Facial Expression Detection

AIM97

June 12, 2019

Contents

Chapter 1

CH1 - Introduction

1.1 Preface

1.1.1 Preface

<You expect employees to have high levels of emotional intelligence when interacting with customers. Now, thanks to advances in Deep Learning, you'll soon expect your software to do the same.>

<Companies have also been taking advantage of emotion recognition to drive business outcomes. For the upcoming release of Toy Story 5, Disney plans to use facial recognition to judge the emotional responses of the audience. Apple even released a new feature on the iPhone X called Animoji, where you can get a computer simulated emoji to mimic your facial expressions. It's not so far off to assume they'll use those capabilities in other applications soon.>

1.2 Image Processing

1.2.1 Brief

definition Image processing is the process of applying some operations to an image to reach an enhanced image that satisfies a certain goal depending on the application in hand. for example if we need to make an application that detects edges within an image we use an image processing technique that is capable of highlighting those edges and make them stand out. the result image is not necessarily a beautiful one from the perspective of a human, but it has to highlight the features of interest within the image that would be used for further processing.

impact Apart from the rule image processing plays in graphics enhancement to make image more visually appealing, Image processing is very important tool that is used for the specially preparation for computer vision and Machine learning, image processing a key preprocessing step to be taken before start in any of the two fields. the key difference between those two purposes is that when we want the image to be more visually appealing our target is a human, a human is the one who should view that image in the end. but when it comes to fields like computer vision or Machine Learning, the target is a computer that is programmed to act based on the content of input image. for this computer to do that it must be able to clearly extract feature of interest from the image, in order to make use of the image, we must have 2 main tasks for image processing:

1. noise removal: to remove the noise (like salt and pepper, or gauessian blur, ...etc) that we estimate to exist in the image so as to refine the features to be extracted from the image.
2. feature extraction: to highlight and evaluate features of interest that exist in the image to be used as input data for computer vision or Machine learning algorithms e.g. neural networks.

these two processes are the most common use cases for image processing, and we will go over them with more detail in next section.

1.2.2 How important is image processing?

The applications of image processing are many we will catch on some applications for noise removal and feature extraction.

Noise removal

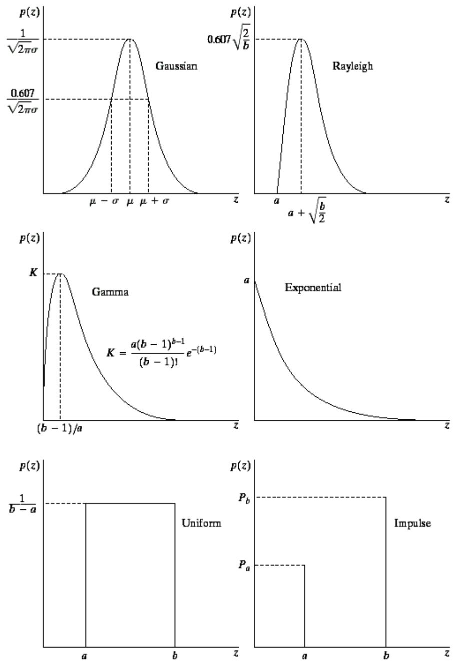
this process is done to make features more clear to refine the quality of extracted features by removing different types of noise (see figure 1.1).
multiple filters exists to restore original image by removing noise as much as possible.

1. Max filter: the output at one pixel is the **maximum** value of the pixels around it.
2. Min filter: the output at one pixel is the **minimum** value of the pixels around it.
3. Median filter: the output at one pixel is the **median** value of the pixels around it.
4. Mean filter: the output at one pixel is the **mean** value of the pixels around it.
5. Gauessian filter: it applies a matrix with values with gaussian distribution(highest weight in the center and weight decreases as we go away from the center) to the current window and the result is assigned to current pixel.

these filters make use of the values of pixels around current pixel in order to detect abnormal changes within the image which is probably noise and based on the values of surrounding pixel a new value is assigned to current pixel which is estimated to be closest to the original value.
Some other filter are used for enhancement can be sharpening filter and.

Feature extraction

this process is done to highlight features of interest in an image



1

Figure 1.1: examples for different types of noise

1.2.3 How is image processing important for our project

The purposes of our project is to recognize the facial expression from face image, for this task multiple image processing techniques have been applied to the input image before extracting features like face landmarks and HOG from the image.

Noise Removal

since we don't expect the input image to be particularly corrupted we only use median filter to remove white noise and salt and pepper noise a gaussian filter was being used at as well but it was inefficient in terms of time so the median filter took its place without a problem. we also use sharpening filter to make face features more clear for the landmark extraction process.

Feature Extraction

unlike CNN (convolutional neural network) model which extracts the features it needs from the image directly, one of the approaches we took requires pre-processing to extract some features from input image we needed two types of features in particular:

1. HOG (Histogram of Oriented Gradients) : feature descriptor which means that it generalize the object in a way that the same object (in this case a person) produces as close as possible to the same feature descriptor when viewed under different conditions.

2. Face Landmarks : those are points in the face that represent the face main features and we need those to estimate the emotion as well.

1.3 Neural Networks

1.3.1 Breif

What is Neural Networks? Artificial neural networks are one of the main tools used in machine learning. As the "neural" part of their name suggests, they are brain-inspired systems which are intended to replicate the way that we humans learn. Neural networks consist of input and output layers, as well as (in most cases) a hidden layer consisting of units that transform the input into something that the output layer can use. They are excellent tools for finding patterns which are far too complex or numerous for a human programmer to extract and teach the machine to recognize.

While neural networks (also called "perceptrons") have been around since the 1940s, it is only in the last several decades where they have become a major part of artificial intelligence. This is due to the arrival of a technique called "backpropagation," which allows networks to adjust their hidden layers of neurons in situations where the outcome doesn't match what the creator is hoping for — like a network designed to recognize dogs, which misidentifies a cat, for example.

Another important advance has been the arrival of deep learning neural networks, in which different layers of a multilayer network extract different features until it can recognize what it is looking for.

1.3.2 Basics of Neural Networks.

a basic idea of how a deep learning neural network learns, imagine a factory line. After the raw materials (the data set) are input, they are then passed down the conveyer belt, with each subsequent stop or layer extracting a different set of high-level features. If the network is intended to recognize an object, the first layer might analyze the brightness of its pixels. see figure 1.2

The next layer could then identify any edges in the image, based on lines of similar pixels. After this, another layer may recognize textures and shapes, and so on. By the time the fourth or fifth layer is reached, the deep learning net will have created complex feature detectors. It can figure out that certain image elements (such as a pair of eyes, a nose, and a mouth) are commonly found together.

Once this is done, the researchers who have trained the network can give labels to the output, and then use backpropagation to correct any mistakes which have been made. After a while, the network can carry out its own classification tasks without needing humans to help every time.

1

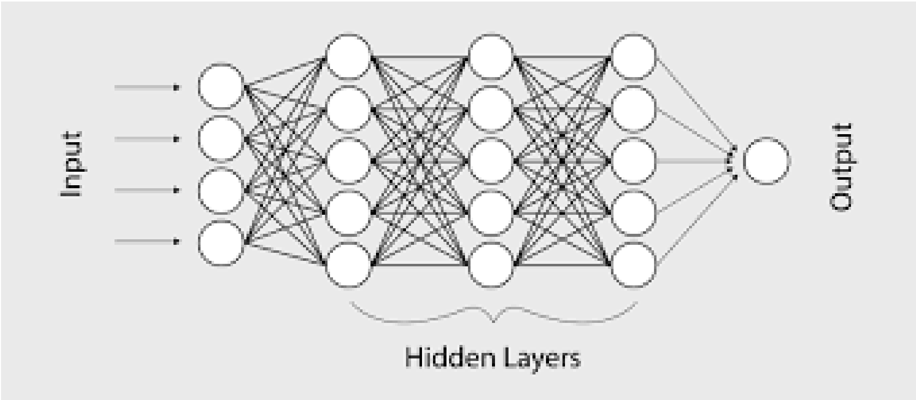


Figure 1.2: example of neural networks

Beyond this, there are different types of learning, such as supervised or unsupervised learning or reinforcement learning, in which the network learns for itself by trying to maximize its score

1.3.3 Why Neural networks are important?

ANNs (Artificial Neural Networks) have some key advantages that make them most suitable for certain problems and situations:

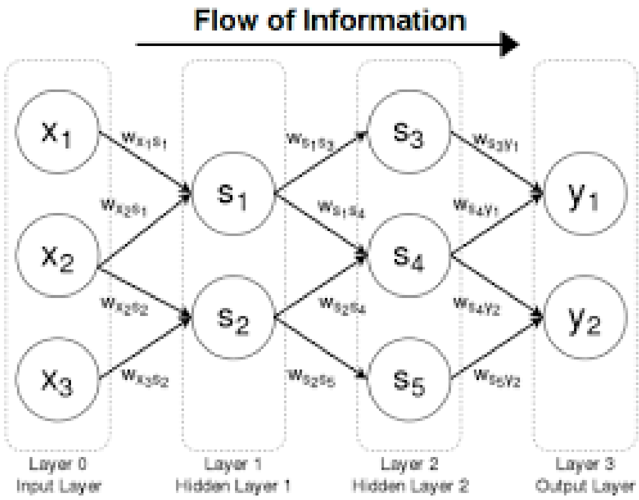
1. ANNs have the ability to learn and model non-linear and complex relationships, which is really important because in real-life, many of the relationships between inputs and outputs are non-linear as well as complex.
2. ANNs can generalize ,After learning from the initial inputs and their relationships, it can infer unseen relationships on unseen data as well,thus making the model generalize and predict on unseen data.
3. Unlike many other prediction techniques, ANN does not impose any restrictions on the input variables (like how they should be distributed). Additionally, many studies have shown that ANNs can better model heteroskedasticity i.e. data with high volatility and non-constant variance, given its ability to learn hidden relationships in the data without imposing any fixed relationships in the data.

1.3.4 Types of neural network

There are multiple types of neural network, each of which come with their own specific use cases and levels of complexity.

feedforward neural network

The most basic type of neural net is something called a feedforward neural network, in which information travels in only one direction from input to output.



1

Figure 1.3: feedforward neural networks

neural network layers are independent of each other; hence, a specific layer can have an arbitrary number of nodes. Typically, the number of hidden nodes must be greater than the number of input nodes. When the neural network is used as a function approximation, the network will generally have one input and one output node. When the neural network is used as a classifier, the input and output nodes will match the input features and output classes.

A neural network must have at least one hidden layer but can have as many as necessary. The bias nodes are always set equal to one. In analogy, the bias nodes are similar to the offset in linear regression i.e. y = mx+b. How does one select the proper number of nodes and hidden number of layers? This is the best part: there are really no rules! The modeler is free to use his or her best judgment on solving a specific problem. Experience has shown that there are best practices such as selecting an adequate number of hidden layers, activation functions, and training methods.

Recurrent Neural Network

A recurrent neural network (RNN) is a type of artificial neural network commonly used in speech recognition and natural language processing (NLP). RNNs are designed to recognize a data's sequential characteristics and use patterns to predict the next likely scenario.

Chapter 2

CH2 - Review

2.1 Preprocessing

2.1.1 Data sets

2.1.2 Face detection

hello face detection

2.1.3 Image enhancements

2.1.4 Landmark extraction

2.1.5 HOG

2.1.6 Data augmentation

2.1.7 Under sampling

it's a technique we can use when the data distribution is imbalanced, first we define what an "imbalance" is, data is called imbalanced if the output class values aren't evenly represented, for example let's say we study data with output feature having 2 possible values, either 0 or 1, the value 0 happens 80% of the time while the value 1 is represented by only 20% of the data, if we start training on this data right away we can expect the trained model to be biased towards the value 0 which appears to be the most common outcome, so even if he value 0 is not ha frequent, the model will assume so because that's what the data say.

the solution for this is to balance the data by removing some instances from the large class(s) until they are close enough to the smaller classes, this way we can decrease the bias of our model towards the large classes, this technique is called **under sampling**.

under sampling should solve the problem with bias in the trained model, but at he cost of decreasing the size of the data set when you drop large portion of it, hence decreasing what he model learns, so it can decrease the accuracy as well, if the biggest problem for your model is bias then under sampling can solve it, however if te biggest problem is in learning itself then under sampling shall make it worse.

one of the data sets we worked was imbalanced which made the trained model more biased towards negative emotions (which was more presented in the data set), so we tried to apply under sampling to solve this problem.

2.2 model

Chapter 3

CH3 - Proposed System

3.1 System Architecture

3.2 Preprocessing

3.3 Model

3.4 Library

3.5 Summary

Chapter 4

CH4 - Results

Chapter 5

CH5 - Conclusion

Chapter 6

References

1. Facial Expression Recognition using Convolutional Neural Networks: State of the Art

2. Y. Tang, “Deep Learning using Support Vector Machines,” in International Conference on Machine Learning (ICML) Workshops, 2013

3. B.-K. Kim, J. Roh, S.-Y. Dong, and S.-Y. Lee, “Hierarchical committee of deep convolutional neural networks for robust facial expression recognition,” Journal on Multimodal User Interfaces, pp. 1–17, 2016

4. Z. Yu and C. Zhang, “Image based static facial expression recognition with multiple deep network learning,” in ACM International Conference on Multimodal Interaction (MMI), 2015, pp. 435–442

5. A. Mollahosseini, D. Chan, and M. H. Mahoor, “Going Deeper in Facial Expression Recognition using Deep Neural Networks,” CoRR, vol. 1511, 2015.

6. Z. Zhang, P. Luo, C.-C. Loy, and X. Tang, “Learning Social Relation Traits from Face Images,” in Proc. IEEE Int. Conference on Computer Vision (ICCV), 2015, pp. 3631–3639

7. B.-K. Kim, S.-Y. Dong, J. Roh, G. Kim, and S.-Y. Lee, “Fusing Aligned and Non-Aligned Face Information for Automatic Affect Recognition in the Wild: A Deep Learning Approach,” in IEEE Conf. Computer Vision and Pattern Recognition (CVPR) Workshops, 2016, pp. 48–57.