Grammatical Facial Expression Recognition with Supervised Classifiers

Yulun (Allen) Zeng , Songzhu (Sean) Zheng

Department of Statistics

Nov. 2016

Outline

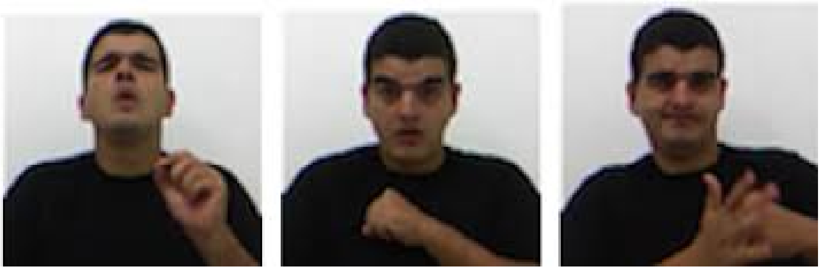
What is a GFE

Grammatical facial expression (GFE) is a kind of sign language widely used by deaf people to communicate.

These languages consist of facial expressions, head movements, and hence are intrinsically multimodal languages

[0] Fernando de Almeida Freitas, et al. (2014)

GFE



[0] Fernando de Almeida Freitas, et al. (2014)

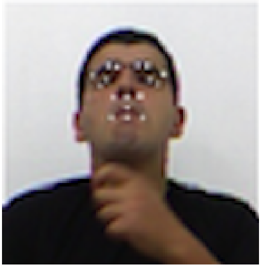
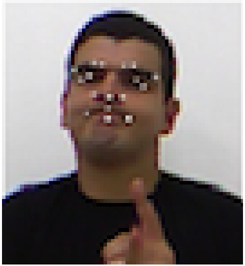
Data Set

- Data set comes from UCI machine learning repository

- The data set is composed by eighteen videos recorded using Microsoft Kinect sensor

- In each video, a user performs (five times), in front of the sensor, five sentences in Libras (Brazilian Sign Language) that require the use of a grammatical facial expression

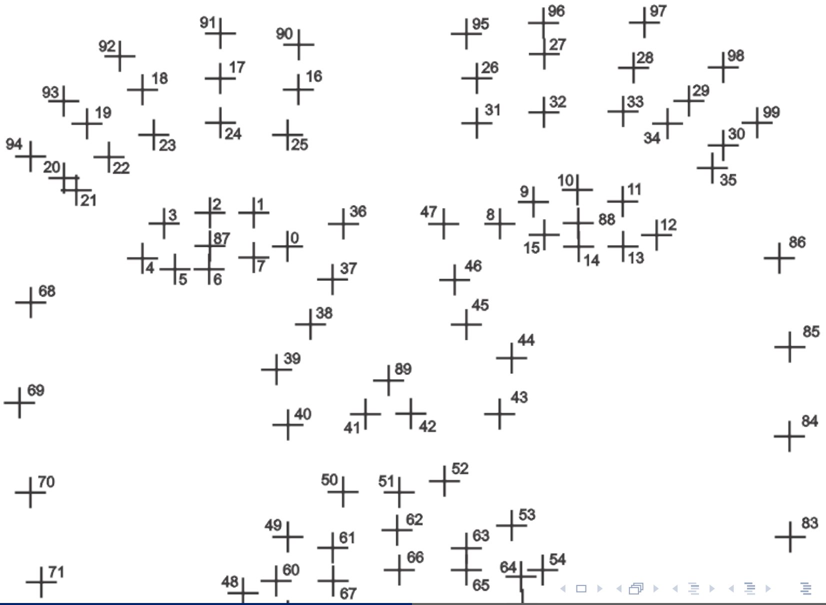
Data Set

Data Set - Kinect Sensor Record

- There are 27 frames captured by the kinect each second
- Each frame contains a picture that includes 100 coordinates(x,y,z) of points from eye, nose, eyebows, face contour and iris
- Each frame contains one facial expression that might or might not be a valid GFE, and are labelled manually by specialists

Points Recorded

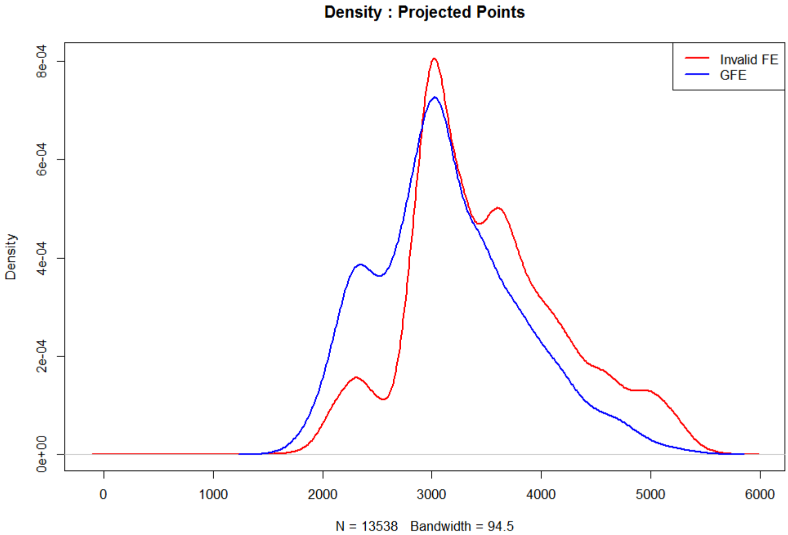


Final Data Set

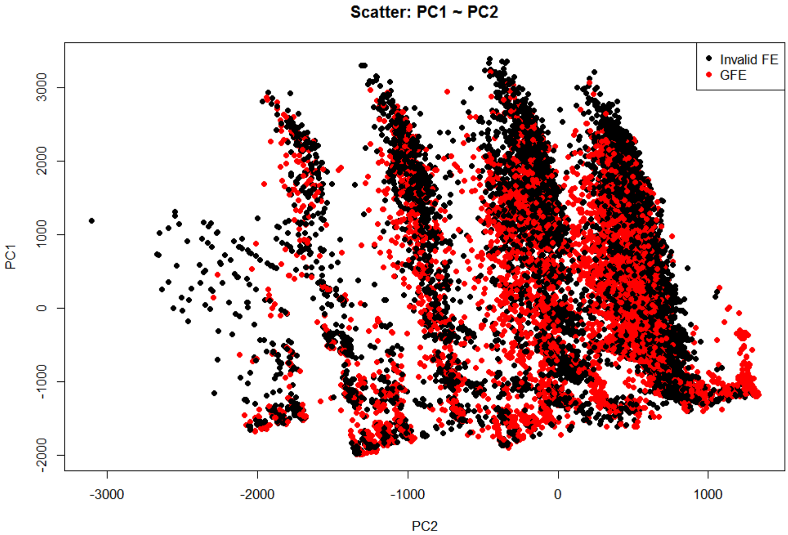
Collapse each frame to a one line vector and attach the label to it. We finally have a 27936×301 matrix.

20951 (75%) of them are used for model training, the rest 6985 are used for testing.

Projecting the Data onto the Mean Difference



First Two PC

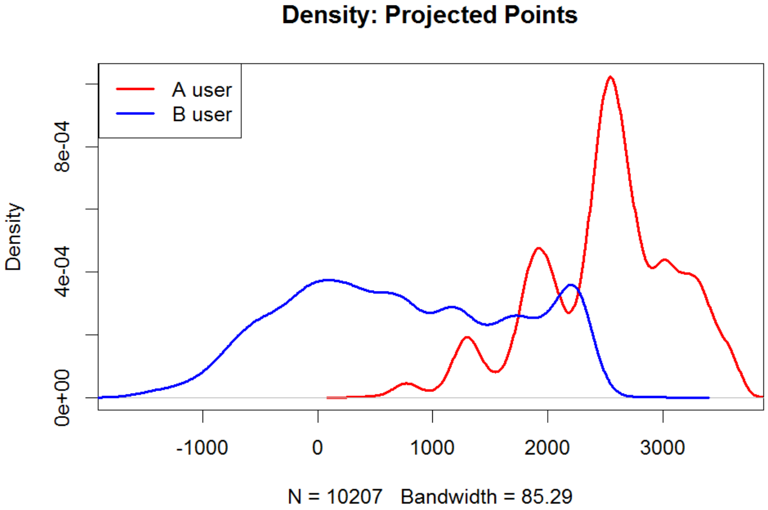

Scatter: PC1 ~ PC2
Invalid FE
GFE
PC1
PC2
3000
2000
1000
0
-1000
-2000
-3000
-2000
-1000
0
1000

Model and Results Comparison

Table: Classifier Comparison

Classifier	Accuracy	Parameter	Training Time
LDA	0.82	Moment Way	14.34s
QDA	0.76	Moment Way	4.33s
Logistic	0.83	$-$	26.98s
Logistic LASSO	0.80	$\lambda = 0.00026$	2889.742s
NaiveBayes	0.58	$laplace = 0$	64.73
SVM	0.86	radical	580.03s
ada	0.86	exponential	1518.81s

Projecting onto User a and b



Model and Results Comparison

Table: Classifier Comparison

Classifier	Accuracy	Parameter	Training Time
LDA	0.9996	Moment Way	14.58s
QDA	0.9757	Moment Way	4.36s
Logistic	0.9997	–	82.38s
Logistic LASSO	1	$\lambda = 0.000226$	
NaiveBayes	0.9933	$laplace = 0$	63.51s
SVM	0.9996	radical kernel	42.14s
ada	0.86	exponential/100 iter	1548.80s

Model and Results Comparison

Table: Classifier Comparison

Classifier	Accuracy	Parameter	Training Time
SVM	0.8802	default setting	469.017s
CART	0.6062	$xval = 5, cp = 0.01$	27.870s
Random Forest	0.9756	50 trees	7.382s
Gradient Boosting	0.9682	$depth = 2, \eta = 1$	19.82s
Multinomial Logistic	–	–	∞

tsne

The scatter at the very begging is a visualization of these high dimensional data set in the 2-D plane through t-stochastic neighbour embedding.

Idea behind tsne

We try to properly reflect the euclidean distance structure of the high dimensional data set, by finding it's 2-D representation through minimize the KL divergence of them two.

t-sne Assumption

Given a point i in the data matrix, the probability that the data point j is in the same cluster as i does is given by:

$$p_{j|i} = \frac{\exp\left(-||x_i - x_j||^2/\sigma_i^2\right)}{\sum_{k\neq i}\exp\left(-||x_i - x_k||^2/\sigma_i^2\right)}$$

We assume a similar structure in the lower dimension:

$$q_{j|i} = \frac{\exp\left(-||y_i - y_j||^2/\sigma_i^2\right)}{\sum_{k\neq i}\exp\left(-||y_i - y_k||^2/\sigma_i^2\right)}$$

y is the lower-dimension representation of the higher dimensional data-points.

t-sne

We wan to find y_is, that minimize the KL divergence of these two density:

$$C = \sum_i KL(P_i||Q_i) = \sum_i \sum_j p_{j|i} \log \left\{ \frac{p_{j|i}}{q_{j|i}} \right\}$$

More Demonstration

Reference

[1] FREITAS, F. A. ; Peres, S. M. ; Lima, C. A. M. ; BARBOSA, F. V. . Grammatical Facial Expressions Recognition with Machine Learning. In: 27th Florida Artificial Intelligence Research Society Conference (FLAIRS), 2014, Pensacola Beach. Proceedings of the 27th Florida Artificial Intelligence Research Society Conference (FLAIRS). Palo Alto: The AAAI Press, 2014. p. 180-185.

[2] Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.