Automatic Analysis of Motivational Interviewing with Diabetes Patients

Xizi Wei, Martin Russell, Peter Jancovic
School of Engineering, University of Birmingham, Birmingham B15 2TT, UK

Introduction

- Motivational Interview (MI) is one kind of goal-driven Clinical Conversion (CCs) between a clinician and patient that seeks to facilitate and engage a patient's intrinsic motivation to change his or her behavior. It's a significant part to support active self-management of diabetes patients' conditions.
- Conventional assessment of MI is based on human analysis of the transcriptions of a clinician-patients dialogue to measure the clinician's compliance with guidelines, which is labor-intensive, time-consuming and financial costing.
- In this study, an automatic assessment system is built to make automatic assessment of MI.

Current work

- Built and adapt a DNN-HMM ARS system to transcribe the speech
- Built a speaker diarization system

Data set:

In-domain data

Recordings of Motivational Interviews provided by Institute of Psychiatry , Psychology and Neurosciences (IoPPN).

- 170 coded 10-minute transcribed recordings of MI sessions with Diabetes patients
- 8 full transcripts of 25 minutes each are available
- For ASR:
  - MI_train: 8h segmented and transcribed data
  - MI_test: 6 interviews,0.7h in total

AMI meeting corpus

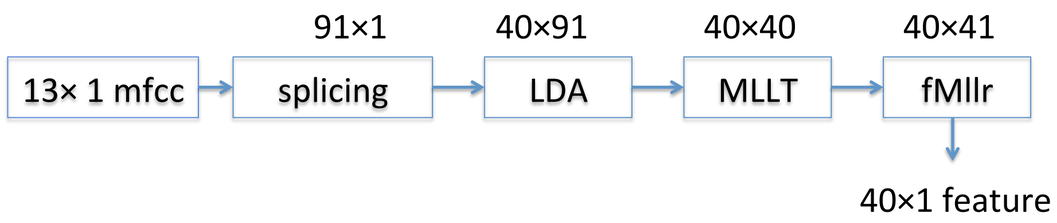
- single distant microphone (sdm) record set – 70h
- spontaneous speech

Building DNN-HMM ASR System:

Sequence of systems for GMM-HMM baseline:

(1) monophone system
(2) triphone system with MFCC + Δ + ΔΔ
(3) triphone system with LDA + MLLT
(4) Triphone system with LDA+ MLLT + SAT

Baseline features for DNN training:



Baseline DNN-HMM system trained on AMI+MI

- State alignments from GMM base LDA+MLLT+SAT system
- Senone set: 3981 senones
- DNN model size: $440\times2048^{6}\times3981$

DNN adaptation

- Adapt baseline DNN using MI only
- Re-alignment from baseline DNN-HMM ASR system
- Senone set selection:
  - 3981 senones
  - Smaller senone set from alignment of MI data-1538 senones

Language model

- 303 out-of-vocabulary words were added to the BEEP dictionary
- A trigram language model is trained on the 171 transcripts of MI recording.

Results:

Model	Senone No.	WER (%)
Baseline DNN-HMM	3981	53.13
MI adapted DNN	3981	47.24
	1538	48.68

Table 1. The ASR results on MI_test.

Building Speaker Diarization System:

1. Use manual segmentation to extract ivector for one interview

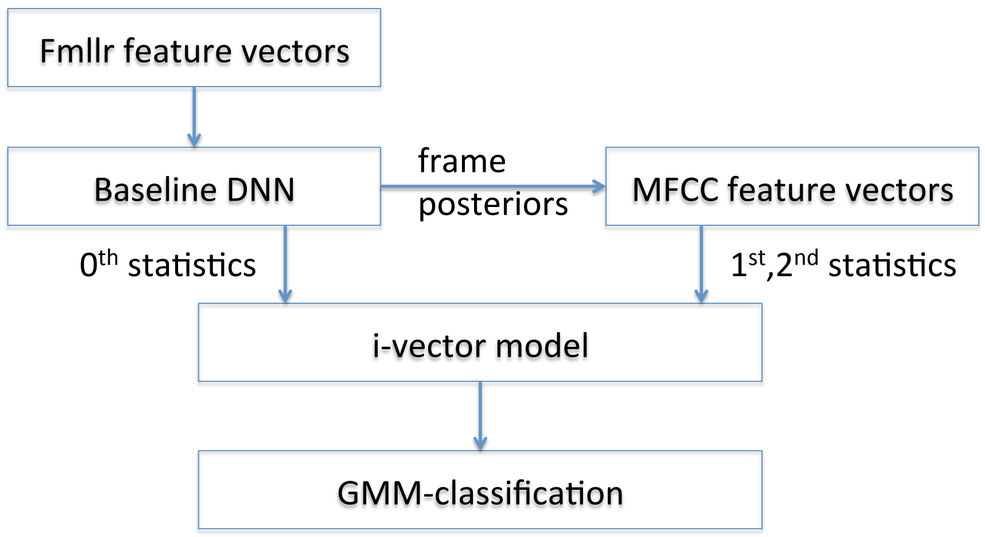


Fig. 1. The flow diagram of the DNN ivector farmework

Results: accuracy 94.8

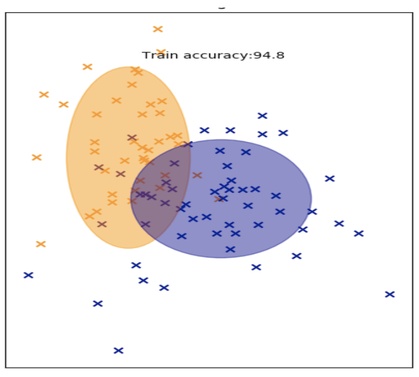


Fig. 1. The classification result of a 2-component GMM

2. Automatic segmentation

- Temporal Segmentation
  - 2 second
- Gmm-ivector
- 2-component GMM classification
- Evaluate the two segmentation
  - Lium
  - Merge the 2s segment

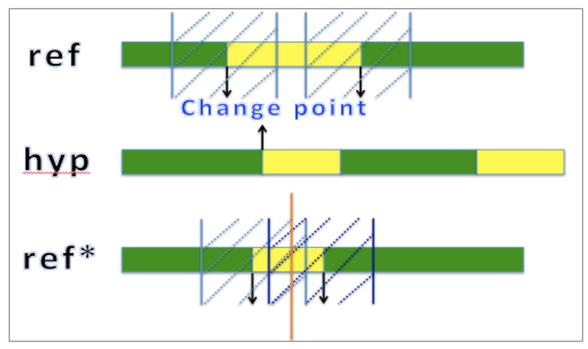


Fig. 2. The evaluation diagram : window size 2s

	H_{hit}	N_F	N_{REF}	PRC	RCL
lium	35	144	93	0.24	0.38
ivector	42	159	93	0.26	0.45

Table 2. the evaluation results of segmentations by lium and ivector: H_{hit} number of correct detected change point, N_F number of change points in the hypothesis, N_{ref} number of change points in the reference, PRC precision , RCL recall

UNIVERSITY OF BIRMINGHAM