Machine Learning Engineer Nanodegree

Capstone Proposal

Swaraj Jena
27 May 2017

Emotion Recognition using Deep Learning

Domain Background

Recently one of the top application of Artificial intelligence is to recognise and understand human faces. One of the advanced development in this field is Emotion Recognition. In addition to identifying faces ,the computer is now able to determine the facial expression and hence the emotion of a person. There are various potential applications of emotion recognition. One possible application is in the area of surveillance and behavioural analysis of people for law enforcement. It can also be used in theme parks or shopping malls to understand customer behaviour ,so that the services can be offered accordingly. It can also be used for the robots to keep track of the mental state of the user so that it can behave appropriately. Emotion recognition therefore plays a key role in improving human machine interaction,

Due to the importance of facial expression in designing Human–computer interaction systems, various feature extraction and machine learning algorithms have been developed for Facial Expression Recognition. Most of these methods are hand-crafted features extractions followed by a classifier such as [1] who's used Local binary pattern feature extractor with SVM classification, Haar[2], SIFT[3], Gabor filters with fisher linear discriminant[4], and Local phase quantization (LPQ) [5]. The recent success of convolutional neural networks (CNNs) in tasks such as image classification has been extended to the problem of facial expression recognition[8]. Unlike traditional machine learning and computer vision approaches where features are defined by hand, CNN learns to extract the features directly from the training database using iterative algorithms like gradient descent.

Problem Statement

The majority of existing techniques focus on classifying 7 basic (prototypical) expressions, which have been found to be universal across cultures and subgroups, namely: neutral, happy, surprised, fear, angry, sad, and disgusted. Main objective of my research is to build a CNN based deep learning model capable of deriving these emotion of a person through pictures of his/her face.Given a picture of a person my model will automatically determine various probabilities of above mentioned emotions and find the best emotion that fits the picture. I will also analyze the situations where the model might fail to recognise emotion correctly.Humans are well-trained in reading the emotions of others. So the main objective of this research is **Can a computer interpret facial expression with human level accuracy?**

Datasets and Inputs

For emotion recognition ,several datasets are available for research,varying from a few hundred high resolution photos to tens of thousands of smaller images. Most famous publicly available datasets are the Facial Expression Recognition Challenge (FERC -2013)[8],Extended Cohn-Kanade (CK+)[9]. In CK+ the facial expressions are posed (i.e 'clean') as all images are captured in controlled environment in a lab, while in FERC-2013 set shows emotions 'in the wild'. This makes the pictures from the FERC-2013 set harder to interpret,but given larger size of the dataset,the diversity can be beneficial for the robustness of the model. Because of this reason for my work I am going to use FERC-2013 dataset.

This FERC-2013 dataset is available in kaggle for public research purpose. This dataset has around 35000 low resolution images and corresponding emotion. All images are grayscale and have a resolution of 48 by 48 pixels.For my model I will be using these image values matrix as the inputs. As we are going to use sequence of CNN in the model,hence we need a lot of training data for the model to generalize. Hence during training, some randomness is introduced in the data augmentation process .

Solution Statement

I will build a deep neural network model using sequence of Convolutional layers,max pooling layers and feed forward layers. A softmax layer will be added in the end to

output various probabilities ,each corresponding to one emotion. **Categorical cross entropy loss** for the output with respect to the correct emotion will be calculated and the loss will be back propagated to train the model. This model will be trained on FERC-2013 dataset. data augmentation in the form of random horizontal shifts, random vertical shifts, and random horizontal flips,rotation ,zoom etc will be applied to the training image to combat overfitting of the model. I will experiment on various combinations of convolutional, pooling and Feed forward layers and will choose the one giving best accuracy score. Accuracy score on validation dataset will be used to assess the model.

For our model categorical cross entropy loss function is calculated as

Loss L(θ) = $-\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} log(p_{ij})$

Where,

i indexes samples/observations

j indexes classes

y is the sample label (one-hot vector)

$p_{ij} \in (0, 1) : \sum_{j} p_{ij} = 1 \ \forall \ i,j$ is the estimated prediction by our model.

To calculate accuracy we need to first find the best emotion for each sample as

$$p\_best_{i} = argmax_{j} \, p_{ij}$$

Accuracy = $\sum_{i} \{1 : p\_best_{i} = y_{i} , 0 : otherwise \}$

Benchmark Model

For benchmark model I will use a single convolution layer to extract features from the image and a softmax layer to predict the probabilities of different emotions. I will find the accuracy on the validation dataset and compare this accuracy result with the accuracy of the final model.

The human accuracy on this dataset is around 65.5% [9].So we will compare accuracy of our model to human accuracy and will ensure accuracy of our model is close to human accuracy.

Evaluation Metrics

I will use accuracy as the most important evaluation parameter on the dataset. Accuracy of the final model on the dataset will be compared to the benchmark model and human accuracy .

In addition to that Confusion matrix for true and prediction emotion counts will be generated . This matrix will be analysed to get idea of how the model is performing on different emotions.

We will also find precision value for each emotion to get some quantitative measure of how well our model is predicting each emotion in comparison to the benchmark model.

Project Design

In the first step I will analyse FERC-2013 dataset . I will also find some emotions which has very less training data and will merge with some other equivalent emotion.

In the next step I will create the benchmark model as mentioned above using keras deep learning framework with Tensorflow backend. Evaluation metrics for the benchmark model will be calculated and will be stored for comparison with the final model.

Our final model will be build using Multiple Convolutional,Pooling and feed forward layers. I will use data augmentation technique to generate more data.Then I will optimize number of various layers so that model can give human level of accuracy. Evaluation metrics for the final model will be calculated and will be compared with the benchmark model.

I will also analyse the result for various emotions and will find it's performance on each of them individually.

Once the model is finalized I will create a web app which can capture a image and send the image to server for processing . Using OpenCV I will process the image and face will be detected. Then I will pass this image of the face to our pre-trained Model to predict emotion. The final prediction will be displayed as emoji to the user. Probabilities of all other predictions will also be displayed to the user using bar chart.

References

[1] C. Shan, S. Gong, and P. W. McOwan, “Facial expression recognition based on Local Binary Patterns: A comprehensive study,” Image Vis. Comput., vol. 27, no. 6, pp. 803–816, May 2009.
[2] J. Whitehill and C. W. Omlin, “Haar features for FACS AU recognition,” in 7th International Conference on Automatic Face and Gesture Recognition (FGR06), 2006, p. 5 pp.-pp.101.

[3] S. Berretti, A. D. Bimbo, P. Pala, B. B. Amor, and M. Daoudi, “A Set of Selected SIFT Features for 3D Facial Expression Recognition,” in 2010 20th International Conference on Pattern Recognition (ICPR), 2010, pp. 4125–4128.

[4] C. Liu and H. Wechsler, “Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition,” IEEE Trans. Image Process., vol. 11, no. 4, pp. 467–476, Apr. 2002.

[5] Z. Wang and Z. Ying, “Facial Expression Recognition Based on Local Phase Quantization and Sparse Represe

[6] B.-K. Kim, J. Roh, S.-Y. Dong, and S.-Y. Lee, “Hierarchical committee of deep convolutional neural networks for robust facial expression recognition,” J. Multimodal User Interfaces, vol. 10, no. 2, pp. 173–189, Jan. 2016

[7] Kaggle. Challenges in representation learning:Facial expression recognition challenge ,2013

[8] P.Lucey, J.F Cohn, T. Kanade: The extended chn-kanade dataset(CK+):A complete dataset for action unit and emotion specified expression. Im Computer vision and pattern recognition workshop(CVPRW),2010 IEEE Computer society conference on page 94-101,IEEE 2010.

[9] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee, Y. Zhou, C. Ramaiah, F. Feng, R. Li, X. Wang, D. Athanasakis, J. Shawe-Taylor, M. Milakov, J. Park, R. Ionescu, M. Popescu, C. Grozea, J. Bergstra, J. Xie, L. Romaszko, B. Xu, Z. Chuang, and Y. Bengio, “Challenges in representation learning: A report on three machine learning contests,” Neural Networks, vol. 64, pp. 59–63, 2015.