CS224S Assignment 4

Milind Ganjoo
mganjoo@cs.stanford.edu

Stephanie Lynne Pancoast
pancoast@stanford.edu

Sebastian Schuster
sebschu@stanford.edu

1 Sample Analysis

We plotted the sample audio files in Praat. The following paragraphs summarize notable observations for each file (plots not included).

Anger The plot of the anger audio sample shows that the speaker's angry voice has a high pitch range, generally a high pitch and very high energy. Also some consonants are pronounced very long and generally all words are pronounced very clearly.

Despair Despair speech on the other hand is characterized by a low pitch, a medium pitch range and medium energy. The pitch also goes down at the end of each word.

Happy In case of happy speech, we also have a medium pitch range and the pitch also goes down at the end of words. The pace is very fast and there are hardly any breaks between the words.

Neutral The neutral sample has low pitch and a low pitch range and medium energy. It has also a rather fast pace and almost no breaks between syllables or words.

Sadness Our example for sad speech is characterized by very low pitch, a very low pitch range, very low energy and a low pace.

2 Experimenting with the classifier and parameters

This section describes our experiments with different classification algorithms and their parameters. We first describe the various experiments performed, and subsequently summarize the results and observations.

L_1 regularization By default, LIBLINEAR uses a L_2-regularized L_2-loss SVM classification algorithm. With the aim of automatically selecting relevant features, we used an L_1-regularized SVM instead, which is able to assign zero weights to irrelevant features in large data sets.

Logistic regression Our data set potentially contains features that contain no discriminatory information; therefore, we also tried a L_1 regularized logistic regression classifier.

Normalizing features We experimented with normalizing the feature values using the following transformation for each feature value f_i:

$$f_i' = \frac{f_i - \mu_i}{\sigma_i}.$$

Neural network-based softmax classifier In contrast to the closed-form logistic regression done in LIBLINEAR, we built a simple 1-layer, L_2 regularized neural network-based softmax classifier (384 input nodes, 5 outputs), and optimized weights using LBFGS. (Note that L_1 regularization would not be useful, since the optimization objective may not be differentiable at all points).

Regularization strength Finally, we also tried different values for the regularization strength C in both the LIBLINEAR and neural network-based implementations.

The results are presented in Table 1. This table shows that using L_1 regularization significantly improves the classification accuracy. Using normalized features further improves classification accuracy

Classifier	Raw	Normalized
SVM w/ L_2-reg	0.30	0.49
SVM w/ L_1-reg	0.44	0.53
LogReg w/ L_1-reg	0.53	0.52
Softmax NN w/ L_2-reg	—	0.61

Table 1: Accuracy for different classifiers using raw and normalized features. For all experiments we used a regularization strength $C = 1$.

(with the exception of logistic regression where we see a very small drop in accuracy).

The neural network classifier achieved an accuracy of 60.61% with a regularization parameter $C = 0.1$, but the best accuracy of 62% was obtained using the $L-1$ regularized logistic regression classifier with normalized features and inverse regularization strength $C = 0.09$.

3 Feature Analysis

The logistic regression classifier we used outputs weights for all of the features. Because we used a L1-loss function, many of these are zeros. More specifically, out of the 384 possible features, only 58, 65, 63, 42, and 63 were used for anger, despair, happy, neutral, and sadness, respectively. 186 of the features were given a zero weight for all five emotions. Also, no features were used for all five emotions. Only three features (MFCC3-skewness, MFCC6-max, and MFCC8-amean) were weighted non-zero for four of the five. We examined, for each emotion, the five strongest (in terms of magnitude weights) to gain insight into which features are indicative of each emotion.

Anger Three of the five heaviest weighted features for anger classification are related to the MFCCs (MFCC2-amean, MFCC4-minPos, MFCC6-max). These are all lower to mid coefficients, indicating basic functionals of the vocal tract characterize angry speech. The other two features: F0-amean delta and energy-min are not surprising. It is expected that energy will impact the classification of anger as anger is usually associated with loud more intense speech. The changing of the $F0$ arithmetic mean over windows captures a feature of the pitch contour.

Despair Three of the five of the top heaviest weighted feature for the despair training are related to the MFCCs delta functionals. The model learns heavy weights for MFCC3-kurtosis delta, MFCC1-stddev delta, and MFCC10-stddev delta. This suggests that the changing of both the vocal tract and the glottal excititation are correlated with despair. Energy-maxPos also is heavily weighted with despair, indicating that when we act despairingly, we adjust where in an utterance we place the most energy. The final feature, voiceprob-delta kurtosis indicates the amount of voicing in depair utterances is somehow correlated with the emotion.

	A	D	H	N	S
Anger (A)	**29**	1	4	0	2
Despair (D)	1	**14**	7	5	8
Happiness (H)	3	8	**18**	4	2
Neutral (N)	0	0	0	**23**	2
Sadness (S)	0	5	2	9	**18**

Table 2: Confusion matrix for emotion classification using the $L-1$ regularized logistic regression classifier.

Happy All five of the top features for the happy utterances involve MFCCs. MFCC1-range, MFCC6-max, MFCC7-amean, MFCC9-linregerrQ and MFCC12-amean receive the top five weights (thought not listed in order of rank here). This suggests that the basic characteristics of our vocal tract behavior and our average glottal excitation correlated with what we consider "happy".

Neutral For neutral speech, the top five features were again all related to MFCCs. MFCC1-amean, MFCC1-linregerrQ delta, MFCC2-max, MFCC4-delta skewness, MFCC8-linrecg2. There is a lack of influence from the higher coefficients, which remains true when examing the top ten weighted features (which are all still derived from MFCC 1 through 8), indicating glottal excitation does not correlate with neutral versus emotional speech.

Sadness The heaviest weighted feature for sadness is the min energy. This is very expected, as we as humans associate sadness with less energy. The other top-four features include MFCC2-kurtosis delta, MFCC8-linregerrQ, MFCC4-skewness delta, and MFCC7-linregerrQ.

4 Error Analysis

We generated a confusion matrix for the emotion classifications which is included in the Table 2. As is consistent with the literature, we observe that neutral speech is the easiest to classify and also the most frequently recipient of a mis-classification. Sadness is misclassified as neutral 9 times, the most of any misclassification pair in our test set. This is not so suprising as neutral and sadness use similarly features in their classifiers. Despair segments are only correctly classified 40% of the time, making the emotion the most difficult to identify.