Transformation Based Feature Selection for Human Emotion

Sarah Brown[1,2], S.R. Prakash[1], Andrea Webb[1], Jennifer Dy[2], Deniz Erdogmus[2]

[1] Fusion, Exploitation, and Inference Technologies Group - The Charles Stark Draper Laboratory, [2]Electrical and Computer Engineering – Northeastern University

Abstract

We pose feature selection as a feature transformation problem by solving an optimization problem for a projection-based criterion with a sparsity inducing norm penalty term. This allows us to solve a convex optimization problem to reduce the dimensionality of the data while maintaining the meaning of the original features. Through the optimization, the high dimensional feature set is projected into a lower dimensional space that optimizes the given criterion, while simultaneously forcing whole rows of the projection matrix to zero, resulting in a projection based on a subset of the feature set.

We apply this to data from a psychophysiological experiment designed to evaluate physiological responses to emotionally evocative stimuli. Physiological responses were recorded from a variety of sensor modalities including electrocardiogram (ECG), respiration, electrodermal activity (EDA), finger pulse (FP), body movement, and pupil diameter. The resultant feature subset both lies in a reduced dimension space and provides insight to understanding the autonomic mechanisms of human emotion.

Technical Approach

Goals:
- Reduce dimensionality
- Maintain feature meaning
- Learn relationships among features
- Retain label information in features
- Eliminate redundant, noisy features

Hilbert-Schmidt Independence Criterion measures independence using the Hilbert-Schmidt Norm of a cross-covariance operator in a Reproducing Kernel Hilbert Space.

Probabilities are unknown, so the empirical estimation is as below, for data X and labels Y.

$$HSIC(X,Y) = \frac{1}{n^2} tr(KHLH)$$

\# of samples

$$K,H,L \in \boldsymbol{R}^n$$

Gaussian Kernel in the transformed feature space

Feature vector for subject i

$$K_{ij} = k(x_i, x_j) = e^{\frac{1}{2}\sum_k^d w_k^T (x_i - x_j)(x_i - x_j)^T w_k}$$

Kernel Function on the labels

$$L_{ij} = l(y_i, y_j) = \delta(y_i, y_j) = \begin{cases} 1 & y_i = y_j \\ 0 & y_i \neq y_j \end{cases}$$

Centering Matrix

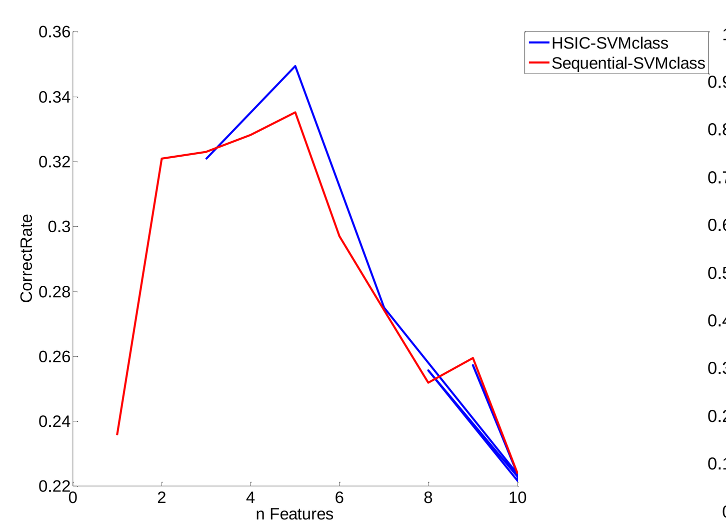
$$H = I - \frac{1}{n} 1_n 1_n^T$$

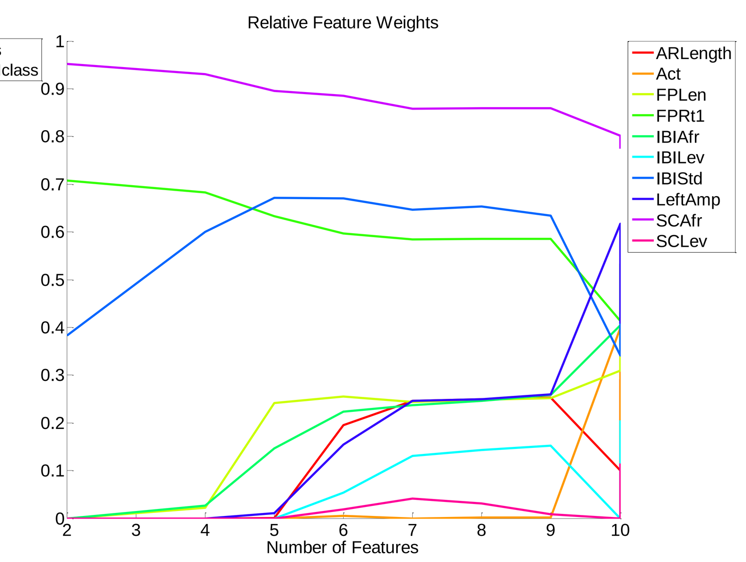
Data Collection

- 32 subjects were each presented with 24 sound and image stimuli
- Stimuli are expected to evoke one of five emotions: Amusement, Fear, Disgust, Sadness, or Anger
- Physiology was collected during the entire experiment
- Current analysis uses features extracted with CPSLAB, averaged across stimulus class and standardized within subject

Features	
ARLength	Abdominal respiration line length for a window beginning at stimulus onset
Act	Gross body movement line length for a window beginning at stimulus onset
FPLen	Line length of the finger pulse signal for a window beginning at stimulus onset
FPRt1	Finger pulse response from the first low point for a window beginning at stimulus onset
IBIAfr	Interbeat interval (derived from ECG) area to full recovery for a window beginning at stimulus onset
IBILev	Interbeat interval (derived from ECG) level for a window beginning at stimulus onset
IBIStd	Standard deviation of Interbeat interval (derived from ECG) for a window beginning at stimulus onset
LeftAmp	Left pupil diameter peak amplitude
SCAfr	Area to full recovery for the electrodermal activity signal for a window beginning at stimulus onset
SCLev	Level of the electrodermal activity signal for a window beginning at stimulus onset

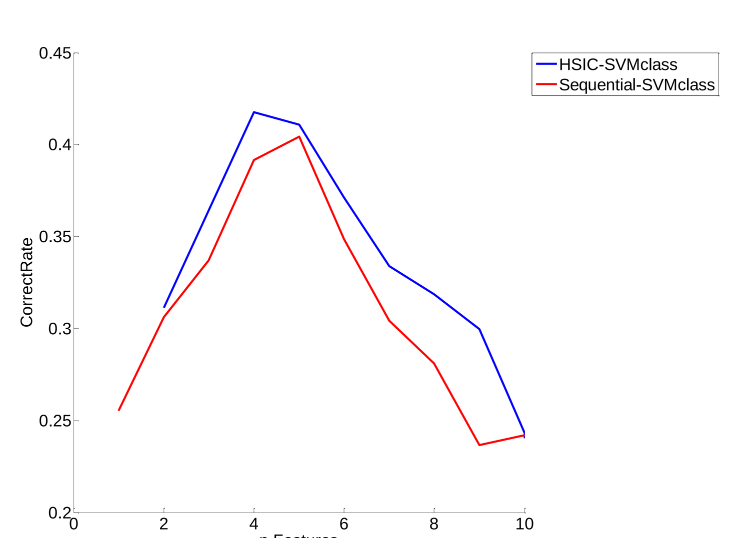
Results- Sounds

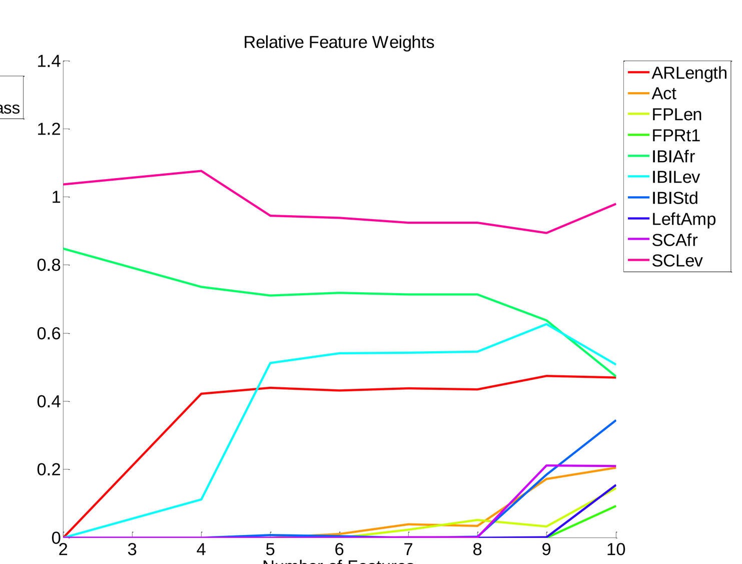




Transform Based Feature Selection				Sequential Forward Search			
Selected Features	Correct Rate	Sensitivity	Specificity	Selected Features	Correct Rate	Sensitivity	Specificity
[4 9]	0.31313	0.193	0.8665	[4 7]	0.31487	0.43033	0.78675
[4 7 9]	0.36247	0.372	0.79383	[4 5 9]	0.36833	0.51833	0.75108
[3 4 7 9]	0.376	0.32633	0.78017	[4 5 7 9]	0.37947	0.54333	0.76933
[3 4 5 7 9]	0.41907	0.39767	0.78275	[4 5 7 8 9]	0.36393	0.424	0.77133
[3 4 5 7 8 9]	0.37013	0.32967	0.77042	[2 4 5 7 8 9]	0.31947	0.35867	0.79325
[1 3 4 5 7 8 9]	0.35353	0.40633	0.82733	[2 3 4 5 7 8 9]	0.3028	0.25567	0.74483
[1 3 4 5 6 7 8 9]	0.3536	0.392	0.83458	[2 3 4 5 6 7 8 9]	0.31413	0.32933	0.74817
[1 3 4 5 6 7 8 9 10]	0.34953	0.33033	0.82633	[1 2 3 4 5 6 7 8 9]	0.3262	0.38567	0.81658
[1 2 3 4 5 6 7 8 9 10]	0.30487	0.30733	0.79175	[1 2 3 4 5 6 7 8 9 10]	0.3044	0.306	0.79067

Results- Images





Transform Based Feature Selection				Sequential Forward Search			
Selected Features	Correct Rate	Sensitivity	Specificity	Selected Features	Correct Rate	Sensitivity	Specificity
[5 10]	0.39828	0.33207	0.89647	[5 6]	0.31531	0.51966	0.84845
[1 5 6 10]	0.4669	0.58517	0.87853	[3 5 9]	0.34276	0.50793	0.80957
[1 5 6 7 10]	0.43655	0.4931	0.89138	[5 6 7 10]	0.45476	0.47034	0.91897
[1 2 5 6 7 10]	0.38952	0.53414	0.80276	[3 5 6 7 10]	0.44034	0.46172	0.87103
[1 2 3 5 6 9 10]	0.39062	0.52655	0.79147	[3 4 5 6 7 10]	0.36048	0.42138	0.84983
[1 2 3 5 6 7 9 10]	0.38641	0.52103	0.73672	[2 3 4 5 6 7 10]	0.34669	0.51931	0.79121
[1 2 3 5 6 7 8 9 10]	0.36959	0.51103	0.7206	[1 2 3 4 5 6 7 10]	0.36703	0.56138	0.75431
[1 2 3 4 5 6 7 8 9 10]	0.3291	0.50448	0.71698	[2 3 4 5 6 7 8 9 10]	0.35552	0.53069	0.75966
[1 2 3 4 5 6 7 8 9 10]	0.32621	0.50862	0.71448	[1 2 3 4 5 6 7 8 9 10]	0.33131	0.50586	0.7156

Formulation

Minimize the independence between the projected features XW and labels, Y

The parameter controls how sparse the resulting matrix will be, and how many features are selected

$$\min_{W \in \mathbb{R}^{q \times d}} -HSIC\,(XW, Y) + \lambda \sum_{j=1}^{d} \|w_j\|_\infty$$

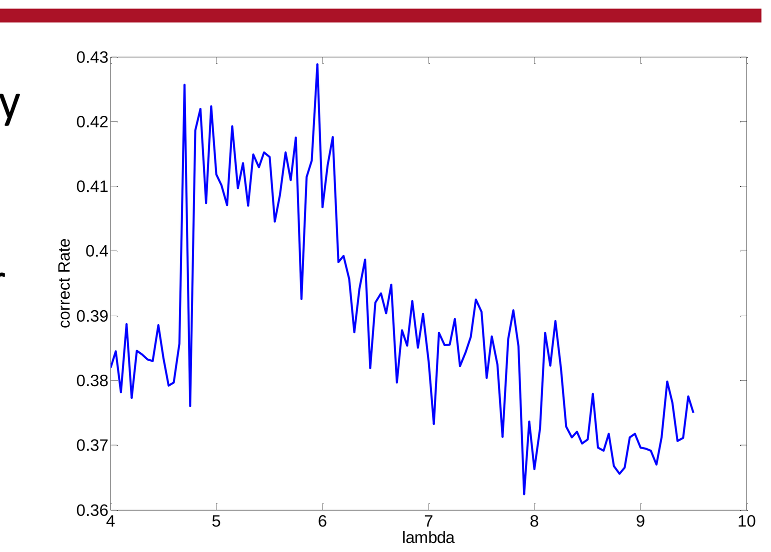
The L_1 -L_∞ norm on the rows of the projection matrix induces row-sparsity

This is solved as a constrained optimization problem

$$\min_{W \in \mathbb{R}^{q \times d}} -HSIC\,(XW, Y)$$
$$s.t. \|w_j\|_\infty \leq \alpha \;\; \forall j \in [1, d]$$

Future Work

- Exploration of column sparsity
- Sensitivity analysis in optimization
- Further analysis of parameter selection



References

Mahdokht Masaeli, Glenn Fung, and JenniferG. Dy. “From Transformation-based Dimensionality Reduction to Feature Felection”. In ICML, pages 751–758, 2010.

Acknowledgements

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-0946746.

DRAPER LABORATORY