

COLLEGE OF ARTS AND SCIENCES

**Machine Learning**

Facial emotion detection using deep learning

and Machine learning methods

April 2022

**Team members**

Fouzy Al Mozayen

Sara Baassiri

**Abstract**

What is machine learning and why is it critical in today’s world? Machine learning is a branch of artificial intelligence that allows applications to become more accurate at predicting outcomes by being programmed to do so. We use prebuilt data (datasets) to train a model based on it to be able to predict new output values. In this project, we apply examples of many machine learning methods using python in addition to a real-world example of how it works with emotion and face recognition. Face recognition has been around for a long time. Taking a step forward, human emotion, as expressed by the face and felt by the brain, can be mimicked using video, electric signal (EEG), or image form. The ability to identify human emotions is critical for modern artificial intelligence. Intelligent systems can mimic and assess facial emotions. This can assist you in deciding. Make well-informed decisions Whether it's to determine intent, promote offerings, or ensure security, threats. Recognizing emotions from photos or video is a simple operation for the human eye, but it proves to be difficult for computers. be difficult for machines to handle and necessitates a variety of image processing algorithms for feature extraction. Several machine learning algorithms are suitable for this job. Any detection or recognition by machine learning requires algorithm training and then testing on a suitable dataset.

**TABLE OF CONTENTS**

1. **Introduction.**
2. **Machine learning methods**
   1. Supervised learning
   2. Unsupervised learning
   3. Linear regression one variable
   4. Linear regression multi variables
   5. Neural networks (Explained in the emotion detection section)
   6. K means
3. **Introduction to emotion detection and face detection**
4. **Image Features**
   1. Landmarks
   2. Feature Descriptors
5. **Theoretical Background**
   1. Affective Computing
   2. Facial Emotion Recognition
   3. Machine Learning
   4. Artificial Neural Networks
   5. Deep Learning
      1. Use of GPU
   6. Convolutional Neural Networks
6. **Tools and libraries**
   1. Python3
   2. TensorFlow Keras
   3. OpenCV
   4. Eel (built with flask)
   5. Matplotlib
7. **Implementation**
   1. Downloading the dataset
   2. Setting the virtual environment
   3. Import libraries
   4. Writing the model training code
   5. Face detection
   6. Python OpenCV camera access
   7. Use of models with OpenCV
8. **Results.**
9. **Conclusion.**
10. **References.**

**Introduction**

ML is a field that focuses on the learning aspect of AI by developing algorithms that best represent a set of data. In contrast to classical programming, in which an algorithm can be explicitly coded using known features, ML uses subsets of data to generate an algorithm that may use novel or different combinations of features and weights than can be derived from predefined principles. In ML, there are four commonly used learning methods, each useful for solving different tasks: supervised, unsupervised, semi-supervised, and reinforcement learning. Many machine learning algorithms will be defined and demonstrated some of those methods are, supervised learning, unsupervised learning, Linear regression with one and multi variables, and K means. In addition to demonstrating logistic regression to be able to train a model for the emotion detection implementation.

**Machine Learning Methods**

**Linear regression single and multi-variable**

The most basic explanation of what linear regression is as follows “linear regression is a linear approach for modeling the relationship between a scalar response and one or more explanatory variables.” (google). Linear regression as the definition states could be used with a single or multi-input value. Single value linear regression methods are crucial tools to interpret one variable input (Continuous variable outcome). In addition, Multiple variable linear regression is a regression model that estimates the relationship between a quantitative dependent variable and two or more independent variables using a straight line. Examples will be given below.

**Code Example one variable:** Text

Description automatically generated

**Code Output:**

Text

Description automatically generated

**Code Example Multi-Variable:**

Text

Description automatically generated

**Code Output:**

Text

Description automatically generated with medium confidence

**Supervised Learning**

Supervised learning is defined as follows it is the machine learning task of learning a function that maps an input to an output based on example input-output pairs. It infers a function from labeled training data consisting of a set of training examples. Examples are given below:

**Code Example:**

Text

Description automatically generated

**Output:**

**Chart, scatter chart

Description automatically generated**

**Graphical user interface, text

Description automatically generated**

**Unsupervised learning**

Unsupervised learning is a type of algorithm that learns patterns from untagged data. The hope is that through mimicry, which is an important mode of learning in people, the machine is forced to build a compact internal representation of its world and then generate imaginative content from it.

**Code Example:**

**Text

Description automatically generated**

**Output:**

Chart, scatter chart

Description automatically generated

**K means clustering**

k-means clustering is a method of vector quantization, originally from signal processing, that aims to partition n observations into k clusters in which each observation belongs to the cluster with the nearest mean, serving as a prototype of the cluster.

Text

Description automatically generated  
**Output:**

Chart, line chart

Description automatically generated

Chart, scatter chart

Description automatically generated

Chart, scatter chart

Description automatically generated

**Emotion Detection Introduction**

Human emotion is implemented in many areas for example the requirement of additional security or information about the person. Emotion detection is seen as a second layer added to face detection as additional security. In addition, it can be used in the domain of business where the business can see what the customer’s response was to their promotion and whether they liked or disliked the product. Emotions can be classified into 7 categories which are angry, disgusted, fearful, happy, neutral, sad, and surprised. Facial muscle contortions are very minimal and detecting these differences can be very challenging as even a small difference results in different expressions and expressions of different or even the same people might vary for the same emotion, as emotions are hugely context dependent. Using deep learning, we generate a model that can identify emotions based on an accuracy percentage. Emotion detection can be split into simple steps:

1) Dataset preprocessing.

2) Face detection. (OpenCV)

3) Feature extraction.

4) Classification based on the features.

**Image Features**

We will look at some learning algorithms that can be used to determine facial features:

1. **Landmarks:**

Chart, scatter chart

Description automatically generated

Face landmarks are extremely important and can be utilized to identify and recognize faces. In the case of phrases, the same landmarks can be employed. A 68 facial landmark detector in the Dlib library determines the location of 68 landmarks on the face. The 68 landmarks are shown in the figure above. Those landmarks can be divided into several specific areas:

1. Right Eye
2. Left Eye
3. Right Eyebrow
4. Left Eyebrow
5. Mouth
6. Nose
7. Jaw
8. **Feature Descriptors**

Good features are those that aid in correctly recognizing objects. The corners and edges of the photos are usually used to identify them. The OpenCV library offers numerous feature detector methods for locating corners and edges in images, such as the Harris corner detector. Many other parameters, such as contours, hull, and convexity, are considered by these feature detectors. The feature detector method detects corner points or edges as key points. The key-surrounding point's area is described by the feature descriptor. The description might include anything from raw pixel intensities to surrounding region coordinates. A local feature is made up of the key point and the descriptor. A histogram of oriented gradients is an example of a feature descriptor.

**Theoretical Background**

A description of relevant topics for this project is offered in this section. This part tries to provide background information on the subjects that will be explored later in the report. A chronological review of disciplines such as affective computing and machine learning is used to provide this backdrop. To explain the principles, a top-down method will be used. In addition, related research to the approach employed in this study is discussed.

1. **Affective Computing**

Affective computing is the study and development of systems and devices that can recognize, interpret, process, and simulate human effects. It is an interdisciplinary field spanning computer science, psychology, and cognitive science.

A picture containing text

Description automatically generated

(The figure represents facial action units)

1. **Facial Emotion Recognition**

Facial Emotion Recognition is a technology used for analyzing sentiments from different sources, such as pictures and videos. It belongs to the family of technologies often referred to as ‘affective computing’, a multidisciplinary field of research on computers’ capabilities to recognize and interpret human emotions and affective states and it often builds on Artificial Intelligence technologies.

1. **Machine Learning**

Machine Learning (ML) is a subfield of Artificial Intelligence. A simple ML explanation is the one coined by Arthur Samuel in 1959: “... field of study that gives

computers the ability to learn without being explicitly programmed”. This statement

provides a powerful insight into the approach of this field. It completely differs from other fields where any new feature must be added by hand. For instance, in software development, when a new requirement appears, a programmer must create software to handle this new case. In ML, his is not exactly the case. The ML algorithms create models, based on input data. These models generate an output that is usually, a set of predictions or decisions. Then, when a new requirement appears, the model might be able to handle it or provide an answer without the need to adding new code. ML is usually divided into 3 broad categories. Each category focuses on how the learning process is executed by a learning system. These categories are supervised learning, unsupervised learning, and reinforcement learning. Supervised learning is when a model receives a set of labeled inputs, which means that they also contain the corresponding belonging class. The model tries to adapt itself in a way that can map every input with the corresponding output class. On the other hand, unsupervised learning receives a set of inputs without them being labeled. In that sense, the model tries to learn from the data by exploring patterns on them. Finally, reinforcement learning is when an agent is rewarded or punished accordingly for the decisions it took to achieve a goal.

1. **Deep Learning**

The latest reincarnation of ANN (Artificial Neural Networks) is known as Deep Learning (DL). According to Yann LeCun, this term designates “... any learning method that can train a system with more than 2 or 3 non-linear hidden layers.”. DL has achieved success in fields such as computer vision, natural language processing, and automatic speech recognition. One of the main strengths of using DL techniques is that there is no need for feature engineering. The algorithms can learn features by themselves over basic representations. For instance, on image recognition, an ANN can be fed with pixel representations of images. Then, the algorithm will determine if a certain pixel combination represents any feature, that is repeated through the image. As the data is processed through the layers, the features will go from very abstract forms to the meaningful representation of objects. DL started to become popular after some better than state-of-the-art results were achieved in several fields. For instance, the first paper containing information about a major industrial application was one related to automatic speech recognition. In that paper from 2012, ANN outperformed Gaussian mixture models in several benchmarking tests. This paper is a collaboration between four research groups: The University of Toronto, Microsoft Research, Google Research, and IBM Research. Two years later, another breakout publication was in the field of natural language processing. This research presented that Long-Short Term Memory (a particular ANN architecture called recurrent neural network specialized in sequences) provided better results than statistical machine translation, which was the default tool for translation at that time. This network was able to translate words and phrases from English to French. Finally, a deep learning technique that is relevant for this project is presented: Convolutional Neural Networks (CNN). A paper published in 2012 by a group of researchers from Toronto University showed results never achieved before in the ImageNet classification competition. This research has become a foundational work on DL. In the 2012 edition, its solution using deep CNN achieved an error rate of 15.3% on the top-5 classification while the second-best achieved 26.2%. More details about CNN are introduced in section 2.5 on page 13. In this research, two concepts, which the ML community has widely adopted, are stressed: the use of the rectified linear unit as the activation function and the use of GPU for training.

* **The use of GPU**

The use of GPU for training has become fundamental for training deep networks because of practical reasons. The main reason is the reduction of the training time compared to CPU training. While different speedups are reported depending on the network topology, it is common to have around 10 times the speed when using GPU.

1. **Convolutional Neural Networks**

A Convolutional Neural Network, also known as CNN or ConvNet, is a class of neural networks that specializes in processing data that has a grid-like topology, such as an image. A digital image is a binary representation of visual data.

Diagram

Description automatically generated

**Tools and libraries**

1. **Python3**

Python is a high-level, general-purpose programming language. Its design philosophy emphasizes code readability with the use of significant indentation. Its language constructs and object-oriented approach aim to help programmers write clear, logical code for small- and large-scale projects.

1. **TensorFlow Keras**

Keras is a deep learning API written in Python, running on top of the machine learning platform TensorFlow. It was developed with a focus on enabling fast experimentation. Being able to go from idea to result as fast as possible is key to doing good research.

1. **OpenCV**

OpenCV is a library of programming functions mainly aimed at real-time computer vision. Originally developed by Intel, it was later supported by Willow Garage and then Itseez, Inc. The library is cross-platform and free for use under the open-source Apache 2 License.

1. **Eel**

Eel is a small library in the Python programming language that allows programmers to create simple Electron-like offline Graphical User Interface (GUI) applications based on HTML and JavaScript with complete access to the capabilities and libraries of Python.

1. **Matplotlib**

Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits like Tkinter, wxPython, Qt, or GTK.

**Implementation Emotion detection**

**Text

Description automatically generated**

**Text

Description automatically generated**

**Results**

Graphical user interface

Description automatically generated

* Press Open Feed to run the camera and detect emotions
* Press Close Feed to stop the camera if it’s open
* Choose the Train option on the navigation bar to be able to go to the training and ML methods screen

Text

Description automatically generated

**Conclusion**

In conclusion, the aim was to demonstrate some machine learning methods and use these methods to create a real-life demonstration of what they could be used for, by training an emotion detection model to be able to detect the emotion on a person's face and check if it’s one of the seven categories that are angry, disgusted, fearful, happy, neutral, sad, and surprised. The project uses deep convolutional neural networks.

**References**