Domain Adaptation Challenge 2013

Advances in sub-space modeling, specifically the i-vector approach, have demonstrated dramatic and consistent improvement in speaker detection performance on the NIST speaker recognition evaluations over the past 4 years. However, these techniques are highly-dependent on having access to large amounts of labeled training data from thousands of speakers each making tens of calls to train the hyper-parameters (UBM, total-variability matrix, within and between covariance matrices). The archive of past LDC data collections has provided such a set of data for the NIST SREs and been used effectively. However, it is highly unrealistic to expect such a large set of labeled data from matched conditions when applying a speaker recognition system to a new application. Thus there is a need to focus research efforts on how to use unlabeled data for adapting and applying i-vector speaker recognition systems.

MITLL has created an experiment design based on LDC telephone corpora which demonstrates the effect of data mismatch on hyper-parameters and poses a challenge task on which researchers can focus. In this experiment, SRE10 telephone data is used as enroll and test set. Specifically, we are using the combined one conversation (1c) and (8c) eight conversation telephone data enroll and one conversation (1c) telephone data test lists.

We have designated two data sets to be used for hyper-parameter training:
- SRE: this consists of all telephone calls from all speakers taken from the SRE 04,05,06, and 08 collections. This will serve as the "matched" hyper-parameter training list.
- SWB: this consist of all telephone calls from the all speakers taken from switchboard-I and switchboard-II (all phases) corpora. This will serve as the "mismatched" hyper-parameter training list.

Below is a table of some key statistics of these two data sets.

Table 1: Summary statistics for the SWB and SRE data sets used for hyper-parameter training.

Data Set	# spkrs	# Males	# Females	# files	Avg files/spkr	Avg phone_num/spkr
SWB	3114	1461	1653	33039	10.6	3.8
SRE	3790	1115	2675	36470	9.6	2.8

These two data sets have very similar statistics and are all telephone speech. The current expectation is they should produce similar results. However, in the DET plots below, it is clearly demonstrated that there is a gap in performance on the SRE10 enroll/test set when hyper-parameters are trained with the different sets. DETs for 1-conversation and 8-conversation enroll, using hyper-parameters trained only on the SRE data or only on the SWB data are shown for three different sites with independent ivector implementations. This demonstrates the performance gap is not a function of particular implementation details (features, SAD, hyper training algorithms, etc).
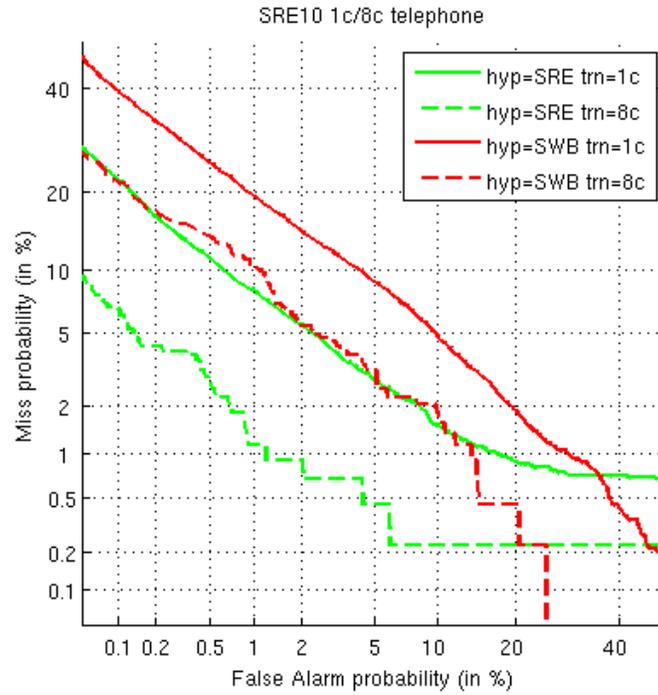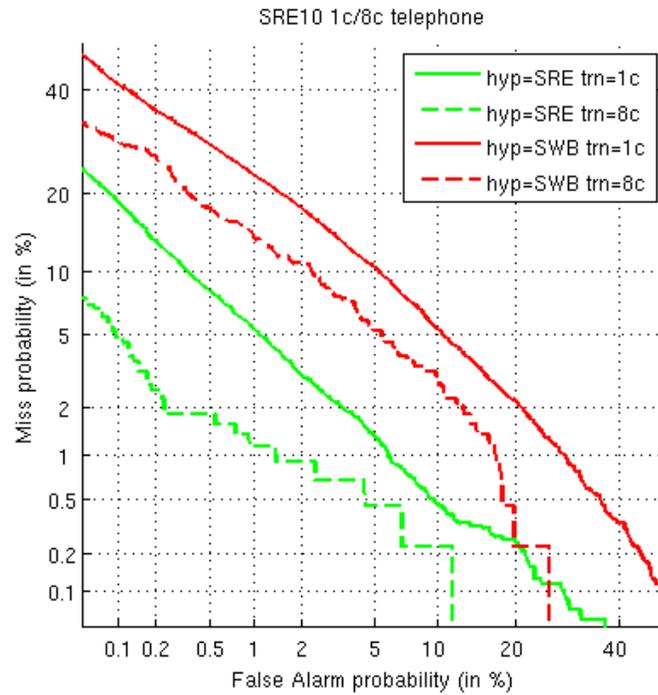
Figure 1: Challenge results from MITLL

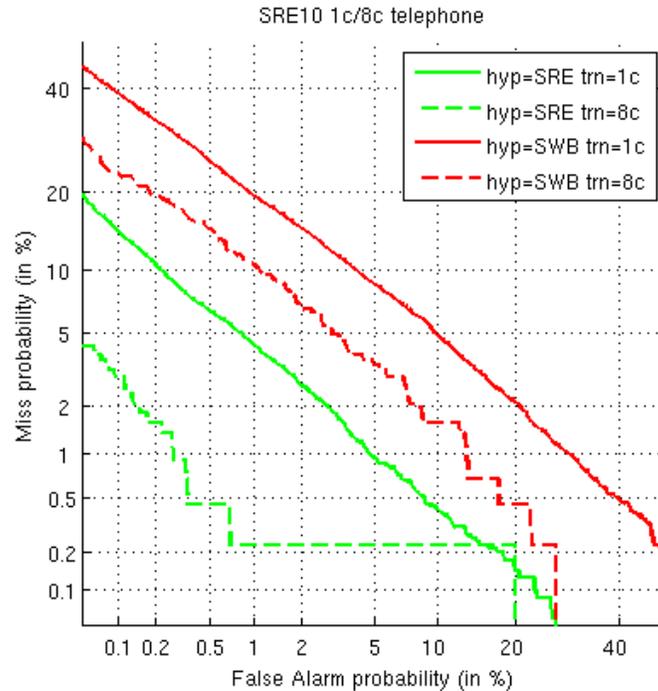**Figure 2 Challenge results from MIT-CSAIL**

Figure 3 Challenge results from JHU-HLTCOE

The main challenge task is to effectively exploit unlabeled data. The aim is to develop approaches and techniques to effectively use the SRE data without any speaker labels to move performance closer to the known gold-standard when the SRE data is used with labels. Only the labeled SWB data and unlabeled SRE data may be used to avoid making this a data-engineering exercise. The aim is to come up with a recipe that can be applied in future situations where only unlabeled data from a new domain is available. Another aim is to develop tools to analyze hyper-parameters and data to detect when a mismatch is present and likely to produce degradation in performance.

The lists, ivectors, and baseline results for this task are available in:

http://www.clsp.jhu.edu/workshops/archive/ws13-summer-workshop/groups/spk-13