The Optimization of Pruned Lightweight Neural Network for Sparse Natural Person Photo Matching Based on Genetic Algorithm

Zhongxuan Liu

Research School of Computer Science,
Australian National University
u6730788@anu.edu.au

Abstract: In the face matching scene, if the same natural person has enough photos, artificial intelligence training is a good solution, the accuracy can be very high. However, neural networks are rarely used in sparse individual photo matching scenes. In this paper, in order to explore the matching effect of neural network in the case of sparse individual photos and the optimization effect of genetic algorithm on lightweight network, a neural network with two hidden layers pruned by distinctiveness is used to test the scene. Because there are too few data in the data set, the effect of the model is not good. In order to avoid the fluctuation of the model result, 10-folds cross validation was used. After 10-fold cross validation, the average accuracy was about 67%. In order to make the model better match the sparse natural person photos, genetic algorithm is used to optimize the neural network. After optimization by genetic algorithm, the accuracy is improved to 77.8%, and precision, recall and F1 increased from 0 to 2/3. The results show that the genetic algorithm has obvious optimization effect on the neural network with distinctiveness pruning.

Keywords: Face Matching, Neural Network, Genetic Algorithm, Pruning, K-fold cross validation.

1 Introduction

Photos are helpful for memory, are the media of social relations, and are also an important channel for self-expression and self-presentation[1]. As the subject of the photo, the people in the photo contain a lot of historical significance and stories. But the historical photos are so old that the people in them are often orphans. When an unmarked photo appears, no one can easily identify the person in the photo. Artificial intelligence may help people solve this problem. The most common method is neural network. Through literature review, it is found that the current research direction of face recognition mostly focuses on how to improve the accuracy of face recognition under extreme conditions. Extreme conditions often refer to large data sets or low quality data. But there is little research on sparse and low quality face recognition. Face recognition of sparse historical photos is the practical application of this scene. Therefore, the research on this topic is of practical significance. However, because the photos are sparse and the number of photos of the same person is small, although the accuracy is not too bad, the actual classification effect is very poor, it is difficult to classify the features into positive class. At this time, the optimization of the model is particularly important, which can improve the performance and accuracy of the model. In this paper, genetic algorithm is used to optimize the neural network, and the results before and after optimization are compared, the accuracy is improved by 11%, and the actual classification effect has been significantly improved.

1.1 Neural Network

In 1943, psychologist Warren McCulloch and mathematical logician Walter Pitts put forward and gave the concept of artificial neural network and the mathematical model of artificial neuron in the paper “a logical calculation of the ideas individual in nervous activity”, which initiated the era of artificial neural network research. Furthermore, Frank Rosenblatt, an American neuroscientist, proposed a machine that can simulate human perception, and called it "perceptron". In 1985, Geoffrey Hinton used multiple hidden layers to replace the original single feature layer in the perceptron. In 2006, Hinton et al. proposed the concept of deep learning[2]. In this paper, the neural network with two hidden layers is used to recognize the face in the photos and complete the matching between the photos.

1.2 Face Recognition

As early as 1965, Chan and Bledsoe published the earliest report on automatic face recognition[3]. In recent years, face recognition research has been favored by many researchers, and many technical methods have emerged. Almost all famous universities of science and technology and IT companies have research groups engaged in related research. Modern face recognition technology can easily recognize the face of modern people, because of the popularity of mobile camera devices, a person can have hundreds of photos, the data set is large enough, coupled with 3D features, the recognition accuracy is very high. However, for historical photos, due to the lack of popular photo taking equipment, the number of photos of a person is limited. Artificial intelligence has difficulties in determining the identity of the people in these photos.

The data set used in the experiment is from a subset of the National Archives of Australia. It's the feature point coordinates of a group of photos. There are 36 photos in total, which are divided into 12 groups with 3 in each group. The first two photos of each group are labeled as matching.

1.3 Pruning

Neural network pruning is a compression method, which involves removing weights from training model. Pruning has many advantages, such as improving generalization performance or being used as a precursor for rule extraction[4]. Pruning is often used in complex models, which can compress the model without excessively affecting the effect of the model. However, there are few pruning methods applied to lightweight neural networks. The Distinctiveness Pruning is used to prune lightweight neural network in this paper.

1.4 Genetic Algorithm

Genetic algorithm (GA) was first proposed by John Holland of the United States in the 1970s. The algorithm is designed and proposed according to the evolution law of organisms in nature. Genetic algorithm is a computational model to search the optimal solution by simulating the natural selection and genetic mechanism of biological evolution. The algorithm transforms the process of solving the problem into a process similar to the crossover and mutation of chromosome genes in biological evolution by means of mathematics and computer simulation. When solving complex combinatorial optimization problems, compared with some conventional optimization algorithms, they usually can get better optimization results quickly[5][8]. In order to solve the problem of neural network model instability in the case of sparse data sample, genetic algorithm is used to optimize the BP neural network in this paper.

2 Methodology

The experiment uses neural network to classify the sparse matching results of historical photos, and Distinctiveness Pruning is used to prune neural network. To improve model classification effect, genetic algorithm is used to optimize the model parameters. In this paper, the experimental method is described from four aspects: dataset, neural network model structure and training, pruning and genetic algorithm.

2.1 The Dataset

Data Description. The data used in the experiment comes from Caldwell's research on sparse historical photos in 2021[6]. The data consists of 36 two-dimensional natural person photos. 36 photos were divided into 12 groups with 3 in each group. Three photos of each group were labeled A, B and C respectively. The natural person in A and B is the same person. The original data is 36 photos, but because this model is back propagation neural network, it does not use pictures as data source. The coordinates of feature points and the distance between feature points of 36 photos are made into two tables. The coordinate table of feature points has 57 rows and 36 columns. Each image has 14 coordinate points, each point is divided into abscissa and ordinate, a total of 28 data(Table 1). Each row of data represents the pairing of each group of photos, so each group has three rows, a total of 36 rows. The coordinate data of two pictures in each row, 28 in each row, a total of 56 rows. The last line is the label of the matching result. So it's 57 lines. The feature point distance table has 183 columns and 36 rows. The reason for 36 lines is the same as above. There are 183 columns in total, because each image has 14 feature points, pairing each other to form 91 feature point distances, and there are two pictures in each row, so there are 182 rows in total. The last line is also the label of the matching result.

Table 1. Description and Statistics of FFMs[6].

Facial Marker	Abbreviation	Bilateral/Median	Maximum		Minimum	
			x	y	x	y
Right exocanthion	rtex	Bilateral	302	552	30	28
Right endocanthion	rten	Bilateral	403	569	44	26
Left exocanthion	ltex	Bilateral	673	585	5	23
Left endocanthion	lten	Bilateral	541	578	57	24
Nasion	nas	Median	456	536	51	21

Sub-nasale	sn	Median	414	753	51	47
Right alare	rtal	Bilateral	357	702	42	43
Left alare	ltal	Bilateral	516	720	62	43
Labiale superius	ls	Median	420	823	52	58
Stomion	sto	Median	416	845	52	61
Labiale inferius	li	Median	408	888	15	63
Pogonion	pg	Median	391	941	53	77
Supramentale	sm	Median	409	911	53	70
Menton	me	Median	399	1025	54	83

Both tables can be used as neural network data sets, but in order to ensure the training effect, we need to calculate the cosine similarity of these features. The cosine similarity between features indicates whether the two features can well reflect the real matching results of two images. Since it is known that A and B are the same person, the higher the cosine similarity between the features of A and B, the easier it is for the neural network to conclude that A and B are the same person.

Table 2. Cosine similarity measures for Distances, Proportions, FFMs[6].

	Distances			Proportions			FFMs		
	A&B	B&C	A&C	A&B	B&C	A&C	A&B	B&C	A&C
Group 1	0.003	0.003	0.001	0.015	0.031	0.036	0.029	0.035	0.001
Group 2	0.004	0.002	0.026	0.026	0.031	0.063	0.058	0.009	0.030
Group 3	0.001	0.003	0.002	0.006	0.013	0.008	0.001	0.030	0.020
Group 4	0.002	0.001	0.002	0.015	0.014	0.039	0.015	0.001	0.019
Group 5	0.017	0.004	0.017	0.607	0.035	0.615	0.006	0.003	0.004
Group 6	0.002	0.002	0.004	0.062	0.010	0.049	0.004	0.007	0.006
Group 7	0.000	0.003	0.002	0.003	0.043	0.035	0.004	0.001	0.008
Group 8	0.073	0.070	0.007	0.256	0.169	0.114	0.031	0.005	0.027
Group 9	0.018	0.019	0.001	0.150	0.201	0.021	0.038	0.044	0.001
Group 10	0.001	0.001	0.001	0.005	0.006	0.005	0.002	0.006	0.007
Group 11	0.000	0.004	0.004	0.005	0.055	0.035	0.000	0.009	0.016
Group 12	0.001	0.006	0.004	0.003	0.026	0.019	0.000	0.045	0.048

Which of the three can better recognize that A and B are the same face, then this value is selected as the data set of neural network. Therefore, a method is designed to evaluate the cosine similarity. Suppose we only consider the cosine similarity of A and B. The cosine similarity of A and B is ranked among A and B, B and C, A and C. When the similarity of A and B is the highest among the three, it is given the weight of 3. If it is the second highest, given the weight of 2. If is the lowest, given the weight of 1. Through this calculation method, the weight sum of distances is 21, the weight sum of Proportions is 19, and the weight sum of FFMs is 22. This shows that FFMs can better show that A and B are the same person, so the neural network selects FFMs as the dataset. After the data set is selected, the data set is segmented. Take columns 1-56 as input and column 57 "outcom" as label.

Data Preprocess. In order to ensure the convenience of subsequent data processing and speed up the convergence speed of the model, the normalization method is used to process the data. In this paper, Min-Max normalization method(Formula 1) is used to process the data.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

Min-Max normalization is a linear transformation of the original data, so that the result value is mapped between [0 - 1]. Where max(x) is the maximum value of the sample data and min(x) is the minimum value of the sample data. One defect of this method is that when new data is added, it may lead to the change of Max and Min, which needs to be redefined. However, there is no new data in this dataset, so this defect can be ignored.

2.2 Original Neural Network Structure and Training

The structure of neural network is back propagation network. It uses two hidden layers, including one input layer and one output layer. Too many hidden layers will increase the training time and over fitting, so the number of hidden layers is two. Since FFMs are selected as the data set, according to the above, there are 56 columns in FFMs. So there are 56 neurons in the input layer. By comparing the effects of different combinations of hidden neurons, according to the results of the validation set, there are 20 neurons and 10 neurons in the two hidden layers. The final output is binary, so the output layer is two neurons.

The selection of training parameters is the result of continuous attempts. Firstly, the activation function and optimizer are determined. The activation function selects the ReLU activation function(Formula 2). The reason for choosing this activation function is the efficient gradient descent and back propagation, which avoids the problem of gradient explosion or gradient disappearance. At the same time, the function does not involve complex calculation, which reduces the calculation cost of the whole network. The optimizer selects Adam. The reason for choosing Adam optimizer is that it is easier to adjust parameters. Because of many parameters, a good optimizer can reduce the workload of parameter adjustment. The training process involves three parameters, epoch, batch size and learning rate. By comparison, if the epoch club is too small, the training result will be poor. After comparing 50 to 200 epochs, we finally decide that the epochs is 80, too much epochs is meaningless and wasting runtime. Batch size is 12. Because of the selection of Adam optimizer, the learning rate is relatively low and better results will be achieved. The learning rate from 0.1 to 0.0001 was tested. Because large learning rate cause over fitting, the final learning rate was 0.0001.

$$f(x) = \max(0, x) \tag{2}$$

In this experiment, because the data set is too small, in order to reasonably evaluate the effect of the model and prevent the occurrence of over fitting, k-fold cross validation is used to evaluate the model. K-fold cross validation is a popular method to evaluate the performance of classification algorithm. A reliable accuracy estimate will have a relatively small variance. It is recommended to perform k-fold cross validation to evaluate the classification mode[7]. This paper uses 10 fold cross validation to verify the model. According to the rule of thumb, the ratio of training set and test set is 3:1.

2.3 Pruning

In order to explore the influence of genetic algorithm on the lightweight neural network after pruning, this paper chooses the distinctiveness pruning method to prune the above neural network structure.

The distinctiveness of the hidden unit is determined by the unit output activation vector on the pattern representation set. That is, for each hidden unit, we construct a vector that has the same dimension as the number of patterns in the training set, and each component of the vector corresponds to the output activation of the unit. In this model, the vectors of the clone units will be the same regardless of the relative size of their output, and can be recognized. Units in the pattern space with short activation vectors are considered unimportant and can be removed. The similarity of vector pairs is recognized by calculating the included angle between them in the pattern space[4]. Distinctiveness Pruning(DP) means that when the angle of two vectors is less than 15 degrees, they are considered similar, and need to remove one vector and add the weight of that vector to the other vector pair. If the angle between the two is greater than 165 degrees, the two are considered contradictory, delete both. The calculation of vector angle can be designed according to the properties of cosine function. The calculation formulas are (2), (3) and (4).

$$\cos\theta = \frac{vectorA \cdot vectorB}{|vectocA| \times |vectorB|} \tag{3}$$

$$\theta = \arccos(\cos\theta) \tag{4}$$

$$angle = \frac{\theta \times 180^{\circ}}{\pi} \tag{5}$$

2.4 Genetic Algorithm

In this paper, genetic algorithm is used to optimize BP neural network. Firstly, initialization is performed. If the maximum number of generations and initial population is too large, the operation time will be too long, And the CPU performance is limited, so the maximum number of generations is set to 10. The initial population is 10. According to the rule of thumb, the chromosome length is 16.

Selection. The purpose of selection is to directly inherit the optimized individuals to the next generation or to generate new individuals through pairing crossover and then to the next generation. The proportional selection method, also known as roulette wheel selection method, was used in the experiment. This selection method is easy to implement and understand[8]. In the proportional selection, the probability of choosing an individual for breeding of the next generation is proportional to its fitness, the better the fitness is, the higher chance for that individual to be chosen(formula 6).

$$p_i = \frac{f_i}{\sum_{j=1}^{N} f_j} \tag{6}$$

p_i is probability of choosing individual i. f is fitness. N is the total number of individuals in the population.

Crossover. It is the process in which two chromosomes (strings) combine their genetic material (bits) to produce a new offspring which possesses both their characteristics. Single point crossover is used in experiment and crossover rate is 0.9. Most of the chromosome pairs will be crossover. A random point is chosen on the individual chromosomes (strings) and the genetic material is exchanged at this point.

Table 3. Example of Single Point Crossover

Chromosome 1	110|111000
Chromosome 2	001|000111
Offspring 1	110|000111
Offspring 2	001|111000

Mutation. Mutation simulates the gene mutation caused by accidental factors in the natural genetic environment, which makes a gene on the chromosome change randomly with a small probability. The chromosome code in the experiment is binary, so there are only two possibilities of mutation: 0 -> 1 and 1 -> 0. The mutation probability in the experiment is 0.1, which can not only increase gene diversity, but also ensure that the genes of high-quality individuals can be retained as much as possible.

Table 4. Example of Mutation

Offspring	11011
Mutated Offspring	11010

In the experiment, genetic algorithm is used to optimize the parameters of BP neural network. It includes the number of hidden neurons, the number of epochs and the learning rate. The input neurons are 56 and output neurons are 2, so the range of hidden neurons is 1 to 100. If the number of epoch is too small, the effect of the model is not good, and if the number of epoch is too large, the effect is not improved, and CPU performance and running time will be wasted. So the range of epoch number is 10 to 100. And the range of learning rate is 0.0001 to 0.01.

3 Result and Discussion

In this section, the results and findings are presented. It includes the classification effect of the basic model after distinctiveness pruning and the classification effect after using genetic algorithm. Finally, the reasons for this result will be discussed.

3.1 Result of Neural Network after DP

The parameter selection of DP is pruning the first layer and removing the weight. The The angle threshold is 15 degrees to 165 degrees. In the 10 fold cross validation, the training accuracy is from 33.333% to 69.697%. The average training accuracy is 63.939%. As shown in Figure 1-1.

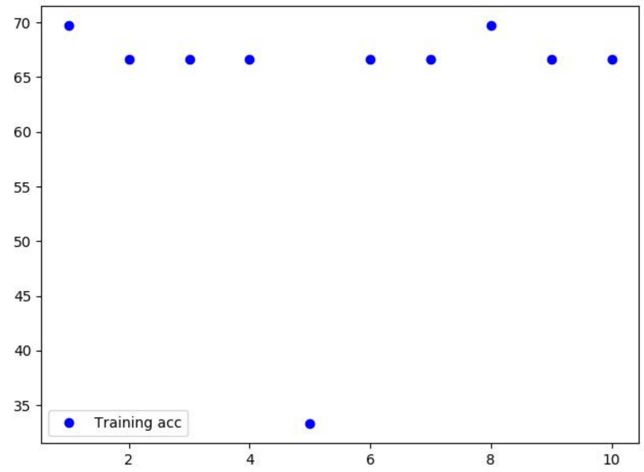

Fig.1-1. The training accuracy of each fold in 10 fold cross validation.

The validation accuracy is 66.667%. The average validation accuracy of 10-fold cross validation is 66.667%. As shown in Figure 1-2.

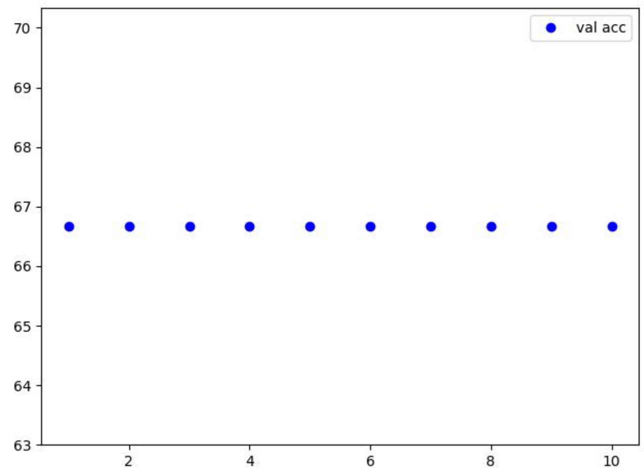

Fig.1-2. The validation accuracy of each fold in 10 fold cross validation.

The training loss values ranged from 0.5263 to 0.6937, and the average training loss of 10 fold cross validation is 0.6460. As shown in Figure 1-3.

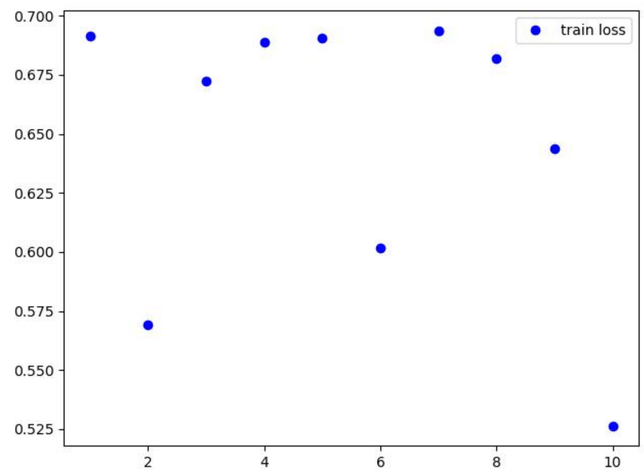

Fig.1-3. The training loss of each fold in 10 fold cross validation.

The validation loss values ranged from 0.6255 to 0.7096, and the average validation loss of 10 fold cross validation is 0.6591. As shown in Figure 1-4.

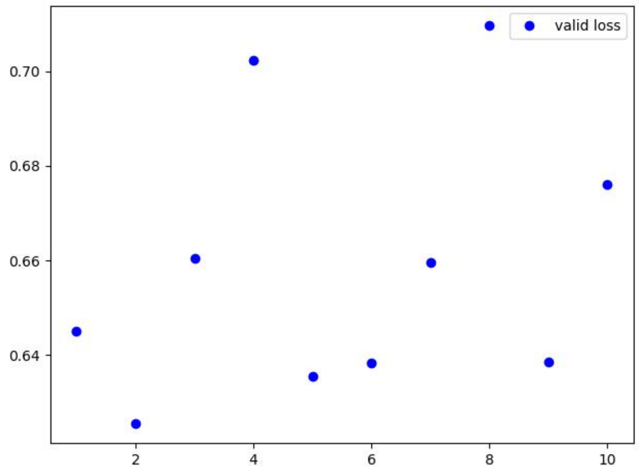

Fig.1-4. The validation loss of each fold in 10 fold cross validation.

Table 5. The Validation Confusion Matrix of Neural Network after DP

		Predict value	
		0	1
labels	0	6	0
	1	3	0

According to table 5, it can be concluded that TP = 0, FN = 3, FP = 0, TN = 6. So we can calculate that precision, recall and F1 are 0.

3.2 Result of Model with Genetic Algorithm

After the pruning BP neural network is optimized by genetic algorithm, the validation accuracy is 77.78%. After 10 generations of evolution, the fitness changes with generations, as shown in Figure 2-1.

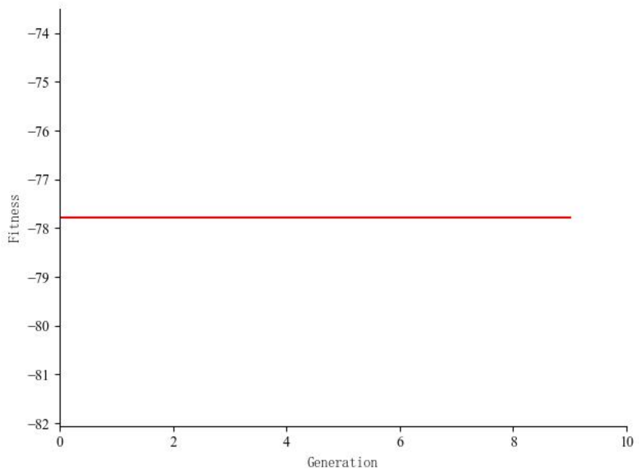

Fig.2-1. Fitness changes with generation

Table 6. The Validation Confusion Matrix of Model with genetic algorithm

		Predict value	
		0	1
labels	0	5	1
	1	1	2

According to Table 6, it can be concluded that TP = 2, FN = 1, FP = 1, TN = 5. So we can calculate that $precision = \frac{TP}{TP+FP} = \frac{2}{3}$, $recall = \frac{TP}{TP+FN} = \frac{2}{3}$ and $F_1 = \frac{2TP}{2TP+FP+FN} = \frac{2\times 2}{2\times 2+1+1} = \frac{2}{3}$.

3.3 Discussion

Due to the small amount of experimental data, the model is light. It leads to the poor results of the model after training. Under such extreme conditions, the overall effect of the neural network after DP pruning is significantly lower than the experimental results reported by Caldwell in 2021, and the validation accuracy is about 67%. Although DP only prunes the similar or contradictory weight vectors, it has a great impact on the lightweight neural network in this experiment, and the accuracy is reduced by 8%. In this experiment, there are only two hidden layers, and each layer has fewer neurons, so the structure is relatively simple. Due to the small number of neurons, although the probability of low weight connection is low, and each weight vector is relatively independent and has its own unique role. When any weight is pruned, the performance of the whole network will be affected.

On the face of it, 67% of the accuracy is not bad. But the accuracy often can not accurately describe whether a model is good or bad in the classification problem. It also needs to calculate precision, recall and F1. By calculation, we find that precision, recall and F1 are all 0, all predictions of the model are negative. Through the observation of the data set, we can find that the positive and negative samples in the data set are not balanced. Among them, negative samples account for two-thirds, and positive samples account for only one-third. There are only 36 data in the total sample, and 12 are positive samples. Therefore, it is difficult to classify features into positive classes. So we can conclude that DP is not suitable for lightweight network.

After the model is optimized by genetic algorithm, the accuracy of classification is improved from 66.7% to 77.8%. At the same time, after calculation, it can be found that precision, recall and F1 are significantly improved, they are all 2/3. This means that the classification ability of the model is greatly improved. Even if the positive and negative samples are not balanced, some features can be correctly classified as positive classes. Although the classification accuracy is not much improved, in the actual scene, the classification effect is far better than the model before using genetic algorithm optimization.

By observing Figure 2-1, we can find that fitness does not change with the generation. This shows that the genetic algorithm is not better because of evolution. The reason for this is that the optimal individual exists in the initial population. As a result, the fitness of offspring generated by subsequent population evolution is not the largest, so fitness will not change with generation. In essence, the problem of small data set affects the evolutionary process of genetic algorithm. Small data set and simple network structure will lead to the premature obtaining of the optimal individual. When the genetic algorithm finds the optimal solution in the initial population, the subsequent evolution process is useless. But even if the subsequent evolution process did not find a better individual, but also had a very good optimization effect.

4 Conclusion and Further Work

4.1 Conclusion

In this paper, neural network is used to solve the classification problem of sparse historical photo matching. Firstly, the most suitable data sets (FFMs) are selected by calculating the weight sum of the data sets. Then the classification task is completed by the neural network with one input layer, two hidden layers and one output layer after DP pruning. By adjusting the parameters, the validation accuracy of 66.7% can be achieved in extreme cases of very few data. However, due to the imbalance of positive and negative samples, precision, recall and F1 are 0, so the classification effect is bad. Compare with previous work, it shows that distinctiveness pruning has a great influence on the lightweight neural network. DP is not suitable for lightweight neural networks. Then, genetic algorithm is used to optimize the model. The experimental results show that the optimization effect of genetic algorithm is obvious, and the validation accuracy is improved by 11%. Precision, recall and F1 are improved from 0 to 2/3. This shows that the classification effect of the model optimized by genetic algorithm has been significantly improved, the new model can classify sparse historical photos well, which proves that genetic algorithm is suitable for optimizing this kind of back propagation neural network.

4.2 Further Work

There are many different ways to advance this work. Datasets, neural network structure and parameters, genetic algorithm can make further efforts to promote the progress of the whole work:

Datasets. For the dataset, we can expand the scale of the dataset and increase the training samples, which can improve the classification effect of the model. In the experimental data, the positive and negative samples are unbalanced, and the negative samples are twice as many as the positive samples. Therefore, the number of positive samples can be increased by over-sampling, and the total number of samples will increase accordingly. At the same time, in order to avoid overfitting caused by the same data, random noise or interference data should be added. Secondly, data mining can be used to explore the relationship between feature points. If we can find stronger relevance, we can also improve the classification effect of the model.

Neural Network Structure and Parameters. For neural network, the classification accuracy of the model can be improved by increasing the number of hidden layers. At the same time, we can adjust the training parameters through more tests to optimize the training effect. For example, in order to prevent over fitting, continue to reduce the learning rate. We can also use the image data set and change the network structure, and use CNN to train the image data for machine learning, so as to reduce the requirements of the previous data processing work.

Genetic Algorithm. For genetic algorithm, we can try more parameter combinations when the running time is not limited. For example, different initial population numbers. This may change the straight line of fitness. Different crossover and mutation methods or probabilities will make genes more diverse, and may produce the optimal individuals with higher fitness, so as to get rid of the local optimal solution and find the global optimal solution. Using high-performance CPU and GPU can complete more generations of evolution in a short time, so high-performance devices can be replaced to complete the genetic algorithm.

These areas are worth exploring. In the future, I will continue to improve this work. I believe that the neural network will achieve higher accuracy in the matching and classification of sparse historical photos, and genetic algorithm will have better optimization effect in the future.

References

1. Van House, N. A., Davis, M., Takhteyev, Y., Ames, M., & Finn, M. (2004). The social uses of personal photography: methods for projecting future imaging applications. *University of California, Berkeley, Working Papers*, 3, 2005.

2. Wu, Y. C., & Feng, J. W. (2018). Development and application of artificial neural network. *Wireless Personal Communications,* 102(2), 1645-1656
.

3. Petrescu, R. V. (2019). Face Recognition as a Biometric Application. *Journal of Mechatronics and Robotic*s, 3, 237-257.

4. Gedeon, T. D. . (1995). Indicators of hidden neuron functionality: the weight matrix versus neuron behaviour. *2nd New Zealand Two-Stream International Conference on Artificial Neural Networks and Expert Systems (ANNES '95), November 20-23, 1995, Dunedin, New Zealand.* IEEE.

5. Kumar, M., Husain, M., Upreti, N., & Gupta, D. (2010). Genetic algorithm: Review and application. *Available at SSRN* 3529843.

6. 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.

7. Wong, T. T., & Yeh, P. Y. (2019). Reliable accuracy estimates from k-fold cross validation. *IEEE Transactions on Knowledge and Data Engineering*, 32(8), 1586-1594.

8. Ding, S., Su, C., & Yu, J. (2011). An optimizing BP neural network algorithm based on genetic algorithm. *Artificial intelligence review*, 36(2), 153-162.

9. Pytorch - K fold cross validation process description and Implementation, https://blog.csdn.net/foneone/article/details/104445320