Report Lab 3 Preparation

Pattern Recognition NTUA
9th Semester

Efthymios Tzinis 031212007
Konstantinos Kallas 03112057

December 19, 2016

Introduction

In this work we prepared steps 1 till 9 from the Laboratory Exercise 3. The main purpose of this Exercise includes the implementation of an automatic emotional recognition system which is dedicated to individual emotion classification of all the song files. All the song files were rated by 3 different people in order to provide an estimation of the valence and activation characteristics. In the early steps of this exercise we performed a statistical analysis of our data in order to see if the ratings are adequately interrelated and sufficient for training our classifier. After that we tried to extract some substantial for our task features by using the MIR Toolbox. Except of the recommende d features to extract, we emphasized in using the MIR Toolbox for specific emotional features through *"miremotion"*. We have used the Weka classifier for various homework exercises and for the 1st Laboratory exercise thus its interface is quite familiar to us. In the report below we describe briefly the procedure we followed. Our code is full of useful comments that are full explanatory.

Step 1

Before proceeding in the experiment there is need to preprocess the data. The data collection that was used for this experiment is 412 music pieces from Beetles songs that are 10-20 seconds long. During preprocess we just reduced the music quality in order to reduce computation complexity. We changed the music pieces from stereo to mono, we reduced the sampling rate in half and we also changed the bits per sample from 16 to 8. After this step we have a wav for each song loaded in Matlab.

Step 2

All music pieces have been tagged by 3 different people on their $Valence$ and *Activation* with a value in $1, 2, 3, 4, 5$. From these values we compute a co-occurence matrix for each labeler 1, 2, 3. As we can see all three labelers' tags follow a distribution that resembles a mixture of Gaussians with different parameters. All three labelers tag most pieces with a valence and activation between 2 and 4.
Labeler 1 seems to not tag many pieces with both low valence and activation. However he seems to tag both activation and valence with equally high values. In contrast labeler 2 tends to tag pieces with a higher activation value than valence. Labeler 3 seems to follow an opposite pattern by tagging pieces with generally higher values of valence than activation.

Step 3

We would like to be able to judge how "good" is our dataset of music pieces as a training and testing set for a classification model.

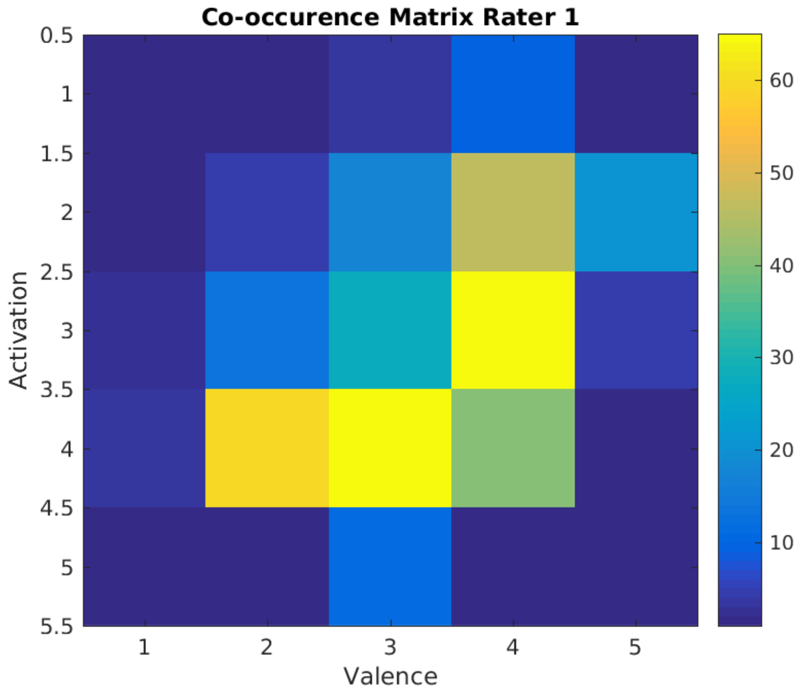

Figure 1

a)

In order to do this we will compute the observed difference between each pair of labelers. Because the observed difference is computed for each music piece we will first show a total observed agreement (mean) for each pair and each dimension (Activation and Valence).

	Activation	Valence
Labelers 1-2	84.59%	81.98%
Labelers 2-3	80.46%	83.31%
Labelers 3-1	79.61%	82.28%

As we can see all pairs of labelers generally agree in their tags.

b)

We demonstrate the 6 difference histograms, 3 for valence and 3 for activation 4. We can observe that most of the differences of all the 3 pairs of labelers are 0 and 1. This is a favorable result in general because it shows a general agreement between the labelers. However we need to use more complicated metrics to be sure that our data is labeled in a uniform way by all our labelers.

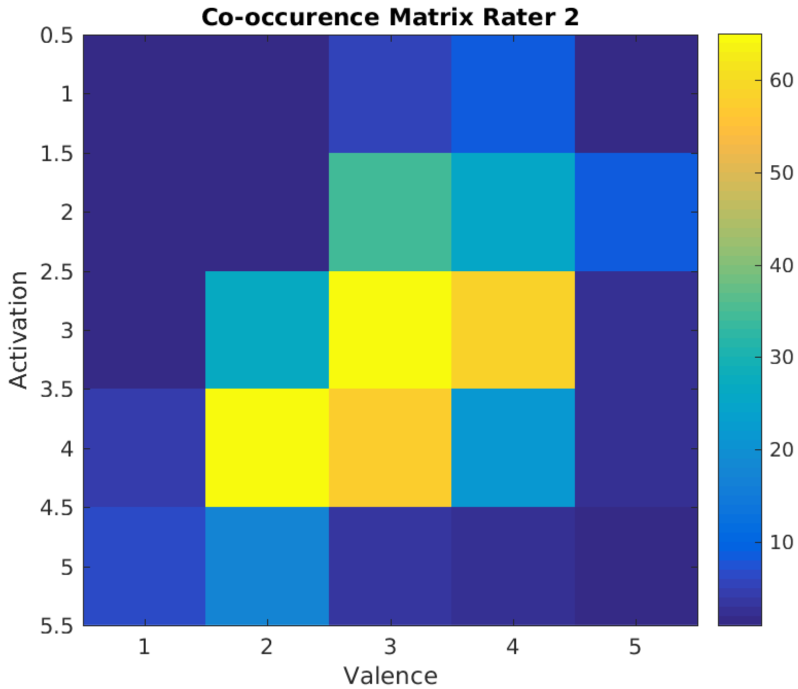

Figure 2

Step 4

We computed the Krippendorff's alpha coefficient for ordinal data among all the raters. We computed the coefficient according to the procedure followed in [1]. The equation for alpha is the following:

$$\alpha = 1 - \frac{Observed\ Disagreement}{Expected\ Disagreement} = \frac{D_o}{D_e}$$

Where D_o, D_e are derived from the following:

$$D_o = \frac{1}{n}\Sigma_c\Sigma_k o_{ck}\delta_{ck}^2, D_e = \frac{1}{n(n-1)}\Sigma_c\Sigma_k n_c n_k \delta_{ck}^2$$

The coefficients are different for the two dimensions of the valence and the activation. Where o_{ck} is the occurrence of the rating pair among all the songs between the 3 raters. For example, lets say the ratings for the valence of a specific song are (3,4,5) then we increase the $o_{35}, o_{45}, o_{34}, o_{53}, o_{54}, o_{43}$ as the matrix would be symmetric. We can refer to $O = [o_{ck}], c, k = 1, 2, 3, 4, 5.$ then $n_j = \Sigma_k o_{jk}$ and $n = \Sigma_j n_j = \Sigma_j \Sigma_k o_{jk}$. Because we have ordinal data the metric will be computed according to the aforementioned notation as:

$$\delta_{ck_{ordinal}} = \frac{\Sigma_{g=c}^{g=k} n_g - \frac{n_c + n_k}{2}}{n - \frac{n_{c_{max}} + n_{c_{min}}}{2}}$$

The term in the denominator: $n - \frac{n_{c_{max}} + n_{c_{min}}}{2}$ is a regulization term in order $0 \leq \delta_{ck} \leq 1$. But we can check if we computed the metric correctly more easily.

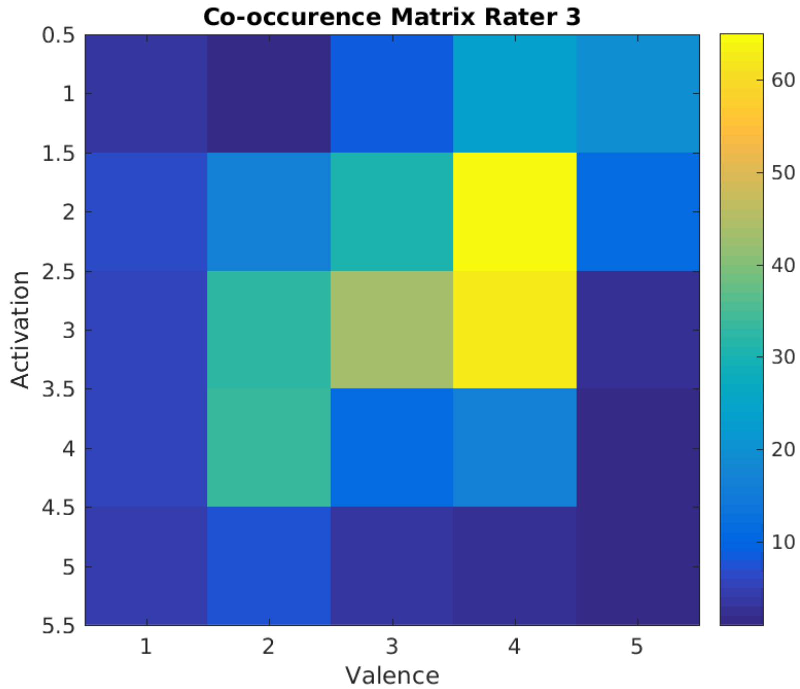

Figure 3

The respective matlab function in our implementation gives us the following results which were cross checked with an automatic Krippendorff's alpha coefficient computing implementation and were the same.

$$\alpha_{Valence} = 0.4745, \alpha_{Activation} = 0.4396$$

According to [1] the coefficient show acceptable data when $\alpha \geq 0.667$ and because of this, before we can properly train our classifier, we will need to throw out the songs with very deviant ratings in valence or activation. Comparing this coefficient with the previous step of Observed Agreement, the latter shows dubious results for the interrelated nature of our data.

In the next steps of the exercise we will try to perform a more sophisticated statistical analysis of the data by throwing away some of the data and then recomputing the same coefficient alpha to show the increment of our belief that the data are not comprised by outliers. This step is not mandatory but we portend that it will increase substantially our classification ratio. Another point that was discussed with the lab assistant and we concluded that a more credible statistical analysis would demand a coalesce of different statistical coefficients in order to verify that the data are actually interrelated or not.

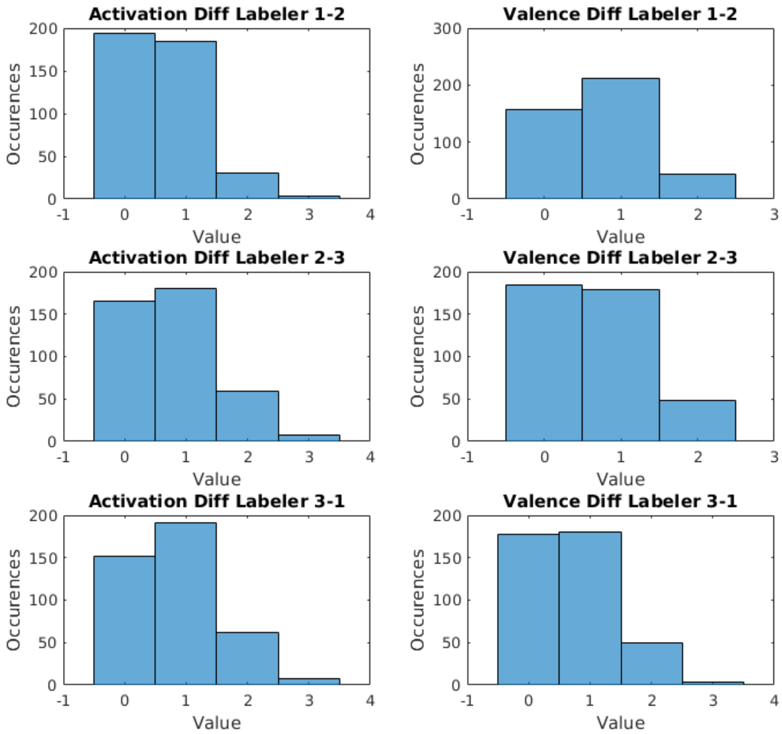

Figure 4

Step 5

As a final estimation of the rating for all our data set we will take the average of the votes of the three raters. Thus, the designated form of rating would be acquired naturally when dividing the sum of votes by 3.

In figure 5 the reader can see the 2D histograms of Valence vs Activation vs Number of Occurrences. It is limpid that the average of the three annotators as a final rating can be much more convenient for representation because all the higher values are concentrated in the middle of the histogram. Consequently, the final rating in both dimensions (valence and activation) demonstrates that both feelings are equally sensible. Of course this is something inviting for the purpose of our music classification task.

Step 6

In this step we used the MIR toolbox to extract some more features from the audio files in order to make emotion classifying easier. We used "miremotion" option as described in [2].
The features that were extracted are the following:

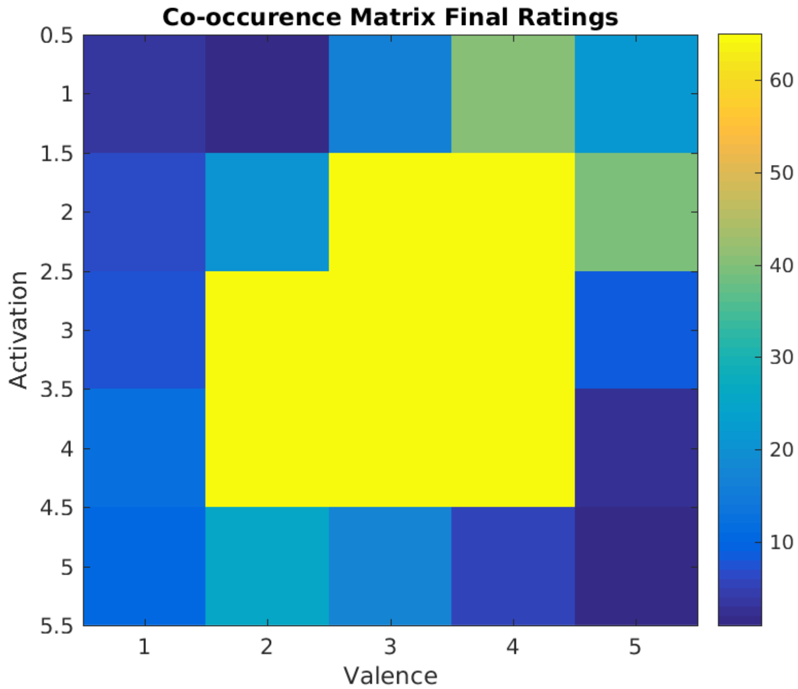

Figure 5

- **Roughness:** The roughness or sensory dissonance was computed for each music piece for each frame. In the end we kept 4 metrics of roughness across all frames as features, the mean , the standard deviation, the mean of the lower 50% values of roughness and the mean of the highest 50% of values.
- **Fluctuation:** The fluctuation or rhythmic periodicity was computed across all frames of all music pieces. The mean and max value were kept for each music piece as features.
- **Key Clarity:** The clarity of the tonal center positions has been computed across all frames and all music pieces. In the end the mean of its music piece was kept as a feature.
- **Modality:** The mean modality (major vs minor) was computed across all frames for each music piece.
- **Spectral Novelty:** The mean of the spectral novelty was computed across all frames of each music piece.
- **HCDF:** The Harmonic Change Detection Function was used in order to measure the change of the tonal content of each frame. The mean across all frames was kept as a feature for each music piece.

Step 7

We also computed the first 13 MFCCs, the first 13 MFCC deltas and the first 13 MFCC delta-deltas for each frame of each music piece. For those 39 coefficients we computed across all frames the mean , the standard deviation and the mean of the 10% lowest and 10% highest values.

In the end we gathered all the above features for each music piece in a vector of 166 dimensions. However before using the feature vectors for classification we shall consider normalizing their values.

Step 8

In the first lab exerice we used the SVMlib and Weka classifiers, as well as a custom knn and bayes classifier that we made with the use of MATLAB. Because of this reusing this classifiers is easy and the only requirement is writing some scripts to format the data in the correct way.

Step 9

We used Weka on the first Lab Exercise so we already have some experience on it.

References

[1] Krippendorff, K. Content analysis: An introduction to its methodology. Thousand Oaks, California: Sage.4 p.241

[2] Tuomas Eerola, Olivier Lartillot, Petri Toiviainen, "Prediction of Multidimensional Emotional Ratings in Music From Audio Using Multivariate Regression Models", International Conference on Music Information Retrieval, Kobe, 2009.