Recognizing Facial Expression: Machine Learning and Application to Spontaneous Behaviour

|  |  |
| --- | --- |
| Marian Stewart Bartlett  https://inc.ucsd.edu/~marni/  Gwen Littlewort  https://www.researchgate.net/profile/Gwen\_Littlewort | Department of Cognitive Science and Psychology, University of California  University of California, San Diego |

We present a systematic comparison of machine learning methods applied to the problem of fully automatic recognition of facial expressions. We report results on a series of experiments comparing recognition engines, including AdaBoost, support vector machines, linear discriminant analysis. We also explored feature selection techniques, including the use of AdaBoost for feature selection prior to classification by SVM or LDA. Best results were obtained by selecting a subset of Gabor filters using AdaBoost followed by classification with Support Vector Machines. The system operates in real-time, and obtained 93% correct generalization to novel subjects for a 7-way forced choice on the Cohn-Kanade expression dataset. The outputs of the classifiers change smoothly as a function of time and thus can be used to measure facial expression dynamics. We applied the system to fully automated recognition of facial actions (FACS). The present system classifies 17 action units, whether they occur singly or in combination with other actions, with a mean accuracy of 94.8%. We present preliminary results for applying this system to spontaneous facial expressions.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Kernel | Adaboost | SVM | AdaSVM | LDApca |
| Linear | 90.1 | 88.0 | 93.3 | 80.7 |
| RBF |  | 89.1 | 93.3 |  |

Table 1. Leave-one-out generalization performance of Adaboost, SVM and AdaSVM’s. AdaSVM: Feature selection by AdaBoost followed by classification with SVM’s.

|  |  |  |
| --- | --- | --- |
| Feature selection | LDA | SVM (linear) |
| None | 44.4 | 88.0 |
| PCA | 80.7 | 75.5 |
| Adaboost | 88.2 | 93.3 |

Table 2. Comparing SVM performance to LDA with different feature selection techniques. The two classifiers are compared with no feature selection, with feature selection by PCA, and feature selection by Adaboost.

The facial action coding system (FACS) is the most objective and comprehensive coding system in the behavioural sciences. A human coder decomposes facial expressions in terms of 46 component movements, which roughly correspond to the 44 facial muscles.

The system was trained on FACS-coded images from 2 datasets. The first dataset was the Cohn Kanade dataset, which contains FACS scores by two certified FACS coders in addition to the basic emotion labels. The second dataset consisted of directed facial actions collected by Hager and Ekman.

The combined dataset contained 2568 training examples from 119 subjects. As above, the system was fully automated. Automatic eye detection was employed to align the eyes in each image. Images were scaled to 192x192, passed through a bank of Gabor filters at 8 orientations and 7 spatial frequencies (4:32 pixels per Cyc). Output magnitudes were then passed to nonlinear support vector machines using RBF kernels. No feature selection was performed, although we plan to evaluate feature selection by AdaBoost in the near future.

We first report performance for generalization to novel subjects within the Cohn-Kanade and Ekman-Hager databases. Generalization to new subjects was tested using leave-one-subject-out cross-validation. The results are shown in Table 3. All system outputs above threshold were treated as detections. Performance was evaluated for thresholds of 0 in the SVM, and then evaluated again for the optimal threshold that maximized percent correct.

The system obtained a mean of 94.8% agreement with human FACS labels. System outputs for full image sequences of test subjects are shown in Figure 1. Although each individual image is separately processed and classified, the outputs change smoothly as a function of expression magnitude in the successive frames of each sequence, enabling applications for measuring the magnitude and dynamics of facial expressions.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| AU | Name | N | P | P(opt) | FA | Hit |
| 1 | Inn. Brow raise | 409 | 90.3 | 92.9 | 0.4 | 71.3 |
| 2 | Out. brow raise | 315 | 91.8 | 92.8 | 1.6 | 62.6 |
| 4 | Brow lower | 412 | 82.7 | 86.8 | 6.9 | 41.0 |
| 5 | Upper lid raises | 286 | 91.2 | 92.9 | 2.1 | 61.9 |
| 6 | Cheek raises | 278 | 92.8 | 93.5 | 1.4 | 70.1 |
| 7 | Lower lid tight | 403 | 85.7 | 88.5 | 4.6 | 52.1 |
| 9 | Nose wrinkle | 68 | 98.7 | 98.8 | 0.04 | 85.3 |
| 10 | Lip raises | 50 | 97.7 | 98.1 | 13.9 | 26.0 |
| 12 | Lip crnr. pull | 196 | 97.8 | 98.0 | 0.04 | 93.4 |
| 15 | Lip crnr. Depr. | 100 | 97.0 | 97.2 | 1.0 | 72.0 |
| 17 | Chin raise | 203 | 87.0 | 92.8 | 7.0 | 40.4 |
| 20 | Lip Stretch | 99 | 94.4 | 96.2 | 6.6 | 41.4 |
| 23 | Lip tightens | 57 | 97.0 | 97.9 | 11.0 | 36.8 |
| 24 | Lip press | 49 | 98.4 | 98.5 | 1.7 | 61.2 |
| 25 | Lips part | 376 | 89.7 | 91.2 | 2.2 | 64.9 |
| 26 | Jaw drop | 86 | 96.7 | 97.1 | 5.9 | 45.3 |
| 27 | Mouth stretch | 81 | 99.2 | 99.2 | 0.04 | 97.5 |
|  | Mean |  | 93.4 | 94.8 | 3.9 | 60.2 |

We presented a systematic comparison of machine learning methods applied to the problem of fully automatic recognition of facial expressions, including AdaBoost, support vector machines, and linear discriminant analysis, as well as feature selection methods.

The machine-learning based system presented here can be applied to recognition of any facial expression dimension given a training dataset. The system was able to detect facial actions in this database despite the presence of speech, out-of-plane head movements that occur during discourse, and the fact that many of the action units occurred in combination.

[1] M.S. Bartlett, G. Littlewort, B. Braathen, T.J. Sejnowski, and J.R. Movellan. A prototype for automatic recognition of spontaneous facial actions. In S. Becker, S. Thrun, and K. Obermayer, editors, Advances in Neural Information Processing Systems, volume 15, pages 1271–1278, Cambridge, MA, 2003. MIT Press.

[2] I. Cohen, N. Sebe, F. Cozman, M. Cirelo, and T. Huang. Learning baysian network classifiers for facial expression recognition using both labeled and unlabeled data. Computer Vision and Pattern Recognition., 2003.

[3] G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. Classifying facial actions. IEEE Transactions on Pattern Analysis and Machine Intelligence, 21(10):974–989, 1999.

[4] P. Ekman and W. Friesen. Facial Action Coding System: A Technique for the Measurement of Facial Movement. Consulting Psychologists Press, Palo Alto, CA, 1978.