Face recognition of historical images by facial features markers, and feature dimensionality reduction

Junwei Xiong[1]

[1] Research School of Computer Science, Australian National University
U6757090@anu.edu.au

Abstract. Predictive classification model from high-dimensional, sparse data poses significant challenge on neural network architecture and accuracy. Identifying people in historical images is a typical task presents the above challenges, sparse images tend to make model computationally intensive and inaccurate. To mitigate the inaccuracy and complexity caused by high-dimensional feature, this paper starts with genetic algorithm to do the feature selection, it reduced the input space to a manageable size. Auto-associative network is also used here, it compressed the input data to the same dimensionality with genetic algorithm's. This paper looks at how spatial features in historical image be selected when applying genetic algorithm and compares the classification result between these two strategies and baseline. The dataset is obtained from Dr Caldwell, it has 36 samples, each sample contains 182 distance values of 14 facial features in two images. It is found that the selected features are far from each other at most time. These auto-associative network achieved 71.0625% while genetic algorithm achieved 78.25% accuracy, which is slightly higher than 75.875% accuracy of baseline with full features.

Keywords: Neural Network, Genetic algorithm, Auto-associative network, Data preprocessing

1 Introduction

There is a saying that "a picture is worth a thousand words." A photograph usually reveals more information compared to the other data formats. Convolution neural network get a huge success in the image recognition field. But there are some drawbacks of CNN, like it requires a large number of training set and it could be slow for the images with high resolution (Romain et al. 2020).

A more light-weight and human interpretable algorithm is preferable. In the forward propagation of DNN, the different levels of features are transmitted to the latter layers. Instead of extracting the features in the DNN like a black box, a standardized terms and methods are proposed. It uses facial markers like exocanthion, nasion, menton, etc to label the human face (Caple and Stephan. 2016). It reduced the image representation dimensionality to 14 dimensions; each dimension represents a facial markers coordinate.

Recognizing faces in two images based on their facial feature marker coordinate is not enough due to the different face poses, image resolution difference and distortions of 3D to 2D photographic compression. Dr Sabrina Caldwell proposed 3 approaches to tackle this problem, the first is calculating the distance between each points, the second is measuring the angle relating to feature points, the last is using cos similarity (Sabrina Caldwell, 2021).

The previous work validates using the distance of feature points as neural network input improves the accuracy, but the limitation of using distance is also discovered. The input data dimensionality grows exponentially with the increase of feature number. With 14 facial feature markers, there are 14C2 = 91 distances are derived, furthermore, 91C2 = 4095 angles are derived.

This paper investigates the use of genetic algorithm and auto-associative network (M.A.Kramer, 1992) to optimize the input data dimensionality, using a subset of features after genetic algorithm feature selection, to predict the target binary variable by the distance derived from the coordinate data subset. The paper additionally used auto-associative network to compress the features to the same dimensionality with genetic algorithm's, and then compared their prediction performance.

2 Data

2.1 data description

The original dataset "FFMs" used in this paper is collected by Dr. Caldwell (2020). The dataset contains 36 rows of data with each row recording the coordinates of 14 facial feature markers (figure 1) in two photographs. Since each coordinate consists of two values, there are 2*14*2 = 56 features in each row. The last column in the dataset is the target variable represented by 1 and 0, 1 means the faces in two photographs are belonging to one person, and vice versa. Therefore, the dimensionality for each data sample is 57.

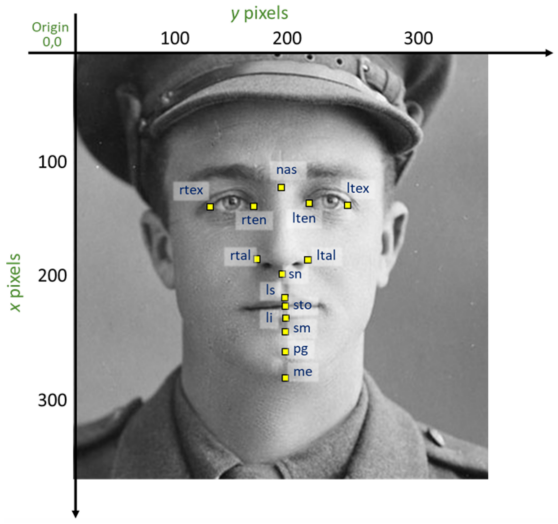

Fig. 1. 14 facial feature markers in an image are represented by (x, y) coordinates, origin is in the top left

2.2 data preprocess: calculating the distance

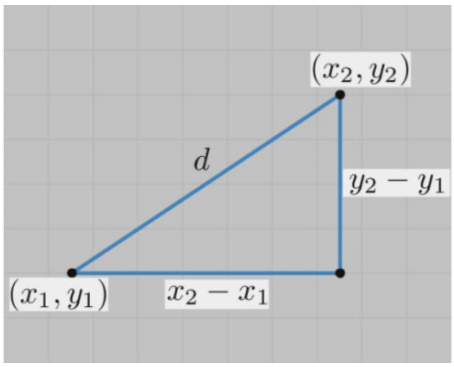

Fig. 2. Distance calculation between two markers example

To calculate the distance between each facial feature markers, all the possible combinations of two markers are computed at first. For each pair of markers, Euclidean distance is used here. The figure 2 is used to illustrate the process of distance calculation. Suppose there are two markers with coordinate (x1, y1) and (x2, y2), based on the Pythagorean theorem, the distance d can be calculated by formula: d = $\sqrt{(x2-x1)^2+(y2-y1)^2}$
This step provides the baseline dataset with full features for the further feature selection and feature compression.

3 Method

3.1 baseline neural network structure

Following Sabrina Caldwell (2021), a two-layer neural network classifier is constructed. The network contains 40 hidden neurons and 1 output neuron. The input neural number is varied depend on the input data dimensionality. For the baseline here, it is 182 as the feature set contains 91 distances for each of two photographs. 30 epochs is chosen to train the network, with a learning rate 0.001.
Each layer is fully linearly connected to the subsequent layer with no lateral, backward or multilayer connections. The input data will be normalized in the first layer. And the relu activation function is used to the hidden layer output. Sigmoid function is used to the output neuron. Adam is chosen as the optimizer.
Since the dataset is small, 8-fold validation is used to mitigate the effect caused by small dataset and measure the performance of neural network.
Note that other hyperparameters are also tested, the neural network with above hyperparameter setting gives the best performance. The 8-fold validation is made 10 times to improve the reliability and reduce the randomness, the average accuracy is 75.875%.

3.2 feature selection by using genetic algorithm

For the genetic algorithm, the feature selection is not used to distance dataset directly. A more interpretable and reasonable way is performing feature selection in the FFMs dataset, then calculate the distances between selected FFMs to get a new distance dataset.
The binary string encoding is used for chromosome, with a length of 14, each bit represents a facial feature. 1 means the facial feature is selected and 0 means not. 30 individuals are initialized randomly at first. Each individual has an own distance dataset according to its chromosome decoding. Chromosome with less than 2 bits of 1 will not happen in our case since the distance calculation requires a minimum of 2 feature points.
The next is to evaluate the chromosomes, the fitness function selection is very important as it determines what kind of individual we want in the last generation. Since the accuracy is not the only one factor we considered, dimensionality reduction is also important and will be evaluated in the fitness function. The aim of this task is to select features as less as possible and remain a good accuracy at the same time. The ratio of selected feature is bringing to the fitness function. After many experiments, the fitness function with formula 1 performs most reasonably.

$$\text{Fitness} = (\text{accuracy} + (1-\text{feature\_ratio})^{0.1}) / 2 \qquad (\mathbf{1})$$

The fitness function can be divided into two parts: classification accuracy and feature ratio score. The accuracy is calculated from 8-fold validation. And feature ratio score is determined by how many features are selected. For example, 10 out of 14, 71% features are selected, then the score is $(1 - 0.86)^{0.1} = 0.88$. The score is 0.94 when there are 6 facial features are selected. The exponent 0.1 in the function is to squeeze the selected feature size in a particular range, we can see this effect when taking 13 features are selected as example, the score of it is 0.77 which is much lower than 10 features are selected. The second effect of the exponent is to avoid the fitness function is biased towards feature ratio score. The dataset has an unbalanced class with 12 1s and 24 0s. At the worst case, the neural network can still get a 67% accuracy without any feature. If the feature ratio score keeps increasing rapidly when ratio is already small, that case could happen.
In each generation, the chromosome has a 90% probability to crossover and 10% probability to mutate. Tournament selection is used to select individuals from population with a tournament size 3. One point crossover is used to generate the offspring since there could be some interdependency between each facial feature marker.
The stopping condition is when it reached maximum 10 generations. The statistic result of generation flow is shown in the figure 3.

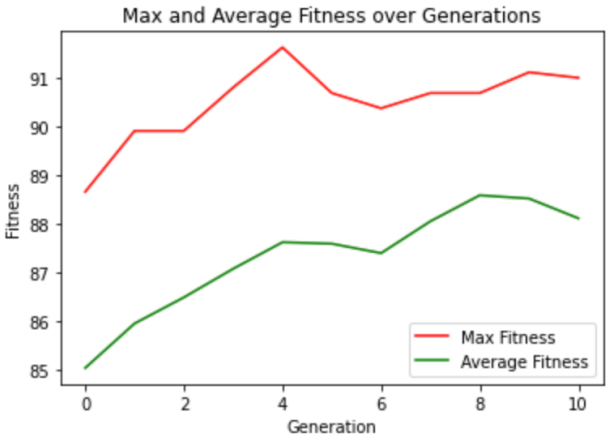

Fig. 3. Max and average fitness over generations

Due to the randomness of neural network, the max fitness in each generation is always higher than the average fitness even thorough the chromosomes converged at last. The genetic algorithm is conducted 12 times. The statistic result of each facial feature marker is in table 1.

Table 1. Total selected facial feature markers count in 12 genetic algorithm rounds

rtex	rten	ltex	lten	rtal	ltal	nas	sn	ls	sto	li	sm	pg	me
2	8	2	1	2	3	7	2	9	3	0	3	5	10

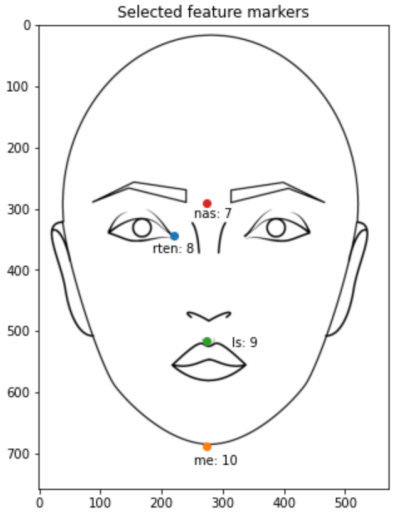

Fig. 4. Top 4 selected feature markers in 12 genetic algorithm rounds

From the table 1 and figure 4, top 4 selected feature markers are right endocanthion (8), nasion (7), labiale superius (9) and menton (10) respectively. The first found is the marker are spatially far away each other, no consecutive markers in the image. And 3 out of 4 markers belong to median feature while the rest one belongs to bilateral feature.
Based on the 4 markers, 6 distances are derived. Then the neural network uses these 6 distances to do the face recognition, after 10 iterations of 8-fold validation, the average accuracy is 78.25%.

3.3 feature compression by using auto-associative network

The auto-associative neural network is applied to distance dataset directly. It's a two-layer structure with one hidden layer. Each layer is fully connected without activation function. The network is trained 300 epochs, with learning rate of 0.001 and Adam optimizer.
Since there are 182 distances in each data sample, the input and output neurons are both 182. For performance comparison, the hidden neuron is set to 6 which is the same dimensionality with genetic algorithm's.

After training, the mean square loss is 665.68, the network is used to encode the distance dataset to 6 dimensionality. With this new compressed feature dataset, same prediction testing condition is applied here. The average accuracy of 8-fold cross validations is 71.0625% which is slightly better than chance. It's lower than the baseline accuracy 75.875%. The result is not optimal but is fairly reasonable, compressing a large dimension data to such a low dimensionality causes a large information loss. We can see this by observing the mean square error, square loss of 665 is very large especially for some distances between two close facial feature markers. Due to some noise in the data like marker measuring error, it makes the data compression much more difficult.

4 Conclusion and future work

This paper presents a genetic algorithm way to do the feature selection, therefore reduce the data dimensionality. This genetic algorithm is performed 12 rounds. Approximately 4 feature markers are selected in each round. The top 4 feature markers account for 59% total selected feature number. Based on the 4 feature makers' spatial distribution, it is found that 4 feature markers are far away from each other. Since some measuring error could happen in the feature marker labelling, select features distributed in a large region can mitigate the error rate and then give a more precise result. Furthermore, 6 distances are derived from these 4 markers. The neural network uses these 6 distances to predict the target variable and achieved a 78.25% accuracy which is surprisingly higher than using full distances. It implies using more features is not always good in classification, some similar or complement features could affect the network feature learning.

Using auto-associative network achieved a 71.0625% accuracy, it's worse than the baseline and slightly better than chance. Without any feature, network can still achieve 67% accuracy since the target variable 0 is two times than 1. The poor performance is caused by the information loss during the feature compression. The large mean square error explains it and supports this conclusion.

Although the genetic algorithm performs better and gives us some implications, a significant drawback of it is its very time consuming. If this algorithm applies to distance dataset directly, it becomes impracticable. Therefore, the future work can investigate some intelligent methods that allow the feature selection on a large feature set directly.

References

1. Arthur, L.G, Simon, M, Romain, T: Data Augmentation for Time Series Classification using Convolutional Neural Networks. ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data, (2016)
2. Caldwell, S. (2021) "Human interpretability of AI-mediated comparisons of sparse natural person photos," CSTR-2021-1, School of Computing Technical Report, Australian National University.
3. Caple, J., & Stephan, C. N. (2016). A standardized nomenclature for craniofacial and facial anthropometry. International Journal of Legal Medicine, 130(3), 863–879.
4. M. A. Kramer (2001), Autoassociative neural networks. Computers & Chemical Engineering, Volume 16, Issue 4
5. Fernandes BJT, Cavalcanti GDC, Ren TI (2014) Constructive Autoassociative Neural Network for Facial Recognition. PLoS ONE 9(12): e115967.