CRF-based Confidence Measures of Recognized Candidates for Lattice-based Audio Indexing

Zhijian Ou, Huaqing Luo

ozj@tsinghua.edu.cn, luohuaqing@live.cn

Department of Electronic Engineering, Tsinghua University, Beijing 100084, China

Motivation

- The use of FB posterior probabilities as the confidence scores seems to be common across various audio indexing systems.
  ☹ A major limitation: its performance for CMs cannot be improved easily.
- Most CM studies seek CMs mainly for the recognized 1-best (**the 1-best case**) rather than for all recognized candidates in a lattice (**the lattice case**).
- How to effectively compute CMs for the lattice case is the main issue studied in this paper.

Computing CMs by combining various relevant features

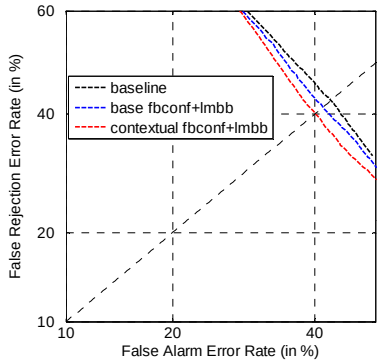
The 1-best case :

J. Fayolle, et al., "CRF-based combination of contextual features to improve a posteriori word-level confidence measures", Proc. Interspeech, 2010.

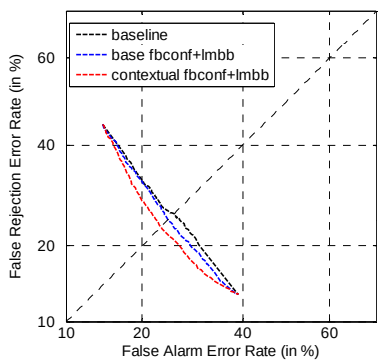
- introduces (linear-chain) conditional random fields (CRFs) to do sequential labeling and uses contextual features.
- The posterior $p(q_n = \text{'correct'} \mid y)$ is used as the CM for the n-th word.

The lattice case :

- It is **not trivial** to extend CRF-based CMs from the 1-best case to the lattice case.
- **First**, (linear-chain) CRFs are probabilistic models suited to label sequence data, while the lattice from the ASR decoder is not sequential.
- **Second**, Contextual features are defined over a word sequence. We need to figure out methods to extract such features from the lattice, which is not as straightforward as from the 1-best transcription.



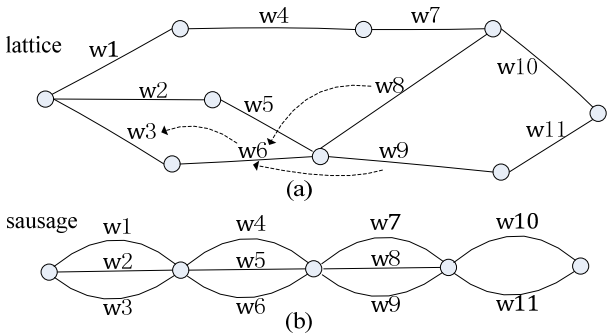
The DET curves for confidence measures



The DET curves for keyword search in audio indexing

CRFs for CMs in the lattice case

1. Reduce lattice to sausage



2. Define CRF over sausage

The conditional distribution $p(q|y)$:

$$p(q|y) \propto exp\left\{\sum_{n=1}^{N}\phi_n(q_n,y) + \sum_{n=2}^{N}\psi_n(q_{n-1},q_n,y)\right\}$$

$$\psi_n(q_{n-1},q_n,y) = \lambda^e_{e(q_{n-1},q_n,y)}$$

edge potential function

$$\phi_n(q_n,y) = \lambda_{\text{other}}1(q_n = \text{'null'}) + \bar{\lambda}_{\text{other}}1(q_n \neq \text{'null'})$$

$$+\sum_{k=1}^{K_n}\sum_{f}\left[\lambda^f_{f(ARC^n_k)}1(q_n = ARC^n_k) + \bar{\lambda}^f_{f(ARC^n_k)}1(q_n \neq ARC^n_k)\right]$$

node potential function

The posterior probability $p(q_n = ARC^n_k \mid y)$ is used as the CM for word arc ARC^n_k

Experimental results

Table 1: performance comparisons between the baseline and the CRF-based approach, using different configurations of features for **CMs** (in terms of EER) and **audio indexing** (in terms of keyword search EER).
#weight denotes the total number of weights used in the CRF model.

features used for the CRF		CM EER	keyword search EER	#weight
baseline		**42.63%**	**25.70%**	
base	*fbconf*	41.58%	24.83%	9
	lmbb	42.96%	26.50%	9
	fbconf+lmbb	41.20%	24.53%	15
contextual	*fbconf*	40.51%	23.98%	297
	lmbb	41.52%	25.50%	65
	fbconf+lmbb	**40.13%**	**23.24%**	**359**