Classification of Human Emotions with CNN deep learning network models

Ceren Umay Özten

I. INTRODUCTION

In classification of human emotions with CNN network models, data characteristics of human expressions are analyzed. Analysis of human expressions, mimics and coordinations of facial muscles is prevalent over computer science discipline and also real life applications. Gaming industry, caregiving industry, artificial intelligence industry and robotic industry are largely benefitted from analysis of human expressions, mimics and coordinations of facial muscles cluing human emotions. In Japan a robotic in guise of sea lian cup cover, can analyze human emotions through human expressions, mimics and coordinations of facial muscles and can react in accordance of patient’s emotive situations. Especially coordinations of facial muscles known as facial expressions route critic steps in examination of human emotions and create reference data for evaluating human emotions across different categories such as disgust, happiness, sadness etc. Facial expressions includes identifying patterns for separation of human emotion categories.

In convolutional neural network CNN architecture, pixel data of images of facial expressions fed into unit neurons throughout deep layers and deep layers result in more complete characteristic pieces of the images that can be classified under specific human emotions. And in classification of human emotions, images of facial expressions is provided by FER-2013 facial expressions image data set. FER-2013 facial expressions image data set pixel data passes through unit neurons of dense CNN layers and encounters training process among unit neurons. Training process has dependency on metric qualifications of CNN dense layered system. Layer metrics, compilation metrics and fit metrics are important integrations of CNN dense layered system. Layer metrics are number of layers, number of unit neurons, filter size, padding, strides, activation function. Focused fit metrics are batch size and epoch number. Meanwhile focused compilation metrics are optimization function and loss function. Fit metrics and compilation metrics set basic mechanics of training process. Layer metrics set synchronization between pixel data of images and CNN network model.

FER-2013 data set

FER-2013 facial expressions data set includes 28709 train and 3589 test samples. Train and test samples represent fundamental human emotions over 7 different folders: angry, disgust, fear, happy, surprise, neutral, sad. Entire samples are in grayscale (48x48) matrix dimensions and samples have not homogeneous distributions over 7 different folders. Happy folder outnumbers other emotion folders in train samples. Train and test samples have a range from photographic realistic samples to 2D cartoon samples due to Google web nature. Samples are extracted from Google web domain.

LITERATURE REVIEW

Creation of CNN network models covers extensive literature research. CNN network models discussed in later concessions, have trails of literature research. Compilation metrics, fit metrics and filter size, number of layers, BatchNormalization-Dropout integrations are critically weighted features throughout CNN network building process. Literature examinations connected to that CNN network building process. And CNN network building process is consisted of experiments. There are five experiments that modifying nuclear CNN model layer metrics. First performance of nuclear CNN model is experimented with fit metrics and compilation metrics. Then experiments over layer metrics are implemented on nuclear CNN model and performance of nuclear CNN model is analyzed during layer modifications.

*First Experiment Literature Review*

In first experiment, convolutional layers are added as block manner with integration of maxpooling layer at the end of each convolutional block and Dense SIFT CNN network model is reviewed for layer organization. Dense SIFT mechanism divides image data into equal (12x12) dimension pixel areas and extracts pixel gradients. And pixel gradients are fed into Dense SIFT CNN network. Dense SIFT CNN network has six convolutional layers as two-stage block manner with integration of three maxpooling layers at the end of each convolutional block. Total six convolutional layer and three maxpooling layer structure Dense SIFT CNN network. Dense SIFT CNN network model has Dropout integration between fully-connected (FC) layers and convolutional layers. Dense SIFT CNN network performs 72.1% accuracy rate on FER-2013 test data []. First experiment’s CNN network model has four convolutional layers as two-stage block manner with integration of two maxpooling layers in contrast to Dense SIFT CNN network model and first maxpooling layer and fully-connected (FC) layer has Dropout integration. Dense SIFT CNN network outperforms first experiment’s CNN network model in comparison of Dense SIFT CNN network accuracy rate on FER-2013 test data and first experiment’s CNN network model val accuracy rate on FER-2013 train data.

*Second Experiment Literature Review*

In second experiment, regularization technique and Dropout-BatchNormalization-maxpooling three-stage implementation are galvanized by deep CNN network of Shimah Alizadeh and Azar Fazel. Deep CNN network integrates L2 regularization technique and Dropout-BatchNormalization integrations as overfitting prevention scheme and deep CNN network has total four convolutional layers and two fully-connected (FC) layers. First convolution layer has filter size of (3x3) with 64 unit neurons, second convolution layer has filter size of (5x5) with 128 unit neurons, third convolution layer has filter size of (3x3) with 512 unit neurons and fourth layer has filter size of (3x3) with 512 unit neurons. Also each convolution layer has strides of 1 and BatchNormalization-Dropout-maxpooling three-stage implementation. First fully-connected (FC) layer has 256 unit neurons, second fully-connected (FC) layer has 512 unit neurons. Fully-connected (FC) layers support Dropout-BatchNormalization integration. Regularization value is 1e-7. Deep CNN network val accuracy rate is 65% []. Second experiment’s CNN network consists of four convolutional layers and two fully-connected (FC) layers. And each convolutional layer pair includes Dropout-BatchNormalization-maxpooling three-stage integration. Convolutional layers have 64-128-256-512 unit neurons from first convolutional layer to last convolutional layer. Fully-connected (FC) layers also have 128-64 unit neurons sequentially. Entire convolutional layer structure has filter size of (3x3) and strides of 2. Regularization value is assigned as 0.01. Especially in comparison analysis, deep CNN network regularization value is critically smaller than second experiment’s CNN network regularization value. Second experiment’s val accuracy performance can not reach deep CNN network’s 64% val accuracy performance.

*Third Experiment Literature Review*

In third experiment Dropout-BatchNormalization-maxpooling three-stage implementation has fully-connected (FC) layers with 1024 unit neurons and Dropout-BatchNormalization integration. And five-layer CNN model of Stanford research paper is used as exemplary structure. Stanford research paper’s five-layer CNN model has three convolutional layers and two fully-connected (FC) layers including last softmax output layer. Three convolutional layers have 32-32-64 unit neurons and filter sizes of (5x5), (4x4) and (5x5) from first convolutional layer to last convolutional layer. Five-layer CNN model implements BatchNormalization-maxpooling integration on convolutional layers. Especially BatchNormalization integration of fully-connected (FC) layers is important part of third experiment and 1024 unit neurons of fully-connected (FC) layers influence third experiment. Five-layer CNN model performs 66.3% accuracy rate on FER-2013 test data []. Third experiment’s CNN model performs approximately 17% on FER-2013 test file. Increase of unit neurons of fully-connected (FC) layers (1024) can not improve overfitting problem.

*Fourth Experiment Literature Review*

ASTESJ research paper’s VGG-16 GAP CNN network results in increase of filter sizes of Dropout-BatchNormalization-maxpooling three-stage implementation, coding of Dropout-BatchNormalization-averagepooling and application of SGD-early stopping technique. In ASTESJ research paper, fully-connected (FC) layers’ except last softmax are omitted from VGG-16 architecture in favour of GAP averagepooling layer against overfitting problem. VGG-16 architecture has block convolutional layer structure with filter size of (3x3) and demonstrates same behaviour as one single convolutional layer with filter size of (5x5) or (7x7). VGG-16 GAP CNN network also integrates SGD optimization function and early stopping together []. Fourth experiment’s CNN network model has filter size of (5x5) in all convolutional layer and averagepooling layer prior to fully-connected (FC) layers. Early stopping technique drops 150 epoch number to less 52 epoch number. Fourth experiment performs 34.30% accuracy rate on FER-2013 test file. In comparison fourth experiment improves accuracy performance over third experiment.

LearningRateSchedular version of fourth experiment’s CNN network model has lower performance than early stopping version of it.

*Fifth Experiment Literature Review*

Nuclear CNN model’s “base model” skeleton is replaced by VGG-16 architecture and transfer learning method is implemented on nuclear CNN model. Pre-trained imageNET neuron nodes are added to nuclear CNN network and nuclear CNN model’s fully-connected (FC) layers extract classification result for added VGG-16 architecture. Stanford research paper’s transfer learning implementation performs 70.2% accuracy rate on FER-2013 test data. Fifth experiment’s VGG-16 implementation performs 63.9% val accuracy rate. Fifth experiment’s lower accuracy rate causes from dimension differentiation of fifth experiment.

DEEP CNN MODELS

**VGG-16**

VGG-16 architecture is popular element across academic landscape covering computer science. Implementation of transfer learning uses a skeleton of CNN network architectures similar to VGG-16 and benefits from pre-trained neuron nodes of convolutional layers with deeper segments-complexities. Especially fifth experiment’s transfer learning uses skeleton of VGG-16 CNN network architecture. Fully-connected (FC) layers of nuclear CNN model is integrated to VGG-16 CNN network architecture and FER-2013 train samples are trained through base layers of VGG-16 architecture.

VGG-16 architecture has 13 convolutional layers and 3 fully-connected (FC) layers that three-stages and two stages of block convolutional layers follow maxpooling layers. Convolutional layers have filter size of (3x3) and strides of 1 throughout the entire architecture. And maxpooling layers have filter size of (2x2) and strides of 2 consistently. VGG-16 architecture does not include higher filter sizes of (5x5) or (7x7) and continues same filter sizes of (3x3) consistency for all convolutional layers. VGG-16 architecture. Huge number of parameters increases connection possibilities of weighted CNN network of VGG-16.

Fifth experiment’s VGG-16 transfer learning implementation performs 63.9% val accuracy rate and 1.3355 val loss. FER-2013 test file samples are combined into batch format for val accuracy rate calculation during training process and 64 batch size and 70 epoch number are assigned as fit metrics. SGD optimization function learning rate is assigned as 0.01 value.

Fifth experiment’s val accuracy performance outperforms rival experiments that is created from building own CNN network models. Transfer learning method demonstrates more efficiency over built-from-scratch CNN network models.

**Nuclear CNN Model**

Nuclear CNN model is start point of experiments except fifth experiment and modification of layer metrics refers to nuclear CNN model. Structure of nuclear CNN model has three convolutional layers and two fully-connected (FC) layers with Dropout or BatchNormalization integrations. Maxpooling layers are embedded between convolutional layers. Unit neurons have 75-50-25 sequential descendancy and convolutional layers have filter sizes of (3x3). SGD, Adam, Adadelta optimization functions are experimented on nuclear CNN model. Epoch number and batch size fit metrics differentiations are also experimented on CNN nuclear model.

**Second Experiment CNN Network Model**

Second experiment’s CNN network model is constructed for analysis of deeper and more complex CNN structures in comparison to nuclear CNN model and examining of regularization technique on decreasing overfitting problem of complex CNN structures. Second experiment’s CNN network model has 64-128-512-256 unit neurons in ascendancy in contrast to nuclear CNN model and results in much more trainable parameters than nuclear CNN model concluded from comparison of 1,626,119 trainable parameters with 48,244. More trainable parameters also make second experiment CNN network model more prone to overfitting problem. L2 regularization technique and Dropout-BatchNormalization integrations are implemented on caused overfitting problem.

**Fourth Experiment CNN Network Model**

GAP averagepooling layer and implementation of early stopping technique with SGD optimization function implies distinction from rival experiments. Similar to second experiment, fourth experiment has four convolutional layers and three fully-connected (FC) layers with last softmax output layer and convolutional layers have Dropout-BatchNormalization-maxpooling three-stage implementation. Just after last convolutional layer Dropout-BatchNormalization-averagepooling three stage implementation creates GAP reaction and averaging over image pixels. Convolutional layers’ unit neurons have 32-32-64-64 sequential repetition and two fully-connected (FC) layers have 1024 unit neurons. Overfitting complexity of fully-connected (FC) layers 1024 unit neurons is expensed by decreasing of convolutional layers unit neurons from 64-128-256-512 to 32-32-64-64 sequence. Fourth experiment’s CNN network model is benefitted from modification of layer metric unit neurons.

ANALYSIS AND RESULTS

Nuclear CNN model is examined on a bit range of optimization functions rmsprop, SGD, Adam, Adadelta. Especially effects of batch size and epoch number fit metrics are analyzed through the optimization functions. First nuclear CNN model

Nuclear CNN model performs higher accuracy rate on train data than second experiment’s CNN network model in contrast second experiment’s CNN network model performs less difference between accuracy rates on train data and validation data that demonstrates better overcome of overfitting problem. Integration of L2 regularization technique and three stage Dropout-BatchNormalization-maxpooling implementation reduce overfitting range and close the gap between train accuracy rate and validation accuracy rate.

Third experiment’s CNN network model performs higher accuracy rate on train data than fourth experiment CNN network model despite its relatively poor performace on test data and reducing overfitting problem. And third experiment’s significant poor performance on test data is caused by density of unit neurons across convolutional layers and additional impact of fully-connected (FC) layers’ unit neurons. Third experiment’s Dropout-BatchNormalization implementation is inadequate for countering complexity of structure. While fourth experiment’s CNN network model decreases overfitting problem in assistance of Dropout-BatchNormalization-averagepooling three-stage implementation and modification of convolutional layers’ unit neurons. Especially on test data layer qualifications of fourth experiment outperforms third experiment. Decreasement of unit neurons on convolutional layers lowers number of connections in CNN network domain plummeting overfitting situation.