Multimodal Probabilistic Learning of Human Communication

Homework 2

Due date: Friday, November 4th, 11:59pm PT

The purpose of this homework is to give you hands-on experience with the process of training, validating and testing classifiers. To simplify the homework, we will be using the same dataset from Homework 1. The qualitative and statistical analysis you performed during homework should give you some insights about the most predictive features. A second document was created (homework2-dataset) describing the input features, the transcriptions and the sentiment annotations.

To successfully finish this homework, you should perform all the steps described below (including Data Preparation, Experiment A, Experiment B and Experiment C) and then prepare a document answering all the questions outlined at the end of this document. Please send your homework responses directly on USC Blackboard in PDF format. You can decide to include separate figures (also in PDF format) by zipping them together. Although your grade will not depend directly on the performance of you classifiers, we plan to share during class the best results.

> ➢ **IMPORTANT: For all experiments, be sure to always use exactly the same data splits for the testing process. This will assure that we can compare results. Specifically, hold out 25% for testing, 25% for validation and use 50% of the data for training. Remember to conduct your experiment on speaker-independent data, i.e. the speakers in the data-splits are disjoint. Using this split of 25/25/50, conduct the experiments 4 times in total (THESE MUST BE ENTIRELY SEPARATE EXPERIMENTS) so that every individual was used in the test set exactly once. Use the Scikit Learn library to run your experiments: http://scikit-learn.org/stable/**

Data Preparation

As a first step, you are asked to import all your audio-visual features and sentiment labels in the correct format for the training/validation/testing process. You should read carefully the separate document describing the dataset (homework2-dataset). To help you with the experiments.
Before importing and preparing the data, it is good to look at your homework 1 results. You want to identify a subset of features that are likely to be helpful for the sentiment classification task.
- We ask you to select at least 3 different acoustic features and 3 different visual features. You are free to use as many feature as you want. Don't forget that each feature should be defined at the segment level, as you did during the homework 1. This process is sometime referred as feature engineering.
- You should import all your engineered features. You should define cell arrays *sequences/data* and *labels,* which should have the same length, equal to the number of valid video sequences in the dataset.
  - Confirm that the data containers hold matrices where the number of columns is equal to the number of engineered features and the rows is equal to the number of segment (different for each sequence).
  - The arrays of *labels* should contain vectors where the number of element is equal to the number of segment (different for each sequence).

- You should take time to read the documentation. Be sure that you understand the parameters of the scikit learn library (identify hyper parameters, etc. for validation/test).

Experiment A: SVM Classifier and Validation Strategies

Your first experiment will be to compare different strategies for automatically selecting the hyper-parameters. As we studied during the course, classifiers such as Support Vector Machine (SVM) have parameters that are not directly optimized during training. For a linear SVM model, the will have a regularization constant C which needs to be automatically validated. A good rule of thumb is to use the logarithmic scale: [10E-2, 10E-1, 10E0, 10E1, 10E2].

You should note that for this homework, we will focus on *early fusion* where all multimodal features are concatenated in larger input feature vectors. Another approach would be to perform *late fusion* where classifiers are first trained for each modality and then combined with a second layer classifier (aka fusion classifier).
- Set the parameters of your learning procedure so that it performs 4-fold testing and hold-out validation (use 25% of the training sequences for validation).
- Set the parameters of your learning procedure so that you will train SVM classifiers with linear kernel and validate the C hyper-parameters with values on the log-space from 0.001 to 100.
- Train, validate and test this model. You should examine the results and plot the respective validation and testing accuracies for each test fold.
- Repeat the same experiment but change the validation mode to 3-fold and retrain the model. Without respecting speaker-independence. Look again at the results and plot the validation and testing accuracies for each test fold. Do you see anything interesting?

Experiment B: Compare Performance of different Modalities

For the second experiment, we want you to experiment with different input modalities. Specifically, we would like you to compare the multimodal classifier you previously trained (in Experiment A) with only acoustic features and with only visual features.
- Create two new feature sets: one for acoustic only features and the second for visual features.
- Train SVM classifiers (using the same test splits as before) using only the acoustic features. You should examine and plot the results.
- Train a second set of SVM classifiers using only the visual features. You should create a comparative chart comparing acoustic-only, visual-only and multimodal.
- Compute the two-sample t-test between the test results of the multimodal classifier and the test results of either the acoustic or the visual classifiers.

Experiment C: Compare Performance of Multiple Classifiers

For your last experiment, you should compare the linear SVM classifier with two other classifiers such as a neural network, naïve Bayes classifier or SVM with RBF kernel.
- Include in your script the parameter structure for the two new classifiers you want to train.
- Train these two next classifiers. Look at the results and create some comparative plots between the different classifiers.

- Optionally, you can start experimenting with classifiers that are also modeling temporal information such as the Conditional Random Field (CRF), Latent-Dynamic CRF (LDCRF) or the Concatenated Hidden Markov Model.

REPORT

For your homework 2 report, you should include the following items:

1. **Data preparation:** you should describe in details how you created your unimodal (or multimodal features). As mentioned in the Data Preparation, you should have at least 3 acoustic features and 3 visual features. Your description should be sufficient for someone else to recreate your feature set. (~0.5 pages; you can add figures too)
2. **Validation strategies (~1 page):** Create the following tables to analyze the effect of validation strategies:
   a. Create a first table showing the accuracies of your linear SVM classifiers when validated with holdout strategy (in Experiment A). You should have three column: training accuracy, validation accuracy and test accuracy. You should have one row for each of the test split.
   b. Create a second table for the 3-fold validation strategy (same layout as the first table).
   c. Do you expect the validation accuracies and test accuracies to be similar? Which validation strategy should give you better similarity between validation and test accuracies? Discuss the differences in these two tables.
3. **Different modalities (~0.5-1page):**
   a. Insert your graph comparing acoustic-only, visual-only and multimodal classifiers.
   b. If some of these pairs are statistically significant (based on the two-sample t-test), include these differences in the graph (using stars *).
   c. Discuss any differences between modalities. If possible, make some references to the qualitative observations you made during the homework 1.
4. **Multiple Classifiers (~0.5-1page):**
   a. Insert a graph comparing the three classifiers you trained and evaluated. Include any statistically significant differences in your graph.
   b. Find which classifier performs the best and write one paragraph describing the **Methodology** you used to train, validate and test this classifier. You should include enough details so that a reader can recreate your experiment.
   c. What would make one of the classifier better in this setup? What would you expect if you had more data? Which classifier would you use in this case?