1. Data Manipulation

Dong-Gyu, Lee

Dept. of Statistics, KU

2020, Jan 14

Contents

Analysis Goal

- Fit Neural Networks using unbalanced data.
- Compare the performance of NNs in balanced and unbalanced data.
- In this time, we will consider various types of CNN(Convolutional Neural Networks).

CNN

- CNN has the following structure:

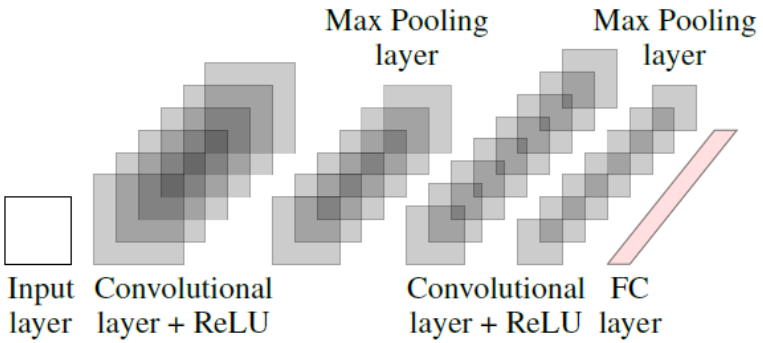


Figure 1: CNN Structure

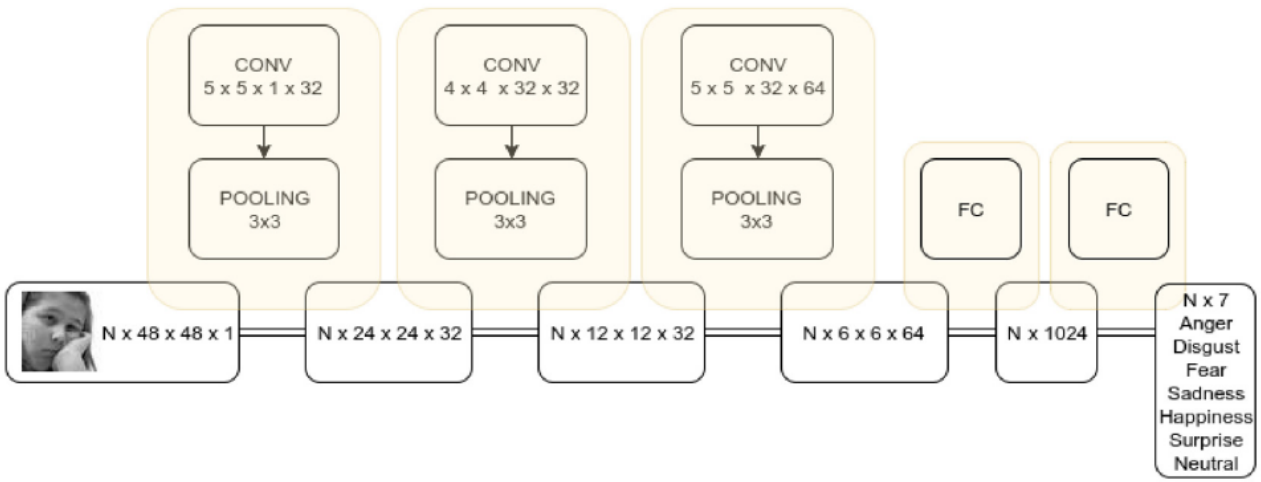


Figure 2: Example of CNN Architecture

FER2013 Data

- The FER2013 used in the ICML 2013 workshop[2] is a dataset to classify face emotion. Abbreviation 'FER' stands for facial emotion recognition.

Figure 3: FER2013 Images

- It has 7 labels : Anger, Disgust, Fear, Happiness, Sadness, Surprise, Neutral
- If you want to get more information, visit the following website : https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge

FER2013 Data

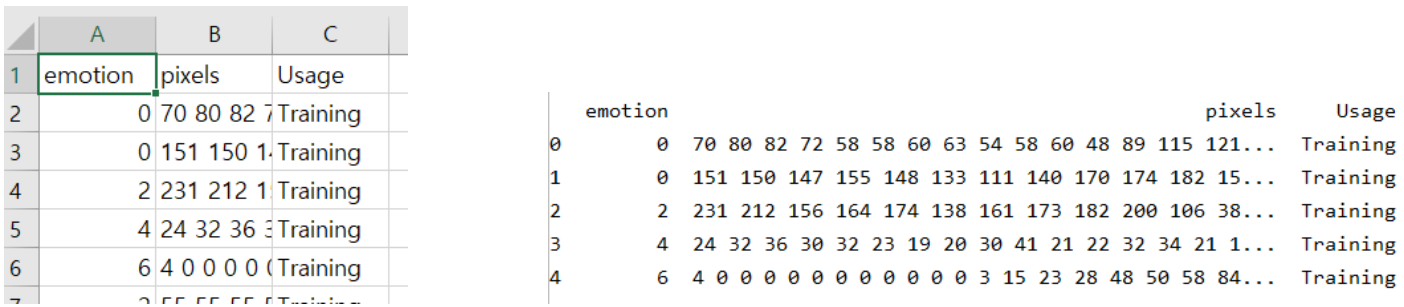
- Using FER2013 dataset, human accuracy was $65 \pm 5\%$
- With the various CNN models, the best accuracy at this point is well over 75%
- The best CNN model under the 'Null' condition in 2013 obtains an accuracy of 60%. Also using an ensemble of such models, the it obtained an accuracy of 65.5% in the same period
- 56 teams submitted on the final dataset. And the accuracy without any constraints exceeded 70% of the first team, and the rest was less than 70%.

FER2013 Data

- The fact that human accuracy was only 65% means that the dataset itself is incorrect.
- Or the answer between the label and the data may be wrong.
- To complement this, the FER+ dataset was released in 2016.

FER2013 Data Structure

- The data dimension is 35887×3 where rows are the data size and columns are 'emotion', 'pixels', 'Usage'.



	A	B	C
1	emotion	pixels	Usage
2	0	70 80 82 7	Training
3	0	151 150 1	Training
4	2	231 212 1	Training
5	4	24 32 36 3	Training
6	6	4 0 0 0 0 (	Training

	emotion	pixels	Usage
0	0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121...	Training
1	0	151 150 147 155 148 133 111 140 170 174 182 15...	Training
2	2	231 212 156 164 174 138 161 173 182 200 106 38...	Training
3	4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1...	Training
4	6	4 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84...	Training

Figure 4: Dataset in Excel & Python

- 0:Anger, 1:Disgust, 2:Fear, 3:Happiness, 4:Sadness, 5:Surprise, 6:Neutral
- 'Usage' is divided into three classes: Training, Private test, Public test

FER2013 Data Structure

- Based on the previous dataset, the number of labels which are divided by purpose of use can be checked as follows :

Happiness	8989
Neutral	6198
Sadness	6077
Fear	5121
Anger	4953
Surprise	4002
Disgust	547
dtype: int64	
Total: 35887	

Happiness	7215
Neutral	4965
Sadness	4830
Fear	4097
Anger	3995
Surprise	3171
Disgust	436
dtype: int64	
Train: 28709	

Happiness	879
Neutral	626
Sadness	594
Fear	528
Anger	490
Surprise	416
Disgust	55
dtype: int64	
Private test: 3588	

Happiness	895
Sadness	653
Neutral	607
Fear	496
Anger	466
Surprise	415
Disgust	56
dtype: int64	
Public test 3588	

Figure 5: The Number of Labels for Each Use

FER2013 Data Structure

- And this can be represented by the following pieplot.

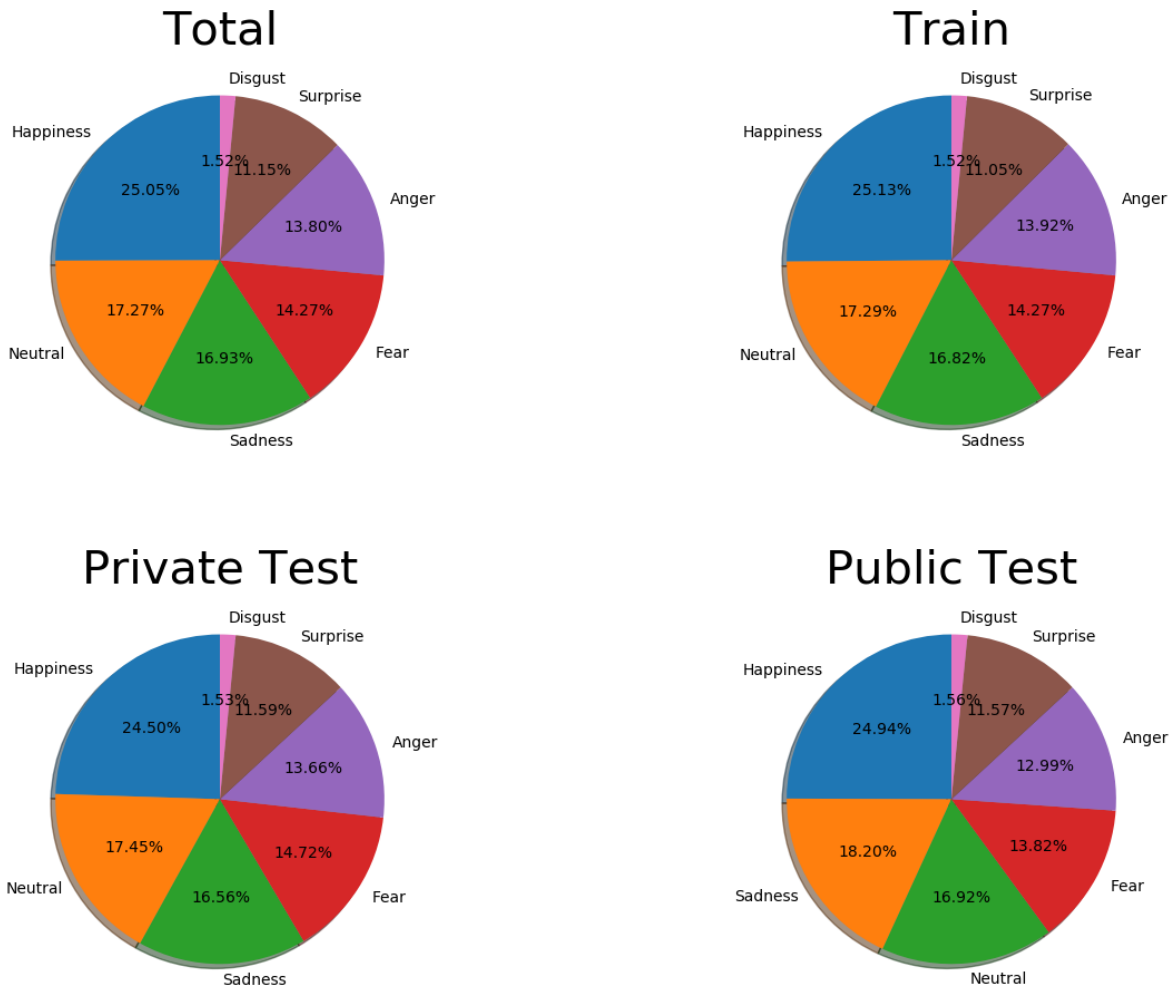


Figure 6: The Ratio of Labels for Each Use

Preprocessing

- First, useless data is hidden in the FER2013 dataset. It is shown as a black image.

Figure 7: Black Image

- The total number of black images is 12, which consists of 11, 0, 1 black images in training, private test, public test repectively.

Preprocessing

- You can also see the the number of data by usage below.

	Before			After		
	Training	Private Test	Public Test	Training	Private Test	Public Test
Counts	28709	3589	3589	28698	3589	3588

Table 1: Data Counts

Preprocessing

- Second, FER2013 is a small data(48×48 pixels) basically. For improvement of model, all images are preferred to change 256×256 pixels[3]

Figure 8: 48×48 Pixels and 256×256 Pixels

- All work is done in Python.

Next

- In preprocessing, data augumentation can be expressed in various ways :
  1. Flip all data left and right for small number of labels.
  2. Make an object(data) through the rotation or translation invariance.
  3. Use the ADASYN, SMOTE, GAN , etc.
  4. Balance the label with other data.
- Also we will cover the various model such as the Alexnet, VGG, ResNet etc.
- And last, we will talk about the various ensembles like voting system, through the combination of above models[1, 4]

Reference

[1] Yang Fei and Guo Jiao. Research on facial expression recognition based on voting model. In *IOP Conference Series: Materials Science and Engineering*, volume 646, page 012054. IOP Publishing, 2019.

[2] Ian J Goodfellow, Dumitru Erhan, Pierre Luc Carrier, Aaron Courville, Mehdi Mirza, Ben Hamner, Will Cukierski, Yichuan Tang, David Thaler, Dong-Hyun Lee, et al. Challenges in representation learning: A report on three machine learning contests. In *International Conference on Neural Information Processing*, pages 117–124. Springer, 2013.

[3] Hong-Wei Ng, Viet Dung Nguyen, Vassilios Vonikakis, and Stefan Winkler. Deep learning for emotion recognition on small datasets using transfer learning. In *Proceedings of the 2015 ACM on international conference on multimodal interaction*, pages 443–449. ACM, 2015.

[4] Alessandro Renda, Marco Barsacchi, Alessio Bechini, and Francesco Marcelloni. Comparing ensemble strategies for deep learning: An application to facial expression recognition. *Expert Systems with Applications*, 2019.