Facial Expression Recognition by Support Vector Machines

By:

Daryl Arredondo
Max Chester
Jiwon Choe
Kai He

Facial Expression Recognition by Support Vector Machines

By:
Daryl Arredondo
Max Chester
Jiwon Choe
Kai He

Online:
< http://cnx.org/content/col11391/1.1/ >

C O N N E X I O N S

Rice University, Houston, Texas

Collection structure revised: December 19, 2011
PDF generated: December 19, 2011
For copyright and attribution information for the modules contained in this collection, see p. 19.

Table of Contents

Chapter 1

The Team[1]

1.1 The Team

Jiwon Choe is a junior ECE student with focus on computer engineering. Daryl Arredondo, Kai He, and Max Chester are junior ECE students with focuses on signals and systems.

[1]This content is available online at <http://cnx.org/content/m41841/1.1/>.

Chapter 2

Introduction[1]

2.1 Introduction

"Reading" someone's face has long been a cheap parlor trick, whereby some claim that a glance at a person is enough for them to tell many things about that person, including their emotional state. With advances in modern computing and signal processing however, it may actually be that we can "train" any computer to more reliably and accurately detect emotions from facial images than a human can. Currently our program only trains a computer how to detect emotions based on images, whereas in the real world humans also analyze tone of voice, language content and body language, among other things to detect emotion. Considering that most current programs only look at facial expressions, there is clearly a large scope to expand the range of data computers analyze in order to produce even more accurate and reliable results.

[1]This content is available online at <http://cnx.org/content/m41855/1.1/>.

Chapter 3

Motivation and Applications[1]

3.1 Motivation and Applications

One of the main reasons we chose this project was because of the huge range of possible applications this has. One of the most clearly evident applications is in regards to artificial intelligence. The famous Turing Test marks the beginning of artificial intelligence as the point when a computer program having conversation with a human can fool the human into believing that the computer is a human. Currently this is done solely through text communication, but imagine how much more effective computers could be at emulating humans if they could take visual cues and use them to adjust how they respond to you. Instead of just responding directly to the user's questions, personal phones could judge the user's emotions and then formulate an appropriate response taking that into account. Advertisements could become even more personally targeted and effective as they could respond to emotional cues to decide for example whether or not to keep playing an advertisement or perhaps what type of advertisement to play. This type of technology could even play a crucial role in keeping us safe, as programs could monitor airline pilots looking for signs that their attentiveness was slipping or that they were falling asleep and once this crossed a certain threshold could provide a stimulus to the pilot to ensure full alertness and greatly reduce the risk of fatigue related accidents. In general, almost any type of interaction between a human and a computing device could benefit from this technology.

[1] This content is available online at <http://cnx.org/content/m41861/1.1/>.

Chapter 4

Implementation[1]

4.1 Implementation

We initially began to explore two different methods of detecting emotions from images of faces. The first, more traditional method takes the top down approach classifying all the different facial expressions associated with emotion and what combinations or patterns of these correspond to specific emotions. Then the program looks for these specific expressions and based on what it sees tries to match these expressions to an emotion based on the given classification system. One example of such a prominently used classification is the Facial Action Coding System (FACS) developed in the 70s to taxonomize human facial expressions. This has the advantages of being quick and relatively easy to implement, but it also suffers from a lack of robustness and has a hard time dealing with different types of faces. Many animators and others use this type of approach, but in the end we decided that for the purposes of our project and our goal of a broad based emotion detector this was not the method for us.

The second method, on the other hand, involves using one of the key capabilities that computers have that humans don't – the ability to quickly intake and analyze large amounts of data – and this was one of the main reasons that we chose it. While the first method involved first coming up with a classification system a priori and then applying it to faces, the second method first 'trains' the program with a large database of images, each image coming with an associated label that indicates what emotion the face is expressing, and then the program is ready to analyze new faces and detect emotions. This approach has the obvious advantage of not requiring any kind of 'given' knowledge or rules – simply working with the provided data. It is uniquely robust, because the algorithm can be continually improved by giving the program more training images and it can very easily be tailored to specific situations of lighting or setting by training it with images with those specific attributes. Furthermore, depending on what you train it with, it could be made to work best only with specific types or groups of people or with a broad range of people. Theoretically, given enough data and time, this type of program should be able to be much more accurate and robust than a human at detecting emotion because it utilizes one of the main strengths of computing relative to the human brain.

[1]This content is available online at <http://cnx.org/content/m41862/1.1/>.

Chapter 5

Algorithm[1]

5.1 Algorithm

This method falls under the broad category of machine learning techniques, but the core of the method we used relies on support vector machines (SVM). In general, a SVM takes a set of input data and classifies each data point into one of two general classes. Geometrically, a SVM can be thought of as constructing a plane dividing a region of space into two separate areas, thus new data will fall on one side or the other of this plane indicating its classification. The svm trainer takes a database of training images and creates an svm that can classify input data into one of two types. But we want to deal with more than two types of data, because there are obviously more than two different emotions so what we do is create multiple svms each one specific to a certain emotion and that classifies into either that emotion or not that emotion so then after running it through multiple svms one of them will come up yes and thus identify the emotion. But what if two emotions are identified by two or more different svms one might wonder? Well then one can run the svm predict and obtain probability estimates for each of the classes in a svm. Thus whichever one has the higher probability estimate is the most likely correct.

But there is another mathematical tool which can help improve the accuracy of our program. If we use kernel functions to map the general set of our problem to an inner product set, we can hope to turn the problem into one of linear classifications. For our project, we tried two different kernel functions. The first, the linear kernel, produced the better results for most datasets, while the Radial Basis Function (RBF) kernel didn't produce as accurate results in general. However, for the FEI database the RBF kernel did produce better results suggesting that more testing is necessary to determine what type of kernels are more appropriate for different types of datasets. This might be because the data in the FEI database is less "easy" though how to define this is difficult.

[1]This content is available online at <http://cnx.org/content/m41865/1.1/>.

Chapter 6

Limitations[1]

6.1 Limitations

One of the main issues associated with this method is finding a good enough dataset because the accuracy of the results is pretty directly tied to the size of the dataset, with more training images usually corresponding to increased accuracy. However, finding large datasets for these purposes proved a harder task than we had initially anticipated. Here we detail the different databases we used for this project:

1) FEI Face Database: This database was acquired from the Electrical Engineering Department of Centro Universitario da FEI located in Sao Paulo, Brazil. It contains 200 individuals, with each of the individuals showing two different emotions, happy and neutral, for a total of 400 images.
2) CMU Multi-PIE Face Database: This database was acquired from Carnegie Mellon University and we used a total of 904 pictures from this database. There were 500 pictures of the same 250 individuals showing both happy and neutral faces, along with 404 pictures of the same 202 individuals showing both disgusted and surprised faces.
3) Japanese Female Facial Expression (JAFFE) Database: This database was acquired online though the images were originally obtained at Kyushu University. It contains 213 total images, of 7 facial expressions posed by 10 different female Japanese models.

[1] This content is available online at <http://cnx.org/content/m41914/1.1/>.

Chapter 7

Results[1]

7.1 Results

From the below results, one can draw some strong conclusions about the utility of our method in different types of situations. When testing the JAFFE Dataset with all 7 emotions, we got strikingly different accuracies when testing with the same people we trained the program on then we tested with different people than the program had been trained on (76.81% to 36.59%). Both of those tests were run using a linear kernel; when we tested the JAFFE Dataset with pre-registered individuals and using a RBF kernel the accuracy dropped by almost 15%.

The next three charts show the results when we just testing with two different emotions on the various datasets. The most striking result here is the difference that the number of pictures you train the program with makes for the accuracy of the result. When testing the JAFFE dataset we trained it with only 8 images and got very low accuracy – 25%, but when testing the FEI dataset we were able to train it with 200 images and got very high accuracy – 80%. Thus the number of images that the program is trained with seems to have a big effect on the accuracy of the program, with more training images being better.

	JAFFE Dataset Testing of Pre-Registered Individuals							
	Linear Kernel: Overall Accuracy 76.81%							
	Predicted Label							
		Happy	Sad	Fear	Distress	Surprise	Anger	Neutral
	Happy	70	10	0	0	0	20	0
	Sad	0	70	0	0	0	30	0
Actual	Fear	0	0	66.6667	0	0	33.3333	0
Label	Distress	0	0	10	70	0	20	0
	Surprise	10	0	0	0	80	10	0
	Anger	0	0	0	0	0	100	0
	Neutral	0	0	0	0	0	20	80
		Percent Chance of Predicted Label						

Figure 7.1

[1] This content is available online at <http://cnx.org/content/m41875/1.1/>.

	JAFFE Dataset Testing of Pre-Registered Individuals							
	RBF Kernel: Overall Accuracy 62.32%							
	Predicted Label							
		Happy	Sad	Fear	Distress	Surprise	Anger	Neutral
	Happy	60	0	0	0	10	30	0
	Sad	0	50	0	0	0	50	0
Actual	Fear	0	0	55.5556	0	11.1111	33.3333	0
Label	Distress	0	0	10	30	0	60	0
	Surprise	10	0	0	0	80	10	0
	Anger	0	0	0	0	0	100	0
	Neutral	0	0	0	0	0	40	60
		Percent Chance of Predicted Label						

Figure 7.2

	JAFFE Dataset Testing of Random Individuals							
	Linear Kernel: Overall Accuracy 36.59%							
	Predicted Label							
		Happy	Sad	Fear	Distress	Surprise	Anger	Neutral
	Happy	16.6667	0	0	0	0	83.3333	0
	Sad	0	16.6667	0	0	16.6667	66.6667	0
Actual	Fear	0	0	100	0	0	0	0
Label	Distress	0	0	0	0	0	100	0
	Surprise	0	0	0	0	66.6667	33.3333	0
	Anger	0	50	0	0	33.3333	16.6667	0
	Neutral	0	0	0	0	50	16.6667	33.3333
		Percent Chance of Predicted Label						

Figure 7.3

JAFFE Dataset Testing Happy/Neutral ONLY				
Linear Kernel: Overall Accuracy 25%				
	Predicted Labels			
		Happy	Neutral	
Actual	Happy	0	100	
Labels	Neutral	50	50	
	Percent Chance of Predicted Label			

Figure 7.4

FEI Dataset Testing				
Linear Kernel: Overall Accuracy 80%				
	Predicted Labels			
		Happy	Neutral	
Actual	Happy	75	25	
Labels	Neutral	15	85	
	Percent Chance of Predicted Label			

Figure 7.5

Multi_Pie Dataset Testing				
Linear Kernel: Overall Accuracy 38.33%				
	Predicted Labels			
		Happy	Neutral	Surprise
Actual	Happy	60	25	15
Labels	Neutral	60	25	15
	Surprise	60	10	30
	Percent Chance of Predicted Label			

Figure 7.6

Chapter 8

Further Improvements and Conclusions[1]

8.1 Further Improvements and Conclusions

One straightforward conclusion we can draw from our project is that when doing emotion recognition with SVMs, larger training sets invariably improve the accuracy of the results. Nevertheless it's currently very hard to find large databases of suitable images, although doubtlessly as computers (and cameras) grow ever more ubiquitous this problem will fade over time. Additionally, it appears that this problem can be at least partially mitigated by using pre-registered users (training the program with images of the same people you test it on). This actually fits in very well with our initial motivation, because we feel like a lot of demand for this type of capability will be for use in personal computing devices of some type, which generally just have very few regular users. Devices could be automatically configured to take a certain number of pictures of a person when they register to use it, such that the device would have high accuracy in detecting that person's emotions.

The question of optimal kernel functions remains unresolved, however, because while the linear kernel worked better for our datasets, we did not rigorously test all the possibilities, and our range of datasets was quite limited. One possible future experiment or project could be to test the different kernels on a wide range of datasets and situations, while holding all other variables constant, to try and see what type of data each kernel works better on.

Lastly, we feel like this type of program has almost boundless potential since it can only get more accurate with increased computing power and larger datasets. One of the main things we are going to be asking of our computers and especially our artificial intelligence in the future is that it can appropriately interact with humans and respond to all of our needs. Being able to recognize human emotions is thus a vital step on the way to fully achieving this goal.

[1]This content is available online at <http://cnx.org/content/m41873/1.1/>.

Index of Keywords and Terms

Keywords are listed by the section with that keyword (page numbers are in parentheses). Keywords do not necessarily appear in the text of the page. They are merely associated with that section. *Ex.* apples, § 1.1 (1) **Terms** are referenced by the page they appear on. *Ex.* apples, 1

Attributions

Collection: *Facial Expression Recognition by Support Vector Machines*
Edited by: Daryl Arredondo, Max Chester, Jiwon Choe, Kai He
URL: http://cnx.org/content/col11391/1.1/

Module: "The Team"
By: Daryl Arredondo, Jiwon Choe, Max Chester, Kai He
URL: http://cnx.org/content/m41841/1.1/
Page: 1

Module: "Introduction"
By: Daryl Arredondo, Kai He, Max Chester, Jiwon Choe
URL: http://cnx.org/content/m41855/1.1/
Page: 3

Module: "Motivation and Applications"
By: Daryl Arredondo, Jiwon Choe, Max Chester, Kai He
URL: http://cnx.org/content/m41861/1.1/
Page: 5

Module: "Implementation"
By: Daryl Arredondo, Max Chester, Jiwon Choe, Kai He
URL: http://cnx.org/content/m41862/1.1/
Page: 7

Module: "Algorithm"
By: Daryl Arredondo, Jiwon Choe, Max Chester, Kai He
URL: http://cnx.org/content/m41865/1.1/
Page: 9

Module: "Limitations"
By: Daryl Arredondo, Jiwon Choe, Max Chester, Kai He
URL: http://cnx.org/content/m41914/1.1/
Page: 11

Module: "Results"
By: Daryl Arredondo
URL: http://cnx.org/content/m41875/1.1/
Pages: 13-16
Copyright: Daryl Arredondo
License: http://creativecommons.org/licenses/by/3.0/

Module: "Further Improvements and Conclusions"
By: Daryl Arredondo, Kai He, Jiwon Choe, Max Chester
URL: http://cnx.org/content/m41873/1.1/
Page: 17
Copyright: Daryl Arredondo, Kai He, Jiwon Choe, Max Chester
License: http://creativecommons.org/licenses/by/3.0/

Facial Expression Recognition by Support Vector Machines
Report of an ELEC 301 Project on Facial Expression Recognition by Support Vector Machines. This collection will walk you through an introduction of the team and motivation behind the project. It will give you the implementation, the algorithm used, limitations, results and conclusions of this project.