Algorithm*

Daryl Arredondo
Jiwon Choe
Max Chester
Kai He

This work is produced by OpenStax-CNX and licensed under the
Creative Commons Attribution License 3.0†

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.

*Version 1.1: Dec 19, 2011 1:16 pm -0600
†http://creativecommons.org/licenses/by/3.0/