# INCREMENTAL ADAPTATION USING ACTIVE LEARNING FOR ACOUSTIC EMOTION RECOGNITION

The 42nd IEEE International Conference on Acoustics, Speech and Signal Processing
ICASSP2017
THE INTERNET OF SIGNALS

**Mohammed Abdelwahab and Carlos Busso**

**Multimodal Signal Processing (MSP) Laboratory**
Erik Jonsson School of Engineering & Computer Science
University of Texas at Dallas

## MOTIVATION

**Background:**

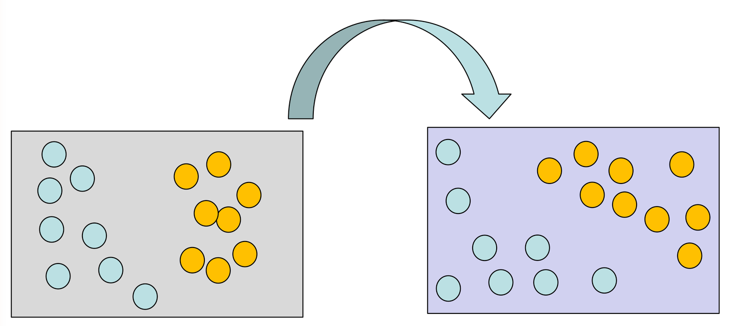
- Classifier performance degrades when training and testing conditions are different.
- Supervised domain adaptation is normally used to improve the base classifier's performance.
- The performance increase depends on the data used for adaptation.

**Proposed Solution:**

- Active learning can be used to annotate the most useful samples to the classifier.
- Adjust hyperplane while maintaining learned information.
- Conservative approach that incrementally modifies the hyperplane with consistent samples.

## Framework

### The Problem



source target

### Domain Adaptation

**Adapt SVM:**

$$f(\boldsymbol{x}) = f^s(\boldsymbol{x}) + \Delta f(\boldsymbol{x})$$
$$= f^s(\boldsymbol{x}) + \Delta \boldsymbol{w}^{\boldsymbol{T}} \phi(\boldsymbol{x})$$

$$min \ \frac{1}{2}\|\Delta \boldsymbol{w}\| + C\sum_{i=1}^{N} \xi_i$$
$$s.t.\ \xi_i \geq 0,$$
$$y_i\left(f^s(\boldsymbol{x}) + \Delta \boldsymbol{w}^{\boldsymbol{T}} \phi(\boldsymbol{x}_i)\right) \geq 1 - \xi_i$$

### Databases

**Source: USC-IEMOCAP**

- 12 hours of recordings
- Scripts and improvised scenarios
- Turns are annotated with emotions
  - Angry, Happy, Sad and Neutral

**Target: MSP-IMPROV**

- Over 9 hours of recordings
- Improvised scenarios
- Turns are labeled with four emotions
  - Angry, Happy, Sad and Neutral

### Proposed Approach

**Active learning**

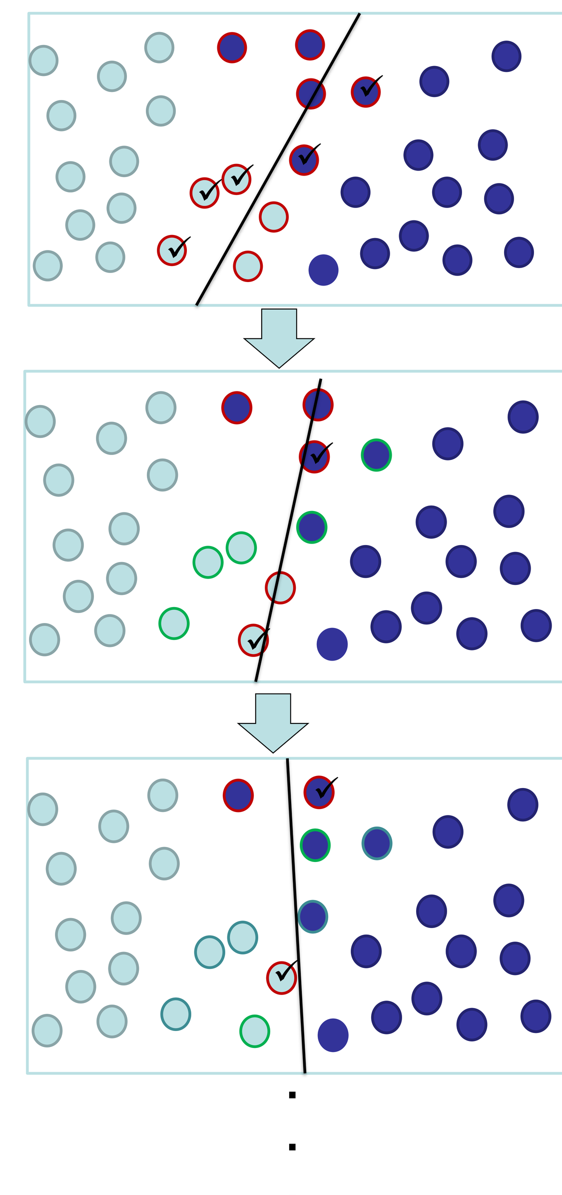
- Identify samples with low confidence
- Annotate samples

***while*** stopping criteria is not met ***do***

- Select subset $N_a$ that the classifier predicted correctly
- Adapt classifier using subset $N_a$

### Stopping Criteria

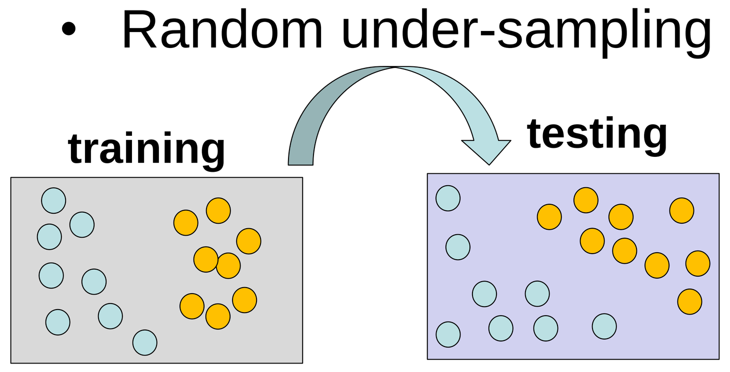
- Criterion 1: $N_a$ doesn't contain labels of all classes
- Criterion 2: $N_a$ contains labels of only one class
- Criterion 3: All samples are used



## Experimental Settings and Results

**Settings**

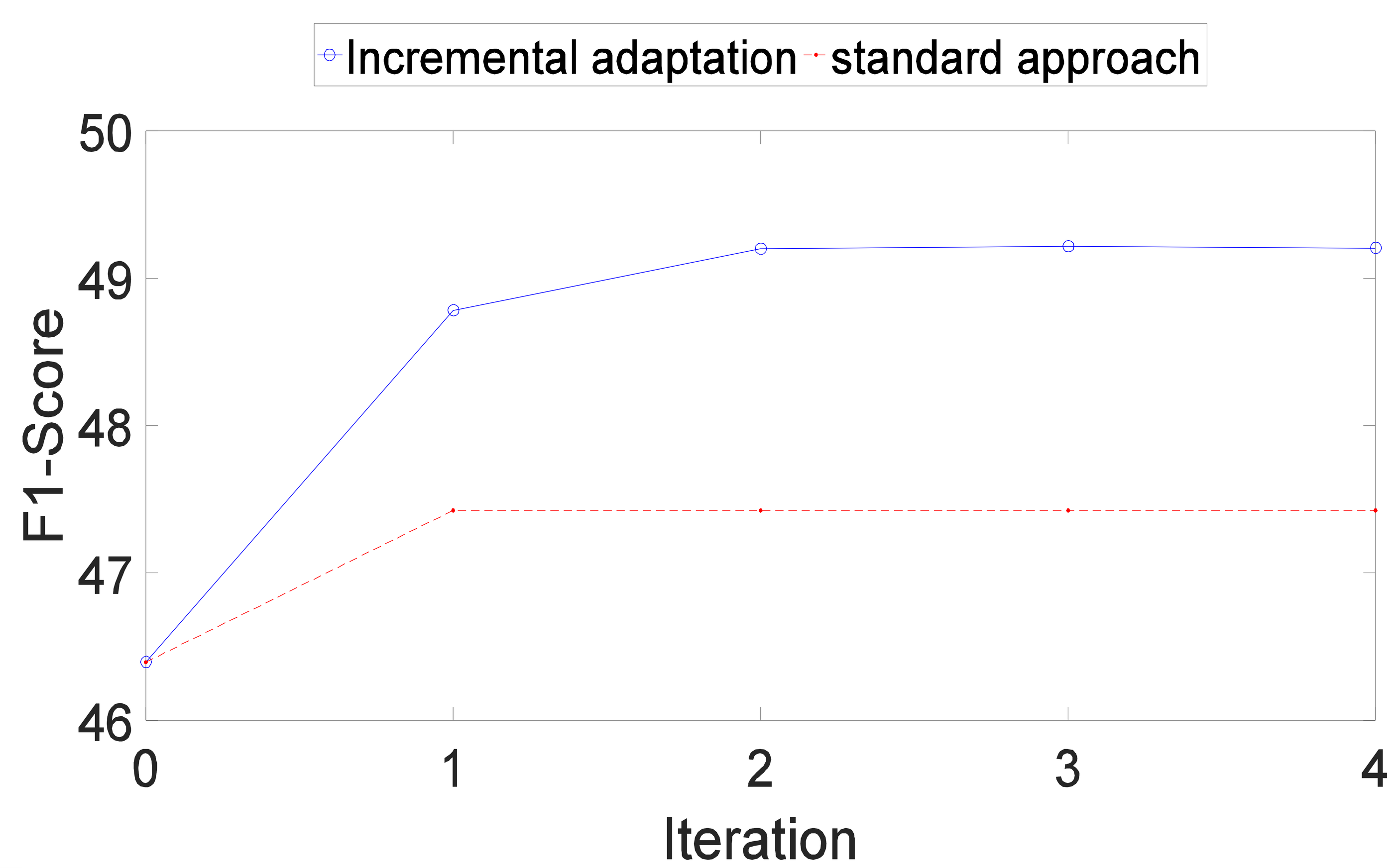
- Interspeech 2013 feature set
- Feature Selection
  - Correlation Feature selection 6373 ⇨ 3000
  - Forward Feature Selection 3000 ⇨ 300
- SVM Classifier with a linear kernel
- Four class balanced classification problem
  - Angry, Happy, Sad, Neutral
  - Random under-sampling




Source USC-IEMOCAP    Target MSP-IMPROV

- Active learning select 200 samples
- Results are the average of 20 trials
- Baseline approach is adapting with all samples

**Results**




**When to stop**

**No adaptation: 45.5 %**
**Baseline approach: 46.7 %**

| criteria | # samples | F1 score After adaptation | # iterations |
|---|---|---|---|
| 1st iteration | 64.4 | 47.78 % | 1 |
| Criterion 1 | 117.8 | 48.28 % | 3.71 |
| Criterion 2 | 123.6 | 48.13 % | 4.71 |
| Criterion 3 | 200 | 45.47 % | 5.71 |

**Key Point**

Carefully selecting the samples used in adaptation yields better performance

## DISCUSSION

**Conclusions:**

- Proposed an algorithm for incremental supervised SVM domain adaptation.
- We showed the importance of selecting the data used for adaptation.
- We used a portion of the labeled dataset, converging to a stable performance after 3 to 5 iterations.

**Future Work:**

- Modify the optimization function so that we can make use of all of the available data.
  - Introducing a variable regularization parameter for each instance.
- Extend the proposed algorithm to other classifiers.

**Acknowledgment: This study was funded by a NSF CAREER award IIS-1453781**