How Well Will the Auto Encoder/Decoder influence the performance of single network on feature classification field.

Yijie Liu,
Research School of Computer Science,
Australian National University,
Canberra Australia

u6890141@anu.edu.au

Abstract. Previously proved that applying threshold have positive effect on single structure network to enhance classification accuracy, however, the result was not as much as expected to be. Hence, in order to promote the learning ability of such network attached with neuron threshold, this paper will demonstrate further implementation of featurizing on the same network structure, where auto-encoder/decoder will be applied.

Keywords: Auto-encoder, classification, facial emotions.

1 Introduction

Facial emotions are in a way the most intuitive and obvious reflection of biological aspects. People convey feelings through their reactions of anger, anxiety, sadness, happiness, surprise, especially when they feel uncomfortable to express them in words. Facial emotions always represent the first reaction to whatever is happening around them, so it makes considerable sense to recognize, digitize and analyze facial movements as a pre-processing step for further AI techniques.

Our task is to analyze a set of movie screenshots including images of people with seven different facial emotions, with a total of 675 experimental data. each sample image is processed in the form of LPQ (Local Phase Quantization) and PHOG (Pyramid of Histogram of Oriented Gradients), and the top five features of the resulting feature data are taken as the features of the sample, respectively. the LPQ is first calculated through the local image window for each grid and then cascade the whole image, PHOG shows better results on low resolution images [1]. (T. Gedeon, 2011)

In order to better featurize the 10 different input sample feature values, here we introduce another un-supervised deep-learning method for pre-training, auto-encoder and decoder. The development of neural networks was facilitated by Rumelhart's introduction[2] of the concept of autoencoder in 1986 and its use for high-dimensional complex data processing. A self-encoding neural network is an unsupervised learning algorithm that uses a backpropagation algorithm and makes the target value equal to the input value. An autoencoder is a type of neural network that can be thought of as consisting of two parts: an encoder function $h = f(x)$ and a decoder that generates a reconstruction $r = g(h)$. Traditionally, autoencoders have been used for dimensionality reduction or feature learning.

Take the data amount into consideration, we generate a suitable cross-validation with k-fold parameter of 5 to our model network[3]. To avoid overfitting and enhance accuracy, another important technique that has been applied to the model is thresholding. Categorization has been shown that the results of trained neural networks do not score model performance[4] (Kogan, 1991), and Gedeon mentioned in another paper that thresholding has been shown to be a good evaluation criterion for related aspects (T.D. Gedeon, 2001).

2 Method

2.1 Data pre-processing

The dataset was generated from screenshots from 10 different movies, each containing several intuitive human emotions, which were then independently labeled by 2 annotators and then converted to numerical data with an accuracy of 6 decimal places.
Pre-processing is essential as the order of magnitude is too small to allow for accurate training results.

$$(x - x.mean())/x.std()$$

x refers to the functional variable in each column. In the dataset, NaN values in columns 6 to 11 need to be removed to avoid uncountable losses and constant accuracy, except for the title, index and movie name. Column 1 consists of the output targets of the network, which are also known as class labels. The class labels refer to 7 different emotions, including 6 basic expressions such as anger, disgust, fear, happiness, sadness, surprise, and neutral classes, respectively.

2.2 Auto encoder/Decoder

For pre-training stage, we apply a "symmetric" network to further characterize the features (introduced in part 1), show a prototype in figure 1. We config the hidden layers consist of 10 and 8 neurons, 4 for the bottleneck, which are linearly connected.
Input data will be fed into the network after standardizing, with respect to each input layer neurons as a tensor unit. Meanwhile, MSE function and Adam optimization was applied during iterating, the epoch number is 100.

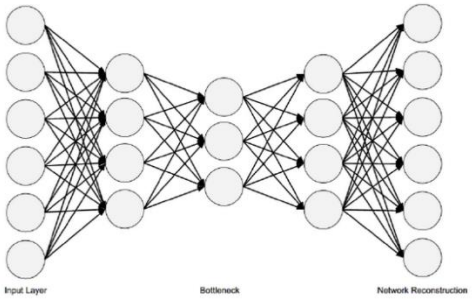

Fig. 1. Auto-encoder/decoder prototype.

We also used weight decay (L2 regularization) to avoid the model overfitting problem. The reasons are as follows:
(1) Explanation in terms of model complexity: smaller weights w, in a sense, means that the network is less complex and fits the data better (this law is also called Occam's razor), and this is verified in practical applications, where L2 regularization tends to work better than unregularized results. (2) The mathematical explanation: when overfitting, the coefficient of the fitted function is often very large, why? As shown in the figure below, overfitting, is to fit the function needs to be concerned about each point, the final formation of the fit function fluctuates greatly. In some very small intervals, the function value changes very dramatically. This means that the function has a very large derivative value (absolute value) in some small intervals, and since the value of the independent variable can be large or small, only the coefficients are large enough to ensure that the derivative value is large. The regularization, on the other hand, is done by constraining the parametrization of the parameters so that they are not too large, so the overfitting can be reduced to some extent.

2.3 Neural Network Construction

The sample network was constructed with one input layer, one hidden layer and one output layer, with parameters of neuron numbers of 10,7,1 respectively, Cross-entropy and SGD will be performed for loss computation and optimizer. Furthermore, a 5-fold cross-validation method stands for enhance learning ability at a finite data scale.

3 Result and Discussion

The result of combining auto-encoder and thresholding in sample network didn’t perform as what we respect, the loss of auto-encoder/decoder remains in a tremendously low level, however, after feeding the output data into sample network, the whole function of network was nearly crushed. We also output the accuracy curve of both situations, shown in Figure 2 and 3.

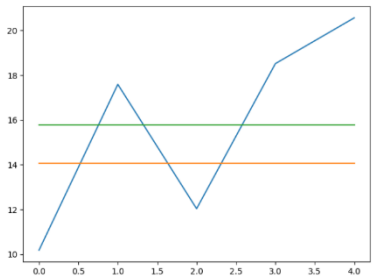

Fig. 2. Data with pre-training

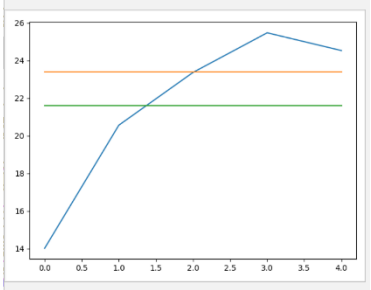

Fig. 3. Data directly fed into network

Green line: testing set accuracy. Orange line: mean accuracy value of cross-validation, y axis represents the accuracy with unit of percent, x axis is series of 5 folder indexes. In Figure 2, it shows a stable result of predicting, with mean accuracy of cross-validation value of 24.53% and testing accuracy of value of 23.40% overall. On the other hand, after pre-training with auto-encoder/decoder, the accuracy of cross-validation and testing process decreased to 15.89% and 14.01% overall.
Not only limited to the shown figures here, the model occurs with over-fitting and under-fitting alternatively even though we set the learning rate in 0.01 at the same time.
However, what do prove here is the conclusion that we concluded before while thresholding between 0.4 to 0.75 the network reached onto peak performance still works here. Therefore, we can further conclude that the model remains a few parts of learning ability, but is not suitable for combining with auto-encoder/decoder.

4 Conclusion and future work

Auto-encoder/decoder mainly contributes in improving images and texts features, which always collaborate with convolutional structure to construct, also it has ability to reduce the dimensionality of complex input [5] which inspired us for applying such network into our model. We initially expect the result to demonstrate a higher performance, however, the drawbacks of such combination appear to be:

1. Does the dimension (10) of the dataset really satisfy the basic input condition of auto-encoder?
2. Does our model have capability to handle such high -level featured data sufficiently?

Our initial network only constructs with 3 layers network, it’s learning ability and stability remains to be poor. Inspired by insufficient dataset, we do get the original image dataset contains 675 pieces of movie screen shots, if we start processing the original images by resizing them into square images

of acceptable size, then these images could be treated as input to be fed with CNN and auto-decoder/encoder. As mentioned, auto-encoder has high ability to do dimensionality reduction for pre-training, it provides with sufficient features to be computed. This scenario was not achieved yet due to low processing speed also limited RAM and memory on lap top.

Emotions are direct interactions of a person towards society, his/her reaction can be various by age, gender, ethnicity, health condition and so on. Meanwhile, for dataset labelling, labellers have possibility to fail in recognising all features when collecting training samples (e. g. low-resolution image), and such an incomplete training data set can result in the decrease in classification accuracy[6] (Li, 2014).

References

[1]A. Dhall, R. Goecke, S. Lucey and T. Gedeon, "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark," 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops), 2011, pp. 2106-2112, doi: 10.1109/ICCVW.2011.6130508.

[2]Rumelhart DE，Hinton GE，Williams RJ． Learning representations by back-propagating errors[J]． Nature，1986，323: 533-536．

[3] Y.Liu, “Influence factors of how neuron threshold affects the overall facial emotion multi-classification accuracy by using cross-validation”, 2021.

[4]Kogan, “Neural networks trained for classification can not be used for scoring,” IEEE Trans. on Neural Networks, 1991.

[5] Yasi Wang, Hongxun Yao, Sicheng Zhao, Auto-encoder based dimensionality reduction, Neurocomputing, Volume 184, 2016, Pages 232-242, ISSN 0925-2312, https://doi.org/10.1016/j.neucom.2015.08.104.

[6]W. Li and Q. Guo, "A New Accuracy Assessment Method for One-Class Remote Sensing Classification," in IEEE Transactions on Geoscience and Remote Sensing, vol. 52, no. 8, pp. 4621-4632, Aug. 2014, doi: 10.1109/TGRS.2013.2283082.