

Gender Classification using a Spider Web Method

Vincenzo Russo¹

¹*Paola Barra, Carmen Bisogni*

Nowadays face classification is a serious problem, in this paper we present an approach based on a feature vector classification extracted using Web-Shaped Model Algorithm [1]. We classify these vectors using a Linear SVM and 2 datasets CelebA, UTKFace and a combined version of the two. In addition we stratify the dataset making it of even number of males and females. There are different approaches on Gender Recognition based on images but we decided to use feature vectors due to the convenience and privacy that they can offer.

I. INTRODUCTION

In this study we intend to classify images of individuals from CelebA and UTKFace datasets, which will be described in detail in the following section.

The purpose of this article is to show the differences between using a stratified sample dataset versus a non stratified dataset for gender classification.

The feature vector was given by the Web-Shaped Model Algorithm [1].

II. THE DATASETS

CelebFaces Attributes Dataset (CelebA) is a large-scale face attributes dataset with more than 200K celebrity images, each with 40 attribute annotations. The images in this dataset cover large pose variations and background clutter. CelebA has large diversities, large quantities, and rich annotations, including:

- 10,177 number of identities,
- 202,599 number of face images, and
- 5 landmark locations, 40 binary attributes annotations per image.

The dataset can be employed as the training and test sets for the following computer vision tasks: face attribute recognition, face detection, landmark (or facial part) localization, and face editing & synthesis.

UTKFace dataset is a large-scale face dataset with long age span (range from 0 to 116 years old). The dataset consists of over 20,000 face images with annotations of age, gender, and ethnicity. The images cover large variation in pose, facial expression, illumination, occlusion, resolution, etc. This dataset could be used on a variety of tasks, e.g., face detection, age estimation, age progression/regression, landmark localization, etc. Some sample images are shown as following:

- consists of 20k+ face images in the wild (only single face in one image)
- provides the correspondingly aligned and cropped faces
- images are labelled by age, gender, and ethnicity

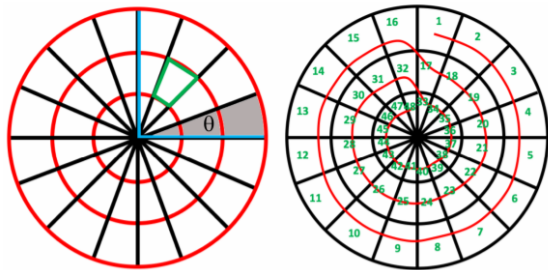


Fig. 3. Left: the web-shaped model applied on the face and centered on the landmark P₃₃; the radius of the web may change according to the farthest landmark from the center; red lines identify the *circles*, blue lines a *quarter*, the gray region is a *slice* of width θ , and green contour identifies a *sector*. Right: the order in which the sectors are analysed to build the pose feature vector, starting from the outer circle of the model.



Fig. 4. An application of the cascade: the first model identifies the face landmarks; the elements of the second one are circles, slices, and their sectors where the procedure places the landmarks to build a feature vector.

III. LANDMARKS EXTRACTION

After merging the two datasets we used the Web-Shaped Model for Head Pose Estimation.[1]

To build a pose feature vector, each landmark needs to be associated with a specific sector of the model. This process is carried out by first detecting the circle, the quarter and then the slice to which a specific landmark belongs to.

For this study we used 4 circles 4 sectors. Giving as output a 64 dimensions vector.

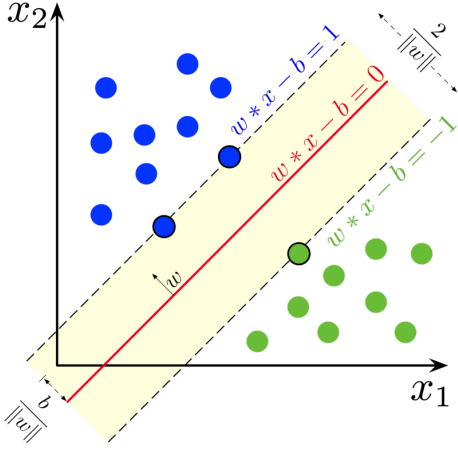


FIG. 1. SVM Example

IV. MODEL USED

Despite the high training cost we decided to use Linear Support Vector Machines because of the simplicity and the good classification generalization for binary problems. Here follows the definition of a Linear SVM:

We are given a training dataset of n points of the form $(\vec{x}_1, y_1), \dots, (\vec{x}_n, y_n)$, where the y_i are either 1 or -1, each indicating the class to which the point \vec{x}_i belongs. Each \vec{x}_i is a p -dimensional real vector. We want to find the "maximum-margin hyperplane" that divides the group of points \vec{x}_i for which $y_i = 1$ from the group of points for which $y_i = -1$, which is defined so that the distance between the hyperplane and the nearest point \vec{x}_i from either group is maximized.

Any hyperplane can be written as the set of points \vec{x} satisfying

$$\vec{w} \cdot \vec{x} - b = 0,$$

where \vec{w} is the (not necessarily normalized) normal vector to the hyperplane. This is much like Hesse normal form, except that \vec{w} is not necessarily a unit vector. The parameter $\frac{b}{\|\vec{w}\|}$ determines the offset of the hyperplane from the origin along the normal vector \vec{w} .

V. FEEDING PROCESS

This process has been repeated for each classification. A matrix of n rows and 65 columns was created from the numpy arrays given by the Web-Shaped Model Algorithm [1].

- The first column which contains the label 0 for female, 1 for male.

- The remaining columns contain the feature vector of 64 elements extracted from the face by the Algorithm [1].

VI. DATASETS STRATIFICATION

In statistical surveys, when subpopulations within an overall population vary, it could be advantageous to sample each subpopulation (stratum) independently. Stratification is the process of dividing members of the population into homogeneous subgroups before sampling. The strata should define a partition of the population. That is, it should be collectively exhaustive and mutually exclusive: every element in the population must be assigned to one and only one stratum. Then simple random sampling or systematic sampling is applied within each stratum. The objective is to improve the precision of the sample by reducing sampling error. It can produce a weighted mean that has less variability than the arithmetic mean of a simple random sample of the population.

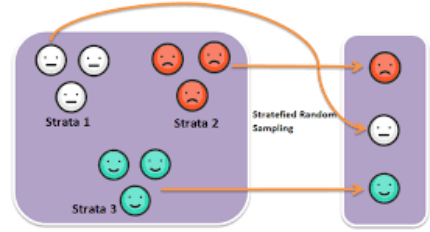


FIG. 2. Stratified Sampling Example

We found the following surplus of individuals by a given gender:

	Surplus	Gender
CelebA	30672	F
UTKFace	901	M
Combined	30878	F

The surplus has been removed from the dataset using random extraction without replacement.

VII. EXPERIMENTS

The experiments took place on a Windows laptop with intel Core i7-8550U, 8 GB DDR4 RAM, 512 GB SSD. Two main experiments were performed:

1. The study of the datasets without stratified sampling
2. The study of the same datasets without stratified sampling.

The datasets used for the experiments are:

1. CelebA
2. UTKFace
3. The combination of both

The estimate training times for the linear SVM were with stratified sampling:

	Time (hours)
CelebA	4,5
UTKFace	0,15
Combined	6

The estimate training times for the linear SVM were without stratified sampling:

	Time (hours)
CelebA	6,5
UTKFace	0,45
Combined	8

VIII. PERFORMANCE METRICS USED

In statistics, mean absolute error (MAE) is a measure of errors between paired observations expressing the same phenomenon. Examples of Y versus X include comparisons of predicted versus observed, subsequent time versus initial time, and one technique of measurement versus an alternative technique of measurement. MAE is calculated as:

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} = \frac{\sum_{i=1}^n |e_i|}{n} ..$$

It is thus an arithmetic average of the absolute errors $|e_i| = |y_i - x_i|$, where y_i is the prediction and x_i the true value. Note that alternative formulations may include relative frequencies as weight factors. The mean absolute error uses the same scale as the data being measured. This is known as a scale-dependent accuracy measure and therefore cannot be used to make comparisons between series using different scales.[2] The mean absolute error is a common measure of forecast error in time series analysis,[3] sometimes used in confusion with the more standard definition of mean absolute deviation.

The same confusion exists more generally.

MAE is not sensitive towards outliers and given several examples with the same input feature values, the optimal prediction will be their median target value.

IX. RESULTS

In this study we intend to classify images of individuals from CelebA and UTKFace datasets. We achieved the following MAE scores:

	MAE (Stratified)
CelebA	0,29
UTKFace	0,31
Combined	0,30

	MAE (Non Stratified)
CelebA	0,28
UTKFace	0,33
Combined	0,30

As further developements we could:

- Normalize the vector data in the range $[0,1]$.
- Choose a better classifier for binary classification task using sparse low dimensional vectors.
- Tweak the linear SVM hyperparameters.
- Using a non-linear kernel for the SVM.

ACKNOWLEDGMENTS

We thank Ph.D. Paola Barra and Carmen Bisogni for giving me the opportunity of working on an interesting problem of gender classification. We thank the BIPLab and Professor Michele Nappi for teaching us the fundamentals of biometrics. We thank again Paola Barra, Silvio Barra, Carmen Bisogni, Maria De Marsico and Michele Nappi for the chance of using this advanced algorithm for landmarks extraction.

-
- [1] P. Barra, S. Barra, C. Bisogni, M. D. Marsico, and M. Nappi, Web-Shaped Model for Head Pose Estimation: An Approach for Best Exemplar Selection (2020).