COSENet: A constant ordinal regression for age estimation

Seungeun Han
Human Robot Interaction Laboratory, ETRI
hse@etri.re.kr

Junseong Bang
Police Science & Public Safety
ICT Research Center, ETRI
hjbang21pp@etri.re.kr

Abstract

Age estimation from face images has attracted attention because it is expected to have many application fields and growing interest. However, It remains challenging because the degree of aging varies from person to person. To address this issue, We propose a COSENet which is a constant ordinal regression with SENet. It can guarantees for rank-monotonicity and consistent confidence score. The result of the proposed model shows a substantial reduction of the prediction error compared to the reference ordinal regression network. we can conclude that our model has a positive effect on the predictive performance of an ordinal regression.

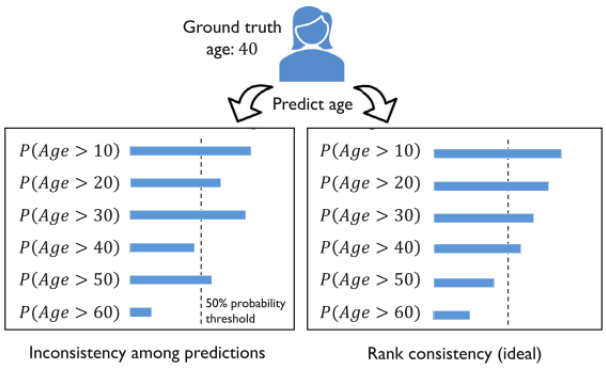

Figure 1. Illustration of inconsistencies

1. Introduction

Age estimation from face images has attracted attention because it is expected to have many application fields and growing interest. In previous methods, age estimation is often cast as a multi-class classification [15] [8] or a metric regression problem [7].

However, The multi-class classification is hard to classify labels to be independent to one another because the age labels have a strong ordinal relationships. The metric regression approach is better to utilize ordinal information for age estimation, but, still has difficulty because the degree of aging varies from person to person. For example, the aging process causing variations in facial shapes, sizes, and texture, has large individual differences due to numerous factors such as genes, diet, and lifestyle, and there are no clear aging characteristics in each age class [1] [6]. Implementations of this approach commonly suffer from classifier inconsistencies among the binary rankings. This inconsistency problem among the predictions of individual binary classifiers is illustrated in Figure 1.

To solve this problem, It would like to use the relative ordinal information among the age labels. Hence the ordinal regression is used at age estimation. [5] [9] [12] [3]. Ordinal regression describes the task of predicting labels on an ordinal scale. Recently, the ordinal regression problem is transformed into a series of simpler binary classification sub problems. For each rank $k \in \{1, 2, , K-1\}$, a binary classifier is trained according to whether the rank of a sample is larger than k. Then, the rank of a sample is predicted based on the classification results of the $K-1$ binary classifiers on this sample. A benefit of this kind of methods is that new generalization bounds for ordinal regression can be easily derived from known bounds for binary classification.

On the other hand, It is important to use the appropriate network for ordinal regression as a series of binary classification. In these days, many Convolution Neural Network(CNN) show improved performance at feature extraction from image data. One of them, The SENet [11] achieve good performance and robust at light change.

So, in this paper, we propose a COSENet which is a constant ordinal regression with SENet. The main contributions of this paper are as follows. First, Implementation of consistent rank logits framework to adapt common CNN architectures, such as SENet [11], for ordinal regression. Second, Experiments on different networks showing that It guaranteed binary classifier consistency improves predictive performance compared to the reference framework for ordinal regression.

2. Related Works

2.1. Ordinal Regression

In ordinal regression, the rank of an object is estimated. Order learning learns ordering relations between instances. By comparing a test instance with references with known ranks, it can estimate the rank of the instance. In other words, it can perform ordinal regression. But in many cases, ordinal regression is converted into multiple binary classification problems.

For instance, Herbrich et al. [13] proposed a method of support vector learning for ordinal regression. In [4], the perceptron ranking (PRank) algorithm proposed by Crammer and Singer is to generalize the online perceptron algorithm with multiple thresholds for ordinal regression. In [14], Shashua and Levin proposed new support vector machine formulations to handle multiple thresholds.

2.2. Ordinal Regression CNN

While earlier works on using CNNs for ordinal targets have employed conventional classification approaches, the general reduction framework from ordinal regression to binary classification by Li and Lin. [10] was recently adopted by Niu et al. [16] as Ordinal Regression CNN (OR-CNN). In the OR-CNN approach, an ordinal regression problem with K ranks is transformed into $K-1$ binary classification problems, with the kth task predicting whether the age label of a face image exceeds rank r_k, $k = 1, ..., K-1$. All $K-1$ tasks share the same intermediate layers but are assigned distinct weight parameters in the output layer.

While the OR-CNN was able to achieve state-of-the-art performance on benchmark datasets, it does not guarantee consistent predictions, such that predictions for individual binary tasks may disagree. This inconsistency could be suboptimal when the $K-1$ task predictions are combined to obtain the estimated age.

Chen et al. [2] proposed a modification of the OR-CNN, known as Ranking-CNN, that uses an ensemble of CNNs for binary classifications and aggregates the predictions to estimate the age label of a given face image. The researchers showed that training an ensemble of CNNs improves the predictive performance over a single CNN with multiple binary outputs, which is consistent with the well-known fact that an ensemble model can achieve better generalization performance than each individual classifier in the ensemble.

Recent research has also shown that training a multi-task CNN that shares lower-layer parameters for various face analysis tasks(face detection, gender prediction, age estimation, etc.) can improve the overall performance across different tasks compared to a single-task CNN.

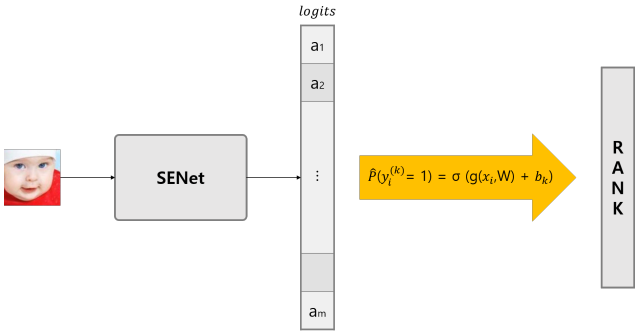

Figure 2. Illustration of the consistent rank logits SENet(COSENet) used for age prediction.

3. Proposed Method

Let $D = \{x_i, y_i\}_{i=1}^{N}$ be the training dataset consisting of N training examples. $x_i \in X$ denotes the ith training example and y_i the corresponding rank, where $y_i \in Y = \{r_1, r_2, ..., r_K\}$ with ordered rank $r_K > r_{K-1} > ... > r_1$. The ordinal regression task is to find a ranking rule h : X $\rightarrow$ Y such that a loss function L(h) is minimized. Let C be a K × K cost matrix, where C_{y,r_k} is the cost of predicting an example (x, y) as rank r_k. That is, $C_{y,r_{k-1}} \geq C_{y,r_k}$ if $r_k \leq y$ and $C_{y,r_k} \leq C_{y,r_{k+1}}$ if $r_k \geq y$. In ordinal regression, where the ranks are treated as numerical values, the absolute cost matrix is commonly defined by $C_{y,r_k} = |y - r_k|$.

3.1. COSENet

3.1.1 label prediction

Given a training dataset D, a rank y_i is first extended into K - 1 binary labels $y_i^{(1)}, ..., y_i^{(K-1)}$ such that $y_i^{(k)} \in \{0, 1\}$ indicates whether y_i exceeds rank r_k. Using the extended binary labels during model training, we train a single SENet with K - 1 binary classifiers in the output layer, which is illustrated in Figure 2.

To achieve rank monotonicity and guarantee binary classifier consistency, the K - 1 binary tasks share the same weight parameters but have independent bias units.

3.1.2 Loss Function

Let W denote the weight parameters of the neural network excluding the bias units of the final layer. The final layer, whose output is denoted as $g(x_i, W)$, shares a single weight with all nodes in the final output layer; K - 1 independent bias units are then added to $g(x_i, W)$ such that $\{g(x_i, W) + b\}_{K=1}^{K-1}$ are the inputs to the corresponding binary classifiers in the final layer.

$$\sigma(z) = 1/(1 + exp(-z)) \quad (1)$$

Equation 1. is the logistic sigmoid function. The predicted probability for task k is defined as Equation 2.

$$\hat{P}(y_i^{(k)}) = \sigma(g(x_i, W) + b_k) \tag{2}$$

For model training, we minimize the loss function as Equation 3, which is the weighted cross-entropy of K - 1 binary classifiers. $\lambda^{(k)}$ denotes the weight of the loss associated with the kth classifier (assuming $\lambda^{(k)} > 0$).

$$\begin{aligned} L(W, b) = &- \Sigma_{i=1}^{N} \Sigma_{k=1}^{K-1} \lambda^{(k)} [log(\sigma(g(x_i, W) + b_k) y^{(k)} \\ &+ log(1 - \sigma(g(x_i, W))(1 - y_i^{(k)})] \end{aligned} \tag{3}$$

For rank prediction (Equation 1.), the binary labels are obtained via Equation 4.

$$\begin{aligned} f_k(x_i) &= 1, \quad if \hat{P}(y_i^{(k)} = 1) > 0.5 \\ f_k(x_i) &= 0, \quad otherwise \end{aligned} \tag{4}$$

4. Experiments

4.1. Dataset

We use The Asian Face Age Database(AFAD) by Niu et al.[19] for our experiments. The AFAD is a dataset proposed for evaluating the performance of age estimation, which contains more than 160K facial images and the corresponding age and gender labels. This dataset is oriented to age estimation on Asian faces, so all the facial images are for Asian faces. It is noted that the AFAD is the biggest dataset for age estimation to date. It is well suited to evaluate how deep learning methods can be adopted for age estimation. There are 164,432 photos in the AFAD dataset. It consists of 63,680 photos for female as well as 100,752 photos for male, and the ages range from 15 to 40. But in this study, we used just 165,501 photos in the range of 15–40 years. Since the faces were already centered, we don't need to preprocessing it. It can be downloaded in https://afad-dataset.github.io/

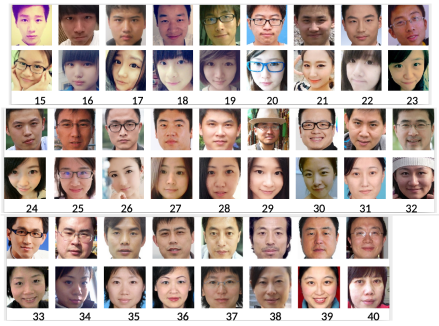

Figure 3. The example of AFAD Dataset

4.2. Neural network architectures

To extract the features for age estimation from face images, we use the SENet architecture, especially SENet modified by Resnet34, which achieves good performance on a image classification tasks. we refer to the original SENet with standard cross-entropy loss. To implement a SENet for ordinal regression using the proposed COSENet, we replaced the last output layer with the corresponding binary tasks.

4.3. Training and evaluation

For model evaluation, we computed the Mean Absolute Error(MAE, Equation 5.) and Root Mean Squared Error(RMSE, Equation 6.), on the test set after the last training epoch.

$$MSE = \frac{1}{N} \Sigma_{i=1}^{N} |y_i - h(x_i)| \tag{5}$$

$$RMSE = \sqrt{\frac{1}{N} \Sigma_{i=1}^{N} |y_i - h(x_i)|} \tag{6}$$

where y_i is the ground truth rank of the ith test example and $h(x_i)$ is the predicted rank.

The model training was repeated three times with different random seeds(0, 1, and 2) for model weight initialization, while the random seeds were consistent between the different methods to allow fair comparisons.

The networks were trained for 50 epochs with stochastic gradient descent via adaptive moment estimation using exponential decay rates $_0 = 0.90$ and $_2 = 0.99$ and a batch size of 256.

Moreover, we chose a uniform task weighting for the cross-entropy of K - 1 binary classifiers.

By hyperparameter tuning on the validation set, we found the optimal learning rate of $\alpha = 5 * 10^{-5}$.

From those 50 epochs, the best model was selected via MAE performance on the validation set. The selected model was then evaluated on the independent test set, from which the reported MAE and RMSE performance values were obtained.

4.4. Results

We conducted three experiments with different random seeds(0, 1, and 2) on one face image dataset to compare the proposed COSENet with other ordinal regression approaches. All implementations were based on the ResNet-34 architecture.

Table 1. shows the result of age estimation on test dataset. Our proposed COSENet shows a substantial performance improvement over CE-CNN and OR-CNN. By a several experiments, we can conclude that our model has a positive effect on the predictive performance of an ordinal regression.

Table 1. Age estimation errors on test datasets.

Method	Random Seed	AFAD	
		MSE	RMSE
CE-CNN	0	3.58	5.01
	1	3.58	5.01
	2	3.62	5.06
	AVG ± SD	3.60 ± 0.02	5.03 ± 0.03
OR-CNN	0	3.56	4.80
	1	3.48	4.68
	2	3.50	4.78
	AVG ± SD	3.51 ± 0.04	4.75 ± 0.06
COSENet	0	3.53	4.76
	1	3.51	4.68
	2	3.50	4.77
	AVG ± SD	3.51 ± 0.01	4.73 ± 0.04

5. Conclusions

In this paper, we proposed COSENet for ordinal regression via extended binary classification for classifier consistency. We implement consistent rank logits framework to adapt common CNN architectures, such as SENet, for ordinal regression. It can guarantees for rank-monotonicity and consistenet confidence score. Moreover, we can see improved predictive performance compared to the reference framework for ordinal regression by using our model. Our method can be readily generalized to other ordinal regression problems and different types of neural network architectures.

References

[1] Karl Ricanek Jr. A. Midori Albert and Eric Patterson. A review of the literature on the aging adult skull and face: Implications for forensic science research and applications. *Forensic Science International*, pages 1–9, 2007. 1

[2] Chu-Song Chen Winston H. Hsu Bor-Chun Chen. Cross-age reference coding for age-invariant face recognition and retrieval. *ECCV*, pages 768–783, 2014. 2

[3] J. Liu C. Li, Q. Liu and H. Lu. Learning ordinal discriminative features for age estimation. *CVPR*, pages 2570–2577, 2012. 1

[4] Singer Y Crammer K. Pranking with ranking. *Advanced in Neural Information Processing Systems*, pages 641–647, 2002. 2

[5] Z. Zhang J. Feng D. Cao, Z. Lei and S. Li. Human age estimation using ranking svm. *CCBR*, pages 324–331, 2012. 1

[6] Leslie Zebrowitz. Reading Faces. Window to the soul? *Westview Press*, 1997. 1

[7] Y. Fu and T. Huang. Human age estimation with regression on discriminative aging manifold. *IEEE Transactions on Multimedia*, pages 578–584, 2008. 1

[8] Y. Fu G. Guo, G. Mu and T. Huang. Human age estimation using bio-inspired features. *CVPR*, pages 112–119, 2009. 1

[9] C. Chen K. Chang and Y. Hung. Ordinal hyperplanes ranker with cost sensitivities for age estimation. *CVPR*, pages 585–592, 2011. 1

[10] L. Li and H. Lin. Ordinal regression by extended binary classification. *NIPS*, pages 865–872, 2006. 2

[11] Peter McCullagh. Regression models for ordinal data. *J. R. Stat. Soc. Ser. B (Methodological)*, pages 109–142, 1980. 1

[12] L. Zhong P. Yang and D. Metaxas. Ranking model for facial age estimation. *ICPR*, 2010. 1

[13] T. Graepel R. Herbrich and K. Obermayer. support vector learning for ordinal regression. *Proc. Int. Conf. Artif. Neural Netw*, pages 97–102, 1999. 2

[14] A. Shashua and A. Levin. Ranking with large margin principle: Two approaches. *NIPS*, pages 961–968, 2003. 2

[15] Z. Zhou X. Geng and K. Smith-Miles. Automatic age estimation based on facial aging patterns. *IEEE T-PAMI*, pages 2234–2240, 2007. 1

[16] L. Wang X. Gao Z. Niu, M. Zhou and G. Hua. Ordinal regression with multiple output cnn for age estimation. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4920–4928, 2016. 2