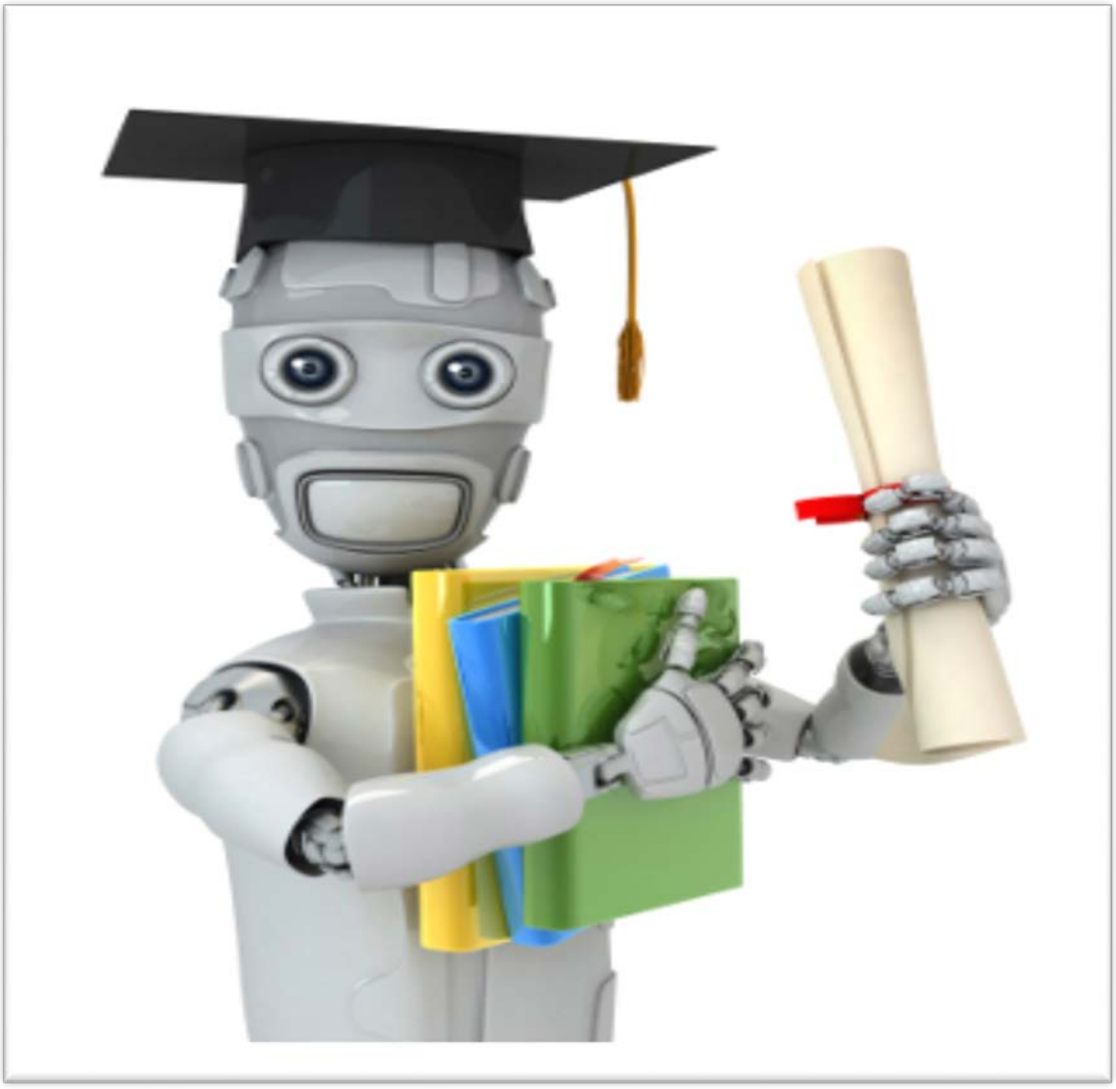
Celebrity Identification and Recognition in Videos

An application of semi-supervised learning and multiclass classification



Contents

Members: **Harshit** (72016) & **Vaibhav Sharma** (72151)

Mentor: Prof. Shobha Bagai, Professor, Cluster Innovation Center

Aim

A media player plugin having features that enables a user to identify a celebrity at any instance of time by clicking on the running video. Unlike Amazon's x-ray the plugin should be able to detect the faces, recognize the celebrities through semi supervised learning algorithms and does not require a pre built-in data for this purpose.

Motivation

An app launched by Amazon X-Ray for movies and television recently got our attention that works on its Kindle Fire series and lets you identify actors with a single tap. We thought that if there was a mechanism working the same way on the internet then the information about a video can be known before-hand. Identification of actors in movies can help in many applications such as video indexing, actor-specific scene retrieval, etc. However, the problem is very challenging due to large variations in appearance, pose, facial expressions, occlusions and camera motion. To solve the problem we initially worked on Semi-supervised learning methods like Modified Adsorption (MAD) and Low Density Separation (LDS), the Algorithms have a lot in common in terms of working and have high accuracy too. This led us to research a bit in the field of integrating both algorithms and making a Hybrid Algorithm "MAD plus LDS" for solving our problem.

Research

The project required deep study of Graph based semi supervised learning algorithms to be able to integrate the two algorithms. We along with our mentors took various steps to ensure our direction of research is both valid and feasible. The initial Aim of integrating both the algorithms was completed from framing the algorithm to successfully implementing it. The results were in accordance with the theoretical expected values and are shown in the later.

The algorithms are briefly explained below :

ADSORPTION –

Adsorption is one such recently proposed graph based semi-supervised algorithm which has been successfully used for different tasks, such as recommending YouTube videos to users and large scale assignment of semantic classes to entities within Information Extraction. Adsorption has many desirable properties:

- It can perform multiclass classification
- It can be parallelized and hence can be scaled to handle large data sets which is of particular importance for semi-supervised algorithms.

Adsorption is a general algorithmic framework for transductive learning where the learner is often given a small set of labeled examples and a very large set of unlabeled examples. The goal is to label all the unlabeled examples, and possibly under the assumption of label-noise, also to relabel the labeled examples.

The method corresponds to an undirected graph $G = (V, E, W)$, where a node $v \in V$ corresponds to an example, an edge $e = (a, b) \in V\ X\ V$ indicates that the label of the two vertices $a, b \in V$ should be similar and the weight $W_{ab} \in R_+$ reflects the strength of this similarity.

We have the nodes and edges to build the undirected graph for the above method which we get from our datasets the following are the sources of our data set –

- IMFDB
- Yale Cropped

LDS –

The goal of semi-supervised classification is to use unlabeled data to improve the generalization. The cluster assumption states that the decision boundary should not cross high density regions, but instead lie in low density regions. We believe that virtually all successful semi-supervised algorithms utilize the cluster assumption, though most of the time indirectly.

We enforced the cluster assumption by the graph based similarities, the results of which we got in the form of a weight matrix having weights based on the similarities of the adjoining nodes. We used Gaussian method instead of cosine similarity to carry out the above mentioned task.

Algorithm –

1. Build nearest neighbor graph G from all (labeled and unlabeled) data.

2. Compute the $n \times (n+m)$ distance matrix D^{ρ} of minimal ρ-path distances according to

$$D^{\rho}_{i,j} = \frac{1}{\rho^2} \ln \left(1 + \min_{\mathsf{p} \in \mathsf{P}_{i,j}} \sum_{k=1}^{|\mathsf{p}|-1} \left(e^{\rho d(\mathsf{p}_k, \mathsf{p}_{k+1})} - 1 \right) \right)^2$$

   from all labeled points to all points.

3. The above distance matrix can be renamed as W_{ab} to be used in the modified adsorption algorithm.

Input:
- **Graph:** $G = (V, E, W)$
- **Prior labeling:** $\mathbf{Y}_v \in \mathbb{R}^{m+1}$ for $v \in V$
- **Probabilities:** $p_v^{inj}, p_v^{cont}, p_v^{abnd}$ for $v \in V$

Output:
- **Label Scores:** $\hat{\mathbf{Y}}_v$ for $v \in V$

$\hat{\mathbf{Y}}_v \leftarrow \mathbf{Y}_v$ for $v \in V$ {Initialization}
$\mathbf{M}_{vv} \leftarrow \mu_1 \times p_v^{inj} + \mu_2 \sum_{u \neq v} (p_v^{cont} \mathbf{W}_{vu} + p_u^{cont} \mathbf{W}_{uv}) + \mu_3$
repeat
    $D_v \leftarrow \sum_u \left(p_v^{cont} \mathbf{W}_{vu} + p_u^{cont} \mathbf{W}_{uv} \right) \hat{\mathbf{Y}}_u$
    for all $v \in V$ **do**
        $\hat{\mathbf{Y}}_v \leftarrow \frac{1}{\mathbf{M}_{vv}} \left(\mu_1 \times p_v^{inj} \times \mathbf{Y}_v + \mu_2 \times D_v + \mu_3 \times p_v^{abnd} \times \mathbf{r} \right)$
    end for
until convergence

Innovation

The project was meant to develop a software with new techniques given that there has not been any attempt to use Semi supervised learning techniques for celebrity identification in videos. We pushed ourselves out of the boundaries of theoretical work to design and develop the workflow of our software. We in our work not only used Semi supervised learning for face recognition but took the thought one step further to use it for famous celebrity identification in videos with absolutely no initial information about video.

Commercialization possibilities

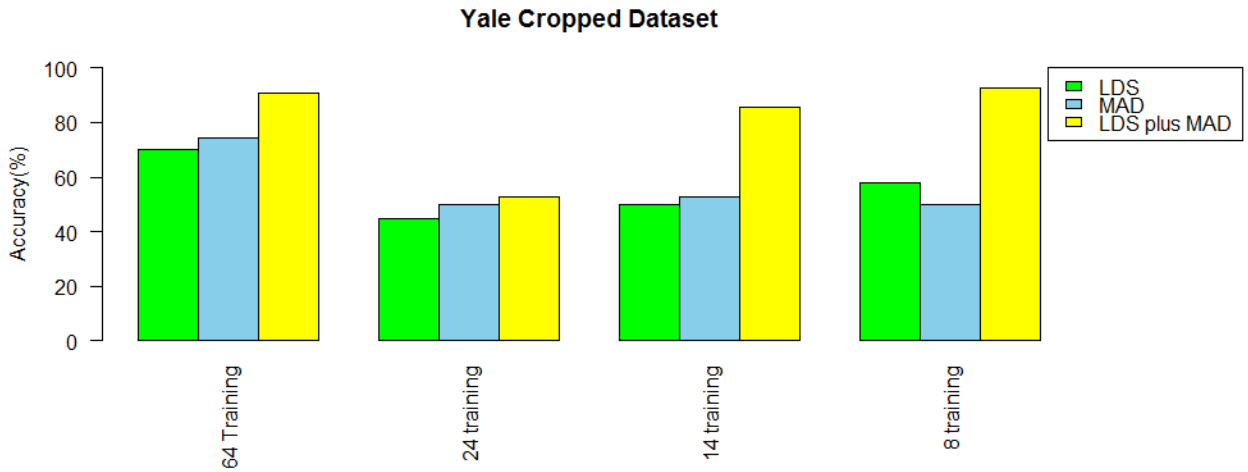
No Software present in the market can outperform the accuracy and speed of our final software even in theoretical terms. The Final Software that we are building is highly flexible and its only competitor would be Amazon's X-Ray app. The app designed by Amazon has all the data previously loaded in it about all the celebrities present in the movie and it recognizes from that set for identifying actors in the scene where the user tapped. Our project is not limited in the aspect of identifying celebrities in those videos for which data is initially present. It can identify actors in any video using the power of semi supervised learning. Our software or recognition method would be useful for the search engines also, as it will help them to provide people with their desired videos.

Results

The hybrid algorithm of MAD and LDS was compared with these algorithms it had better average accuracy than both the algorithm taken alone. The algorithm was tested on two datasets Cropped Yale and IMFDB.

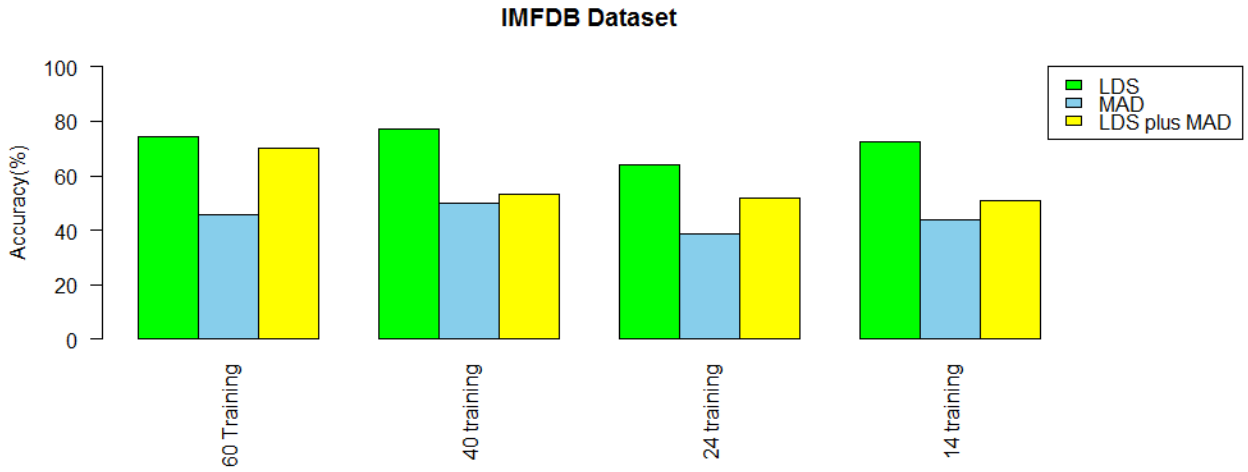
Cropped Yale

The extended Yale Face Database contains 16128 images of 28 human subjects under 9 poses and 64 illumination conditions.



IMFDB

Indian Movie Face database (IMFDB) is a large unconstrained face database consisting of 34512 images of 100 Indian actors collected from more than 100 videos. All the images are manually selected and cropped from the video frames resulting in a high degree of variability in terms of scale, pose, expression, illumination, age, resolution, occlusion, and makeup.



INTERPRETATION

The results shown above clearly depicts that in some of the data sets LDS + MAD has a better accuracy than LDS and MAD indivisually while there are some data sets for which it lies in the intermediate. As per the testing done till now we could speculate that LDS + MAD can prove to be a better choice than any of them indivisually and could handle the data more proficiently in case we have no previous knowledge of our data.In order to completely determine where our proposed algorithm stand among LDS and MAD we need to run it on the same data sets on which they are tested. This is a part of our research , while on the other hand the application part working fine along with the GUI.

APPLICATION

The application mentioned in the aim has been built and is working. The major parts to focus on the application part are the following:

1. Generated some data by collecting each frame from a playing video, this provides us with a sufficient UNLABLED data. We get around 200 frames per minutes. This video editing and manipulation part was done using openCV library in python.

2. We gathered some of the LABELED data from IMFDB and even generated some of ours and labeled it ourselves.

3. Most confident images were selected using the scikit library of python.

4. These data were later fed into the LDS algorithm to get the weights.

5. While these weights were in turn fed to the modified adsorption which classified our images.

Note: Please let us know if you want to see the program running or if we could just send you the screen shots.