Journal of Theoretical and Applied Information Technology
20th May 2016. Vol.87. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

EVALUATING THE PERFORMANCE OF DEEP SUPERVISED AUTO ENCODER IN SINGLE SAMPLE FACE RECOGNITION PROBLEM USING KULLBACK-LEIBLER DIVERGENCE SPARSITY REGULARIZER

[1]OTNIEL Y. VIKTORISA, [2]ITO WASITO, [2]ARIDA F. SYAFIANDINI
Faculty of Computer of Computer Science, Universitas Indonesia, Kampus UI Depok, Indonesia
E-mail: [1]otniel.yosi@ui.ac.id , [2]ito.wasito@cs.ui.ac.id, [2]arida.ferti@ui.ac.id

ABSTRACT

Recent development on supervised auto encoder research gives promising solutions toward single sample face recognition problems. In this research, Kullback-Leibler Divergence (KLD) approach is proposed to obtain penalty of sparsity constraint for deep auto encoder learning process. This approach is tested using two datasets, Extended Yale B (cropped version) and LFWcrop. For comparison, Log and εL_1 also employed as sparsity regularizers. Experiment results confirm that KLD has better performance in image classification of both datasets compared to Log and εL_1.

Keywords: *Single Sample Face Recognition, Deep Auto Encoder, Kullback-Leibler Divergence, Sparsity, Sparsity Regularizer*

1. INTRODUCTION

As one of many topics in computer vision, face recognition has been an interesting topic and applied to solve some of real world problems. Face recognition is basically used for identifying face structure from images or videos [1]. Although a lot of solutions have been given for face recognition problems, some of them, like single sample face recognition remain as an open case. It is because face identification process in single sample face recognition uses only one face image, due to lack of information.

One of latest researches in single sample recognition [2] uses supervised auto encoder for face image labelling. Auto encoder is built from Artificial Neural Network (ANN) structure. For fine tuning process, it utilizes backpropagation process [3].

Auto encoder is usually used to reconstruct input data and minimize its reconstruction error. Research [2] measures many aspects in learning data with supervised auto encoder, such as its performance, activation function, number of layers, hidden units and weight parameters. Supervised auto encoder achieves a better performance compared to other known methods in terms of similarity preservation in its objective function alongside with reconstruction error and sparsity. Most of factors in auto encoder have been evaluated in research [2]. However, sparsity constraint and different sparsity regularizer utilization are not evaluated.

It is common to use sparsity constraint in auto encoder. Sparsity forces several hidden units in hidden layer of auto encoder to be inactive [4]. This situation is achieved by giving penalty to most of hidden units. That penalty is produced by sparsity regularizer. A large number of hidden units is needed for auto encoder to learn the structure of data [3].

Based on research [2] and [3], this research evaluates the use of sparsity constraint in supervised auto encoder for single sample face recognition. Kullback-Leibler Divergence (KLD) will be used as sparsity regularizer. The proposed framework will be tested using two datasets, Extended Yale B (cropped version) and LFWcrop datasets. Another sparsity regularizers mentioned in [4], Log and εL_1, are used as comparison to KLD.

2. RELATED WORKS

2.1. Supervised Auto Encoder

Architecture of basic supervised auto encoder is given in Fig. 1. Unlike unsupervised auto encoder, supervised auto encoder uses two images, one for data sample and the other one for data label. Probe image, with variation of illumination, expression, pose, and so on is usually used for data sample.

Journal of Theoretical and Applied Information Technology
20th May 2016. Vol.87. No.2

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Gallery or database image is usually used for data label. Auto encoder structure learns data sample and label it with the most similar data label. The objective function of supervised auto encoder is presented in (1).

$$\min_{Wb_f b_g} \frac{1}{N} \sum_i \begin{pmatrix} \|x_i - g(f(\bar{x}_i))\|_2^2 \\ + \lambda \|f(x_i) - f(\bar{x}_i)\|_2^2 \\ + \alpha(KL(\rho_x \| \rho_0) + KL(\rho_{\bar{x}} \| \rho_0)) \end{pmatrix} \quad (1)$$

where the first part of the equation compute the reconstruction error and the second part compute similarity preservation (an original idea proposed by [2]). The third part uses KLD as sparsity regularizer. N denotes the number of sample images, λ is weight of similarity preservation and α is weight of sparsity. ρ is mapped average activation of the gallery and probe image. The value of W (weight), b_f (bias for encoder) and b_g (bias for decoder) are subjects for observation. The network structure uses hyperbolic tangent as its activation function.

2.2. Sparsity Regularizer

Sparsity regularizer used in supervised auto encoder to computer penalty for hidden units. Log and εL_1 (an approximation of L_1 norm) are examples of sparsity regularizer [4]. Log regularizer function is displayed in (2) while εL_1 in (3).

$$S(p) = \sqrt{p^2 + \varepsilon} \quad (2)$$

$$S(p) = \log(1 + p^2) \quad (3)$$

where ρ is the mapped average activation.

3. METHODS

The main purpose of this research is to evaluate the performance of deep auto encoder with KLD as sparsity regularizer in classifying image based on single sample face recognition. For that purpose, a deep auto encoder with initial weight sampled randomly from uniform distribution given in (4) is used [2][5][6]. For another parameter learning, L-BFGS algorithm is used [7].

$$-\sqrt{\frac{6}{d_h + d_x}}, \sqrt{\frac{6}{d_h + d_x}} \quad (4)$$

where d_h is the dimension of hidden units and d_x is the dimension of visible units (features of data sample).

In order to evaluate the performance of deep auto encoder, (1) is being modified and changed into (5).

$$\min_{Wb_f b_g} \frac{1}{N} \sum_i \begin{pmatrix} \|x_i - g(f(\bar{x}_i))\|_2^2 \\ + \lambda \|f(x_i) - f(\bar{x}_i)\|_2^2 \\ + \alpha(S(.) + S_{\bar{x}}(.)) \end{pmatrix} \quad (5)$$

where S(.) denotes the sparsity regularizer.

Before being processed using deep auto encoder, face imaged in both dataset are resized into 32x32 pixels.

4. DATASET

4.1. Extended Yale B

Extended Yale B contains of face images taken from 28 subjects with various poses and illumination. This dataset is available online [8]. In this research, cropped version of Extended Yale B is used [9].

4.2. LFWcrop

LFWcrop is cropped version of LFW dataset [10]. LFWcrop is available online [11].

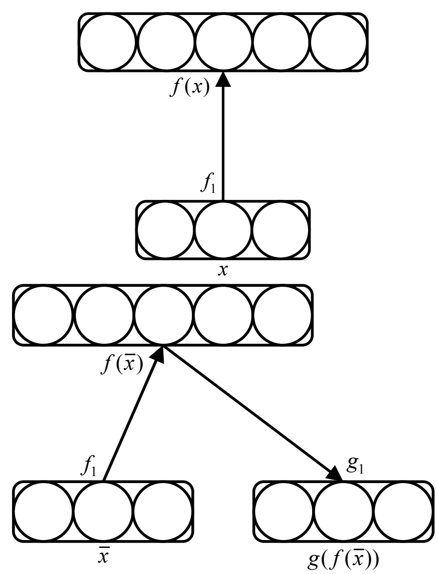

Figure 1. Supervised auto encoder architecture

It contains of face images taken from web. Some subjects may have more numbers than others. LFWcrop exhibits more natural image condition compared to Extended Yale B.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5. EXPERIMENT AND RESULT

Using initial parameter values that have been determined, the result shows that KLD has the best accuracy among three sparsity regularizers. The complete result is listed in Table 1.

Table 1. Average Accuracy of Deep Auto Encoder using KLD, Log, εL_1 Sparsity Regularizer

	Extended Yale B	LFWCrop
KLD		
Avg. Accuracy	77.8	97.68
Std. Dev.	18.13	3.13
Log		
Avg. Accuracy	77.45	97.63
Std. Dev.	18.3	3.17
εL_1		
Avg. Accuracy	77.3	97.58
Std. Dev.	18.23	3.39

For Extended Yale B dataset, standard deviation of average accuracy in each sparsity regularizer is considered high. Some of data sample in Extended Yale B has low illumination. It gives low accuracy in labelling process. Since deep auto encoder performance using LFWCrop is considered high, this research decides to use Extended Yale B for further experiments, to increase the average accuracy.

To optimize the average accuracy of deep auto encoder model trained and tested with Extended Yale B dataset, error threshold, weight of similarity preservation (λ), weight of sparsity (α) and number of hidden units are modified.

Based on experiment result with several error threshold ranged from 0.01 until 0.05, KLD, Log and εL_1 give similar result. The highest accuracy is obtained with error threshold 0.05. It is showed in Fig. 2.

Fig. 3 shows relation between λ modification and model accuracy. λ value addition increases the accuracy of model with KLD. Model with Log and εL_1 shows local minima accuracy in $\lambda = 5$. Therefore, in order to optimize model with KLD, λ value should be increased.

Fig. 4 shows how α value adjustment affects model accuracy. For model with KLD and Log, both of them reaches its optimum accuracy in $\alpha = 0.001$, while model with εL_1 is optimum in $\alpha = 0.01$.

Based on graph presents in Fig. 5, each number of hidden units gives different performance for KLD, Log and εL_1. For example, with 512 hidden units, model with Log has the highest accuracy and εL_1 has the lowest accuracy.

On the contrary, with 2048 hidden units, KLD has the highest accuracy and εL_1 has the lowest accuracy. Despite the rank variation, both three performances incline. With greater number of hidden units, results a better accuracy.

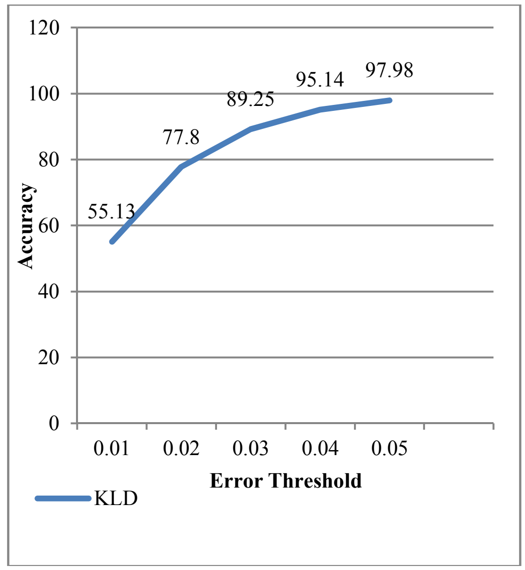

Figure 2. Graphic Of Accuracy Compared To Error Threshold

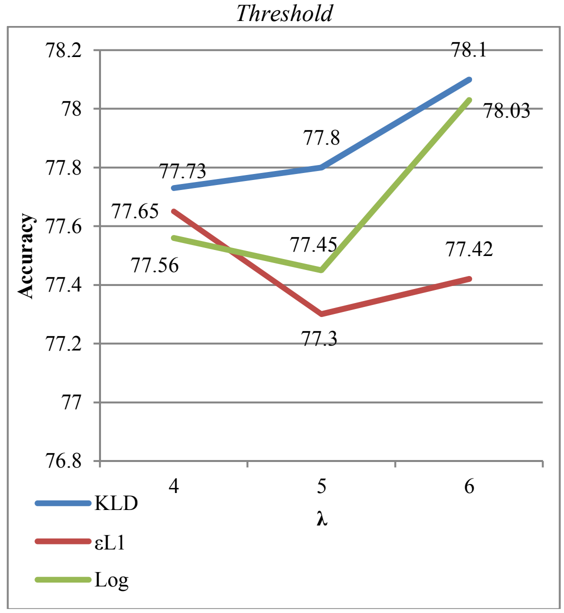

Figure 3. Graphic Of Accuracy Compared To Λ

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

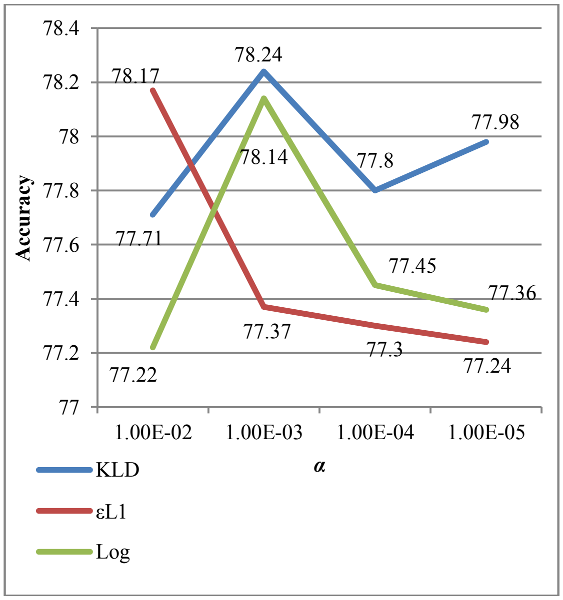

Figure 4. Graphic of Accuracy Compared to α

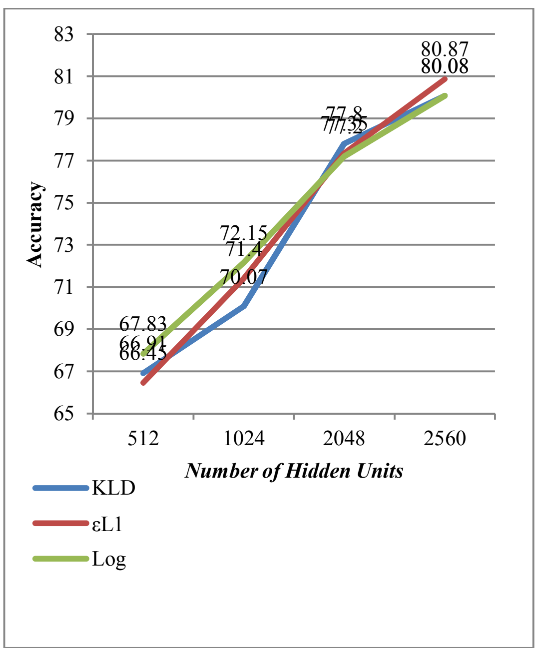

Figure 5. Graphic of Accuracy Compared to Number of Hidden Units

6. CONCLUSION

As a parameter, penalty value used in sparsity should be considered as important in order to obtain an optimal performance. This research concludes that KLD is recommended as sparsity regularizer for supervised deep auto encoder since it shows the best accuracy among other sparsity regularizer examined in this research.

REFERENCES:

[1] Ozdil, Ahmet and Metin Mete Ozbilen. "A survey on comparison of face recognition algorithms." *IEEE 8th International Conference on Application of Information and Communication Technologies (AICT),* 2014, pp. 1-3.

[2] Gao, Shenghua, et al. "Single Sample Face Recognition via Learning Deep Supervised Autoencoders." *IEEE Transactions on Information Forensics and Security,* 2015, pp. 2108-2118.

[3] Ng, Andrew. "Sparse Auto Encoder." *CS294A Lecture notes* 72. 2011.

[4] Zhang, Li and Yaping Lu. "Comparison of auto-encoders with different sparsity regularizers." *International Joint Conference on Neural Networks (IJCNN),* 2015, pp. 1-5.

[5] Glorot, Xavier, and Yoshua Bengio. "Understanding the difficulty of training deep feedforward neural networks." *International conference on artificial intelligence and statistics*, 2010, pp. 249-256.

[6] Bengio, Yoshua. "Practical recommendations for gradient-based training of deep architectures." *Neural Networks: Tricks of the Trade*. Springer Berlin Heidelberg, 2012, pp. 437-478.

[7] Liu, Dong C., and Jorge Nocedal. "On the limited memory BFGS method for large scale optimization." *Mathematical programming* (45), 1989, pp. 503-528.

[8] Georghiades, Athinodoros S., Peter N. Belhumeur, and David J. Kriegman. "From few to many: Illumination cone models for face recognition under variable lighting and pose." *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 2001, pp. 643-660.

[9] Lee, Kuang-Chih, Jeffrey Ho, and David J. Kriegman. "Acquiring linear subspaces for face recognition under variable lighting." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005, pp. 684-698.

[10] Sanderson, Conrad, and Brian C. Lovell. "Multi-region probabilistic histograms for robust and scalable identity inference." *Advances in Biometrics*. Springer Berlin Heidelberg, 2009, pp. 199-208.

[11] Huang, Gary B., et al. *Labeled faces in the wild: A database for studying face recognition in unconstrained environments* (1/2), University of Massachusetts, 2007.